.

THE USE OF CONCORDANCE PROGRAM IN STUDENTS' RESEARCH

Yanty Wirza, M.Pd
Indonesia University of Education
yanty_wirza@yahoo.com

This study examined the use of Corpus Corcordance Program in students' research reports along with some difficulties they encountered in doing the studies as well as in analyzing the corpus data. This investigation serves as one of the efforts to improve the newly-introduced Corpus Linguistics subject offered to undergraduate students at English Literature Department at Indonesia University of Education. There were two classes under this investigation involving 58 students. In their research reports, the students were expected to choose a topic and analysed corpus data of at least 100.000 words, quite a small amount of data for corpus studies. However, this amount was considered adequate for the purpose of practising the program. Interviews and one-on-one discussions were also carried out to provide proper guidance and find out the difficulties encountered by the students in doing this task.The results show that despite some limitations in administering the subject especially the inavailability of computer laboratory and corpus software, the students demonstrated extra efforts in finding the free corcordance program from the internet and used it to analyze the data. However, most students found it hard to satisfy the amount of words to analyze due to some technical problems. In analysing the data, the students related the words or espressions of their own selection to relevant theories. Some students came out with really interesting research findings highlighting the students' personal interests in the research topics.

Background

The advanced developments in the computational linguistics have brought some significant changes in the curriculum of English Language Education – Indonesia University of Education that starting from 2006, the undergraduate students were offered to take the course of Corpus Linguistics. This four-credit subject was immersed into the curriculum to serve the purpose of introducing undergraduate students to the more intensive and extensive use of corpora and concordance programs that have become much more widely available especially through the internet connections. Researchers now believe that this can lead to a new literacy that can contribute to process-oriented approach to language learning (Sullivan, 2007).

.

During relatively short period of socializing the subject of Corpus Linguistics and what it has to offer, many lecturers and students still do not know much about the use and implication as well as relevance of corpus for language learning and pedagogy. Now that we have come to the third year, we tried to explore what we could do with corpora and concordance programs, given that we only searched for the frequency of the words in the last two years.

History of Corpus Linguistics

Corpus (stems from Latin "body) refers to a collection of linguistic data upon which some general linguistic analysis can be conducted (Meyer, 2004); it serves as a methodological basis for pursuing linguistic research (Leech, 1992). With this definition in mind, we are not talking about one or two texts but a large amount of linguistic data. The Expert Advisory Group on Language Engineering Standards (EAGLES) defines a corpus quite generally, saying that it "can potentially contain any text type, including not only prose, newspapers, as well as poetry, drama, etc., but also word lists, dictionaries, etc." ("Corpus Encoding Standard": http://www.cs.vassar.edu/CES/CES1-0.html)

The history of Corpus Linguistics can be viewed consisting of two parts; early and modern days. Early corpus linguistics is the term used to describe corpus linguistics before the advent of Chomsky. Chomsky invalidated the corpus as a source of evidence in linguistic inquiry. With his competence-performance model of language, he insisted that performance is a poor mirror of language competence, on which the model of language

.

should be based upon (McEnery and Wilson, 2007). To balance Chomskian's rejection towards corpus data, Malmkjer and Anderson (1991, Eds) propose their arguments:

1. Corpus data are first-hand textual data that cannot be meaningfully analyzed without the interpretative skills of the analyst using the knowledge of and about the language. In other words, the corpus data are not the performance per se.

2. Corpus studies in most cases flourish in countries where English is not a native language and serves as the second or foreign language because the needs for corpus evidence are greater for them than those of a native speaker.

3. For the purposes of language learning, the language inputs of how the language is used both by native speakers and by learners are relevant. Thus, to rely on the competence and intuition along is not suffice since native speaker's competence of the language can be incomplete in relation to other aspects of language use such as social and cultural phenomena.

4. The development of branches of linguistics such as sociolinguistics, psycholingistics, pragmatics, and discourse analysis have made significant use of performance-based data.

5. The Natural-language Processing (NPL) by computer should process unrestricted naturally occurring data driven from performance data as authentic textual data.

Despite Chomsky's criticisms, linguists kept on conducting research using performance data and came up with major development in language studies. Some of the early corpus linguistics works are the study of language acquisition from the early to middle of 19th

.

century which were based on parental diaries recording the child's locutions as well as large numbers of children utterances and expressions aimed at establishing norms of language development. These studies were typically longitudinal studies involving a small number of children.

Other studies which contributed to the language learning are those conducted in the field of language pedagogy. The corpora were used to provide vocabulary lists for foreign language learners. The word counts derived from these studies were important in defining the goals of vocabulary control movement in second language pedagogy. Other examples using corpora were the study of comparative linguistics, syntax, and semantics.

Modern corpus development rocketed significantly since the era of computer technology. Now the term corpus is almost synonymous with the term machine-readable corpus. The technology in the computer system makes it possible for corpus linguists to carry out various processes. One of the processes which can be done is concordance program. Through this program, the computer has the ability to search for a particular word, sequence of words, or perhaps a part of speech in the collection of texts. Then it can retrieve the all examples of how the word is used in contexts. The machine can find the relevant texts and display it to the user. It can further calculate the number of occurrences of the word to gather the information about the frequency of the word used in contexts. The advantages of machine readable corpus are derived from the capability of automatic processing, as described above, and automatic transmission which includes transferring a text either locally (e.g. from the computer storage), or other electronic links such as internet connections.

.

Since the computer era, there are many computer corpora of modern English. To name a few are the LOB Corpus(1978), which is a corpus of printed British English compiled in order to match as closely as possible the Brown Corpus of American English (1961) that includes certain registers. London Lund Corpus (1982) is a corpus of approximately 500.000 words of spoken English, transcribed in detailed prosodic notation. It was computerized at Lund University, Sweden. The example of the corpus of literature is The Leuven Drama Corpus which consists of 1 million words of British dramatic texts. Another large and growing collection of machine-readable is The Oxford Text Archive which includes texts of various languages and various historical periods.

These machine-readable corpora are not without limitations. The flaws come from, among other things, the amount of words accommodated by the machines. Another reason for their limitation is the time when the corpora were compiled. Thus corpora should be always updated to represent the language use and phenomena of the time. The types of genres and registers as well as the medium of the language use (spoken vs written) covered by the corpora also mark their constraints.

The Use of Corpus-Based Approach in Teaching and Learning: Some Studies

As estimated, the application of corpora and corpus programs to the education sphere develops more rapidly as cheaper and more powerful hardware is coming within the educational budgets. The use of concordances as language-learning tool is currently a major interest in Computer-Assisted Language Learning (CALL) (Malmkajer and Anderson, 1991). In the same vein, Milroy (1987) asserts that with better capability of computer to

.

handle big amount of data, corpus could be an important and practical tool of analysis. In addition to that, according to Bahtia (2002 in Carmen, 2009) corpus offers multi-perspective model for discourse analysis:

- **textual perspective**, helps students identify grammar items through statistical frequencies, collocational patterns, context-sensitive meanings and discoursal uses of words
- **genre perspective**: provides students with exposure to recurrent lexico-grammatical patterns across different academic text types (genres)
- **social perspective**: raise learners' awareness of how speakers' different discourse roles, discourse privileges and power statuses are enacted in their grammar choices.

Nowadays, there are a lot of corpus-based research carried out which take foci of different areas in language education. A study by Liu and Jiang (2009) which examines the use of corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts in China leads to the conclusion that the approach has several positive effects. Those effects are the improvement of the students' command of lexicogrammar, the increasing critical understanding of grammar, and the enhancement of learning skills. Charles (2007) conducted a research on top-down and bottom-up approaches to academic writing that used corpus to teach rhetorical functions. The study shows that the two approaches provided the enriched inputs necessary for the students to make connections between general rhetorical purposes and specific lexico-grammatical choices. In addition to that, a study by O'Sullivan (2007) developed corpus consultation literacy that can enhance a process-oriented approach to language teaching and learning. It is envisaged

.

that this consultation program would contribute to the establishment of a sound theoretical and pedagogical foundation.

Using the corpus programs in education requires no specific and in-depth knowledge of computer science or mathematics. What matters the most is that the students should understand (1) when it is appropriate to use the programs, (2) what data are needed, (3) how to interpret the results (McEnery and Wilson, 2007). This last requirement of interpreting the results of corpora is really important in flourishing the students' academic skills.

The Study

The study was conducted to figure out the use of concordance program in the students' research. The subject was aimed at familiarized the students with the growing trends to use corpus programs in the field of education, as suggested by McEnery and Wilson (2007) above and the multi-perspective use of corpus as proposed by Bahtia (2002). After dedicating several sessions on the theory of Corpus Linguistics and its applications in many branches of language studies, the students were to have some practice using the concordance software. The decision to select the concordance software program was because it was mostly use in education. This software has the ability to search, retrieve, calculate, and sort certain words/phrases in a very short time and display the results of a given search in KWIC (Key Word In Context). However, due to some limitations, the students were assigned to find the free-trial concordance software program from the internet.

.

The participants of this research were two classes of fifth semester students. After practicing using the concordance software program, they were requested to make a research proposal in which the data were taken from corpora. The topics of the research were not restricted so that the students could choose the genres of their own selection. This was aimed also to have more engaging classroom discussions providing that the choices made by the students would vary. Various applications of the concordances software program would give the students exposure to different lexico-grammatical patterns across different text types (genres) (Sullivan, 2009). This would make the subject more interesting to them.

In their research, the students were inquired to analyze corpora of no less than 100.000 words. This is quite small amount of words for a corpus analysis; however, this is viewed to be sufficient for a practice of small-scale research. With some limitations to the computer and internet access experienced by some students, they even could not manage to fulfill the amount.

The expected outcome of this study was to have the students analyze some language phenomena using the concordance software program and relate their data to relevant theories to come up with sound analysis. Intensive one-on-one discussions were held to help them with the software application, relevant theories, as well as the analysis. Some of the students' research reports can be seen in the following table:

Title	**Genre**	**Analysis of the word/words in contexts**

.

The word “love” in Shakespeare poems	Literature	The word ”love” in Shakespeare were found in expressing the love to God, male-female love, and love to nature.
The use of Word “Holy” in religious speeches	Speeches	The word “Holy” were found both in Christian and Islamic religious speeches with higher frequency found in Christian speeches.
The use of word “Blue” in Hinduism	Articles	The word “Blue” had significant meaning and representation in Hinduism which carried the meaning of purity and honesty.
The use of word :”Cool” in different contexts	Articles	The word “Cool” with the meaning of “alright” and “Okay” were more frequent than that of expressing certain degree of temperature.
The use of the word “Battle” in children stories	Literature	The use of the word “Battle” in children stories under investigation indicated that the word represented neutral meaning, without any indication of violence.
The use of the word “Black” in Obama Speeches	Speeches	The word “Black” in Barrack Obama political speeches indicated that the racial issues were central in the American politics. The word “Black” used by Obama showed more positive images of Afro-Americans
The meaning of the word “Build” in education contexts	Articles	In education contexts, the word “Build” was utilized to mean “develop”, “establish”, and “make by putting parts”.
The use of the	Articles	The word “female” outnumbered the word “Woman” in

.

words "Woman" and "Female" in Jakarta Posts		Jakarta Posts articles to show the roles of female in the contexts.
The study of the word "Hot" in different contexts	Articles	The word "Hot" was mostly associated with the contexts of politics, weather, and technology.

In their efforts to conduct the research, the students encountered some problems related to their limitation to the computer and internet access and the analysis ability due to insufficient literature review. This is in line with what McEnery and Wilson (2007) assert that the most important thing in using corpus software is the ability to interpret the results displayed by the program.

Other things that also contributed to the success of the Corpus Linguistic subject are the availability of the supporting facilities; for example sound and stable internet connection and the software (rather than the free-trial software from the internet).

Conclusion

Corpus Linguistics, with more affordable software – some are even free for 30 days trial on the internet – have provided easier ways to analyze the language phenomena. Even though it is not popular yet in Indonesian contexts, the subject accompanied with sufficient practices using the software could be a promising tool of analysis which can be used in education contexts for teaching and learning purposes. There are and will be more corpus-

.

based research in the future whose representativeness and validity are beyond doubt. In tertiary education contexts, corpus-based research should be encouraged as the new literacy and trends in writing their final research paper.

.

REFERENCES

Charles, Maggie. 2007. 'Reconciling Top-Down and Bottom-Up Approaches to Graduate Writing: Using Corpus to Teach Rhetorical Functions'. In *Journal of English for Academic Purposes*, Vol 6 No.4 p 289-302

Gay, L.R. et al. 2006. *Educational Research: Competencies for Analysis and Application*. 8th Ed. Pearson Education Ltd, Ohio.

Liu, Dilin and Ping Jiang. 2009. 'Using Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ASL Contexts'. In *Modern Language Journal*, Vol. 93 No.1 P 61-78

Malmkjer, Kirsten and James M. Anderson (Eds). 1991. *The Linguistics Encyclopedia*. Routledge, London.

Mc. Enery, Tony and Andrew Wilson. 2007. *Corpus Linguistics*. Edinburg University Press, Edindurg

Meyer, Charles F. 2004. *English Corpus Linguistics: An Introduction*. CUP, London

O'Sullivan, Ide. 2007. 'Enhancing a Process-Oriented Approach to Literacy and Language Learning: the Role of Corpus Consultation Literacy'. In *ReCALL*, Vol. 19 No. 3 p 269-286

Perez-Llantada, Carmen. 2009. 'Textual, Genre, and Social Features of Spoken Grammar: A Corpus-Based Approach'. In *Language Learning & Technology*, Vol. 13 No. 1 p 40-58