Research Paper Linguistics

How Using Corpus Linguistics Can Develop L2 Writing Skills

Hamzeh Moradi | Ph.D Research Scholar in Linguistics Panjab University, Chandigarh, India

ABSTRACT

Corpus linguistic as a relatively new field of language study has commenced its investigation as a filed complementary to the mainstream of artificial intelligence, general linguistics, computational linguistics and applied linguistics with direct engagement with the computer technology in the field of linguistic research. Corpus linguistics offers a satisfactory justification in the teaching of writing skills as it can easily depict the patterns of authentic language use by means of the analysis of actual language use. The main objective of the present study is to demonstrate the most common corpora and introduce inexpensive, effective and user-friendly software programs for the development of language skills, specifically writing skills. It presents some practical instances on how application of corpus linguistics can develop learners' writing skills. As an example, concordance makes it possible to see any word or phrase in context in order to notice what type of company it retains. Therefore, the learners, for instance, are able to see the differences between words which they generally mixed up.

KEYWORDS Corpus; database; language; generation; application; writing

1. Introduction

Over the decades the field of linguistics has progressed through a prolonged process of cognitive enterprise for creating conceptual links with various other branches of human science. It has recently taken a new approach to investigate how theories about different facets of human language can be attested with evidence of actual language use. The new direction of language research has added an additional aspect to the discipline associated with traditional linguistics. The introduction of computer and new technology has contributed to linguistics to progress and work out with the various techniques and tools to collect and investigate examples of actual language use from different dimensions of linguistic exercises, language skills and language output. The new approach to linguistics and language studies has provided adequate scopes for the immediate and direct use of linguistic information and evidences in typical activities and works associated with language technology and linguistics.

The advancement and progress of computer technology within the last decades has eventually introduced a new approach to linguistics. As a result of innovation in computer technology relatively new field, namely Computational Linguistics has developed as a significant area of Artificial Intelligence, which aims to look at language as a substantial instrument of human communication which is directly related to human cognition. Corpus linguistics is a significant area of computational linguistics which has an important role in the filed by providing a large amount of empirical language databases collected in a systematic approaches from different filed of actual language use.

Almost always there is a substantial linguistic and cognitive determination to communicate by means of language across space and time. There is also a technical enthusiasm to produce and develop an intelligent computer system to effectively be able to make linguistic interactions with human. With these determination and motivations, linguists and computer scientists, recently have come together at the point to develop new technology or systems like information extraction, language understanding and generation, machine translation, speech recognition and generation, computer-assisted language learning and teaching, and etc. which eventually benefits and contribute the advancement of the different facets of human life.

2. The Discussion

Corpus linguistics can be defined as the study of language based on the actual production of language use stored in corpuses computerized databases for linguistic investigations. Corpus linguistic is the study of language via computational analysis of a large quantity of collections of written data or texts and as well recordings of speech. It is the study of language through analysis of naturally produced language samples; usually special computer software programs are used for the analysis of data. Therefore, corpus linguistic is an approach to collect and analyze data both qualitatively and quantitatively rather than merely a theory of language. It can be applied to give justification towards language features and to test various hypotheses formulated in different areas of linguistic research. For instance, corpora related to levels of proficiency of language learners can offer useful information for second/foreign language acquisition research. With the help of historical corpuses, we can tack the development of particular characteristics in English history, such as emergence and use of verbs "wanna" and "gonna". The most significant benefit of corpus-linguistic approach is that linguistic researchers do not necessarily need to rely merely on native speakers' knowledge or their own intuition; Instead they are able to draw on a large quantity of authentic and natural produced language data obtained from various waters or speakers in order to evaluate, approve or reject their hypotheses about particular language features based on an empirical function.

There are various sorts of corpora which among them the most effective and influential are as follows: a) general corpora, like: Bank of English or British National Corpus, they include substantial variety of spoken and written, and as well different types of texts, produced by speakers with different social class and age. B) Synchronic corpora, like: Frown and F-LOB, record language data obtained at a particular point of time, for example, written American and British English of the early 1980s. c) Historical corpora, like: Helsinki Corpus Of English Text and A Representative Corpus of Historical English Register (ARCHER), include corpus texts from earlier periods. They normally extend and cover several decades or even centuries. D) learner corpora, like Cambridge Lerner corpus (CLC) and international corpus of learner English (ICLE), which demonstrate different regional varieties of language. There are as well some other significant varieties of specialized corpora, such as Michigan Corpus of Academic Spoken English (MI-

CASE).

Moreover, corpora can be analyzed with the help of concordancing programs which can significantly list rules and patterns of special registers. According to Greenbaum (1991), for the analysis of professional data or texts a language corpus of 20000 to 30000 is adequate. Nowadays, there exists such kind of programs, such as WordSmith Tools (Scott, 2012), which is an incorporated sets of programs to investigate how words or phrases behave in texts, as shown in Figure 1 below:

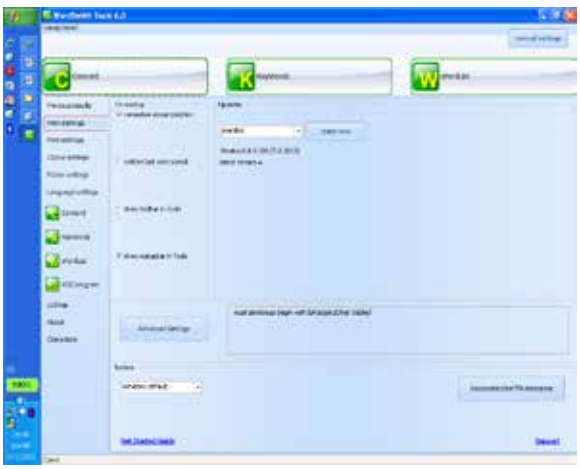


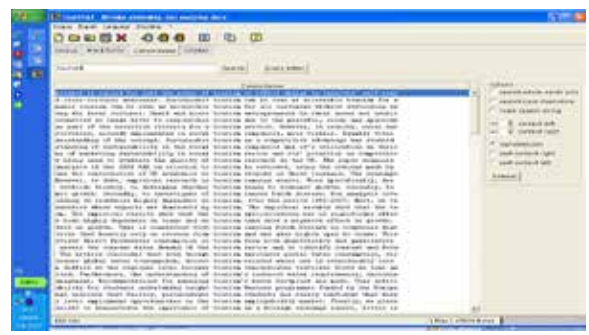
Fig. 1. WordSmith Tools

This particular type of program assists language learners to discover how words are employed in authentic texts and therefore how they can be utilized in their own texts. The wordList program can easily provide an index or list for all words or word clusters in a given text, they can be arranged in alphabetical or frequency order.

The concord or concordancer helps to observe any word, word cluster or phrase in the context, so that one can easily observe the type of company that it retains. This has several benefits for second language learners and teachers, since they can notice the differences between words which they may sometimes mixed up, words such as excited and exciting; they are able to see the patterns of different forms or structures; they can notice the connotation meanings which a word gets due to its relation and occurrence with other words, for example: the word "commit" usually carries negative and undesirable implications due to its frequent co-occurrence with a set word related to crime, fraud or murder; the can recognize the most suitable words or phrases to use, for example, whether to use proposition in or to with the words engage, specialize or interested; they are also able to look at the particular meanings which words can acquire on specific disciplines, for example: structure, sentence, floor or wall. As Wu (1992) asserts, only when words are, in their regular and frequently used environment, demonstrated in their regular forms and structure, demonstrated in their usual patterns of relation, they can effectively be learned, interpreted in a proper way and used correctly.

Finally, with the help of KeyWords, individuals are able to find the key words in a given text. Key-words are useful in providing a method to characterize a text. KeyWords makes a comparison between two pre-existing word lists, which have been created by WordList program. One of the word-lists is presumed to be long enough to be considered as a reference file. The other word-list is exactly on the basis of the text which learner is going to study. The purpose is to find out the words that characterize the text which learner is interested in, this text is automatically presumed to be shorter than the second chosen text.

Nevertheless, WordSmith is not free program and it must be paid for. Fortunately, still there are some other useful software programs for the analyzing of corpora, software programs such as TextSTAT, which is free of charge software and can be used freely, it is a very simple and user-friendly software which can be used for the analysis of texts. TextSTAT can read plain text files (in various encodings) and as well as HTML files (right away the internet) and it can also create word lists of word frequency and concordances from the given files (see Figure 2 below).



Conclusion

Based on the above mentioned discussion and theory, to boost and accelerate the development of writing skills there are some activities which should be considered and done by language teachers. The most substantial ones are as follows:

- Learners need to see real texts of everyday use, like repots, essays, an letters
- With the help of WordList, they are le to see the most common words or word clusters in writing, for example: word related to formal letters.
- Concordance can assist them to discover and observe the most common words and the collocation of words.

To sum up, corpus linguistics and specifically the software programs and application used for the analysis of corpora provide a worthwhile tool for second/foreign language teaching which not only the learners can benefit from, but the language teachers themselves will also be benefited from various significant advantages of software programs. With the help of corpus linguistics and new computer technology teachers can enhance and promote learners' language skills in various effective ways.

REFERENCES

Greenbaum, S. (1991).The development of the international corpus of English. London: Longman. | Scott, M. (2012). WordSmith Tools. Version 6.0. UK. | Hunston, S. (2002) Corpora in Applied Linguistics. Cambridge: Cambridge University Press. | Dash, N.S. (2005) Corpus Linguistics and Language Technology. New Delhi: Mittal Publications. | TextSTAT. (2012). Retrieved March 11, 2013, from http://neon.niederlandistik.fu-berlin.de/en/textstat/. | Wu. M. H. (1992). Towards a contextual lexico-grammar: an application of concordance analysis in EST teaching. RELC Journal, 23(2), 18-34. |