THE IMPLEMENTATION OF CORPUS-BASED TECHNIQUES TO ANALYZE LITERARY WORKS

Rakhmatova Mekhriniso Muhsinovna
Associate Professor, Ph.D.

Usmonov Amon Aminovich
A Master Student of English Literature Department

ABSTRACT:

Corpus linguistics is now known to expand our knowledge of a variety of phenomena in linguistics and literature. However, few studies have been undertaken to test its effects on literature studies. The field of study that analyzes literary texts by using corpus linguistic techniques is called corpus stylistics. It extracts literary meaning by using corpora. This paper sets out to explore three corpus-based techniques: *concordances, frequency lists, collocate lists* in the analysis of literary works. The basic question is whether these techniques is helpful to gain more objective and perceptive knowledge of literary texts. Seven English-translated stories of Utkir Hoshimov (*The butterfly and the tank, Disturbed dream, Why, Happiness, Solar scales, Three kinds of people, The last victim of the war*) are laid to corpus linguistic analysis. Due to the corpus-based techniques large amounts of literary data are analyzable. Otherwise, the data is not more than several lines of poetry or short extracts of narrative.

Keywords: Corpus linguistics, literature, corpus stylistics, corpus-based techniques, concordances, frequency lists, collocate lists

INTRODUCTION:

Corpus linguistics is currently enlarging our understanding of several phenomena in linguistics and literature. By way of computerized data, large numbers of words are analyzed with efficiency and accuracy, Corpus-based methods have been adopted both directly and indirectly in language teaching, data-driven learning and also in lexicography [Ridwan Wahid, 2011]. Along with them, the use of corpus linguistics has also been extended to literature. Literary analysis, either quantitative or qualitative, can result in a very detailed stylistic descriptions of literary texts that can either complement conventional interpretations or enable insights not possible through intuition alone [Adolphs, Svenja, 2006].

The analytic potential of certain corpus techniques might sound great in gaining some insightful knowledge that can be used in generating a sort of rigorous understanding of literary meanings and the way language is organized in a literary text [Khalid Shakir Hussein, 2015]. Using corpora, huge quantities of literary language is completely reachable and open to systematic analysis and detailed description [Khalid Shakir Hussein, 2015]. However, Intuition does not work under the unprecedented huge amounts of corpora.

MAIN PART:
4.1 Concordances:

The term *concordance* is encountered by almost everyone who works on any area in corpus linguistics. Concordance analysis is probably the most known corpus analytic tool. There are a lot of corpus linguists who tried to

offer definitions to what a concordance refers to. According to John Sinclair [1991]:

A concordance is a collection of the occurrences of a word-form, each in its own textual environment. In its simplest form it is an index. Each word-form is indexed and a reference is given to the place of the occurrence in a text.

Word-form or *lemma* is "a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling" [Francis and Kucera, 1982] and can be looked for independently.

There are various concordance programs that are also referred to as KWIC (Key Words In Context) [Scott, 2010], because they all referred to a list of all the occurrences of a word-form in a particular corpus, along with its context in which it usually occurs to the left and right of a search word.

Table(1) below shows a concordance in KWIC format. In this example the search word *happiness*, as it occurs in Utkir Hoshimov's stories, is presented at the center of a fixed context of words and characters.

Concordance programs are highly effective corpus analysis tools that enables researchers and scientists to bring the examples of a particular word or character present in the original context.

4.2 Frequency lists:

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, 2007]. In a word frequency analysis, we are given data about the number of times each word is used in the content. Table(2) below shows the top 15 words in Utkir Hoshimov's selected stories.

1	1204	the
2	506	a
3	495	and
4	463	he
5	431	to
6	365	in
7	355	was
8	303	of
9	292	his
10	266	i
11	253	it
12	253	she
13	234	s
14	215	you
15	207	her

However, this list is not very informative for linguists, because it contains a large number of function words (e.g. "the," "a," "to"). If these words are removed and only the content words remain, a different picture begins to emerge, which is shown in the following table(3).

1	said	155
2	Nasiba	67
3	Munira	47
4	eyes	45
5	Ismail	44
6	mother	44

From this list one can identify some of the main characters in the story and many other words that are probably used to describe them.

4.3 Collocate Lists:

Collocates constitute the words that surround a particular search word [Scott, 2010]. The phenomenon of Collocation, as described by Firth [1957:14], takes into account the very fact certain words tend to occur in combination with each other within certain linguistic contexts. Therefore, a

collocate is most definitely a word that exists in the surrounding environment of another word [Baker, 2006:37].

Collocate lists are closely related to concordances. The latter produces the actual occurrence of a search word accompanied with its textual environment, so what is displayed is the search word within its whatever context [Khalid Shakir Hussain, 2015]. The emphasis on the collocate list, however, is not the search word but the company-words as distributed around [Scott, 2010: 121-22]. For instance, the software *AntConc* provides researchers with an independent window within which collocating occurrences and their frequencies can be set in columns and rows.

Collocational analysis is helpful for various reasons. Pezik [2011: 456] captures a basic reason:

First of all, no description of language can be complete unless it does some justice to its phraseology. Language is highly idiomatic and lexicality (of which collocation are the most important aspect) has become a level of linguistic analysis... Collocations have also been found to be a revealing source of information about discourse-specific metaphors... Studying collocations is also crucial in identifying selectional restrictions and semantic prosodies, the latter of which can be defined as the *attitudinal load* of certain lexical items.

CONCLUSION:

The use of corpus-based techniques in literary analysis has a number of benefits. Such techniques prove us to not fully satisfied with the intuitive critical interpretations of a literary text. By virtue of such techniques, the linguists are provided with neutral and impartial insights into the literary texts under investigation. Moreover, because of the user-friendly corpus software, it is much easier to

bring out some invisible meanings that could be missed or unrecognized.

These three techniques tackled above have their own role in investigating hidden meanings of literary works, Linguists can bring some examples of a particular word or phrase by the use of concordance programs, check the frequency of a word by word frequency analysis tool, and finally search a word that occur in combination with other words within certain linguistic contexts.

REFERENCES:

1. Adolphs, Svenja (2006). Introducing Electronic Text Analysis. London and New York: Routledge.
2. Baker, P., Hardie, A., & McEnery, T. (2006). A glossary of corpus linguistics. Edinburgh: Edinburgh University Press.
3. Firth, J. R. (1957), "Linguistic Analysis as a Study of Meaning", in Palmer, F.R. (ed.), Selected papers of J.R. Firth, London and Harlow, Longmans, Green and Co. Ltd: 12 – 26.
4. Francis, W. & Kucera, H. (1982). Frequency analysis of English Usage: Lexicon and grammar. Boston: Houghton Mifflin.
5. Khalid Shakir Hussein (2015). The potentialities of corpus-based techniques for analyzing literature. Thi-Qar University, Collage of Education, English Department, Iraq.
6. O'Keefe, Anne, Michail Mccarthy, and Ronald Carter. (2007). Fron corpus to classroom. Cambridge: Cambridge Up.
7. Pezik, P. (2013). Comoutational ans corpus linguistics. Retrieved from http://www.pezik.pl/wp-content/uploads/2011/07/new-ways.pdf
8. Ridwan Wahid (2011). The use of corpus-based techniques in literary analysis: Exploring Learners' Perceptions. International Islamic University Malaysia.

9. Scott, M. (2010). Wordsmith tools (Version 5.0). [Computer software]. Liverpool: Lexical Analysis Software.
10. Sinclair, John. (1991). Corpus, Concordance, Collocation. Oxford: Oxford University Press.