

A Corpus-based Analysis of Lexical Cohesion in Charles Dickens's *Great Expectations*

Muhammad Ajmal

Associate Professor, Department of Linguistics & Literature, Qurtuba University of Science and Information Technology, Dera Ismail Khan, KP, Pakistan.

Samina Rana

Associate Professor of English, Higher Education Department, Punjab, Pakistan.

Safia Siddiqui

Assistant Professor, Department of English (Literature), The Islamia University of Bahawalpur, Punjab, Pakistan.

Email: safia.siddiqui@iub.edu.pk

(Corresponding Author)

- p-ISSN: 2663-3299
- e-ISSN: 2663-3841
- L-ISSN: 2663-3299

Abstract: This study endeavours to dissect the lexical items in Charles Dickens's *Great Expectations*. The researchers applied corpus-based analysis to get the statistical results and the software which has been used as a tool in this study is AntConc 3.5.8. Three main features have been investigated in detail which are collocations, function words and their frequency of these. This statistical study will help the literary critic, further statistical results about the style of cohesive devices and it will also lead the students to consider the text from a more syntactical perspective. A quantitative analysis follows a qualitative one as the corpus-based results are shown in a statistical way and then interpreted qualitatively.

Key Words: Corpus Analysis, Cohesive Devices, *Great Expectations*, Textual Meanings

Introduction

Charles Dickens is one of the eminent and big names in English and colonial literature whose literary writing proved mind-blowing effects on the minds of that era through the reflection of the society. Corpora have brought a revolutionary change in English applied linguistics research over the last few decades. Text and discourse analysis both are analyzed through the application of corpora. By using corpora, text and discourse analysts collect the data on a larger scale and can generalize their results with authentic figures on a bigger scale as compared to those

researches in which the corpus technique is not used. Corpus-based analysis helps linguistic researchers to explore the relationships among the word to create meanings on the basis of their association at a different level in text and discourse (Sinclair, 2004). Conversation and academic writing are two different things and these texts can be proved through corpora on the basis of their frequency of phrases (Biber and Conrad, 1999). Now in critical discourse analysis, the relationship is explored among three aspects

discourse, power and ideology and this analytical practice is done through corpora. British newspaper corpus reported on behalf of their analysis that women are mostly represented in terms of their physical appearance like the adjectives describing women as "beautiful", "pretty" and "lovely" which are used as collocations whereas on the other hand adjectives referred to men collocates as "great", "key", and "main".

Adjectives with their different association showed the patriarchal ideological beliefs in British society. Corpus software has been utilized in examining the text. The primary goal of this study is to interpret literary work through the language in a text. The researcher selects only one text from Dickens's literary work and uses the corpus-based method to the analysis of Charles Dickens's *Great Expectations*.

Objectives of the Research

The main objectives of the researcher are,

- 1) To identify the lexical patterns in Dickens's *Great Expectations*
- 2) To consider the textual meanings of those patterns, suggest in this literary text

Questions of the Research

- 1) What lexical patterns are characteristic of Charles Dickens's novel *Great Expectations*?
- 2) What textual meanings do those patterns suggest in this literary text?

Significance of the Study

The current study can help the learner in learning how to study the word pattern in texts by using corpora. It may also inspire the learner to promote an interest in corpus-based analysis. This study will also motivate the researchers to explore a new way of studying through this medium.

Review of Literature

In Linguistics, the word Corpus can be defined as a body or the structure of language on a specific topic. The word Corpus is used to describe a group of language units that can be in form of text and speech as well. Corpus Linguistics is the new linguistics methodology that investigates naturally occurring language on the bases of computerized corpora. The analysis of the text is obtained through the help of a computer, with software like AntConc, COCA, BNC and Wordsmith and then the quantitative data is investigated. Corpus Linguistics analysis is always based on the examination of some kind of frequencies. Stubbs and Halle (2012, p1) define Corpus Linguistics as it is a computer-assisted method to deal with the larger quantity of linguistics analysis in a text of novels. Sinclair (1997) views Corpus Linguistics as the study of language structures of authentic language in use.

No doubt, the studies of patterning and repetition performed a great role in textual formation (Jakobson, 1960), but the best work was done in English cohesion by Halliday and Hasan in their influential research (1976). According to them, cohesive devices perform a very strong bond in making the textual and contextual meaning of a text and discourse. For this purpose, both types of cohesive categories: grammatical and lexical cohesion play a very significant role to fill the gaps.

References in cohesion are those items which are referred towards those objects that are already mentioned before and after in the text or discourse to interpret the related meaning of those objects and to avoid repetition. The names given to these references are personal references i.e. Ali, you, she, the pen, it), comparative references i.e. better, stronger) and demonstrative (i.e. this, that, here, there). Meanings can be interpreted in two ways, one is an internal

reference in which the reference exists in the text and the second one is out of the text. In other words, the internal reference is called the endophoric reference and the exophoric reference is that in which the reference is beyond the text (Halliday & Hasan, 1976). It is used as a relationship between the facts and the things. References may create a relationship at a huge distance in a text. As defined by Crystal (2008) anaphora is a tool for defining a linguistic item depending on some earlier expressed item which is referred to it as the antecedent. Deixis is defined as it is the name of determiners, personal pronouns, possessive pronouns, and interrogative pronouns (Bussman, 1996).

Substitution & Ellipsis

These two discourse markers are associated with words instead of meanings. Substitution is defined in such a way that some words can be replaced by other words while in the process of ellipsis, the word is omitted instead of using other words. In other words, the linguistic items have zero replacement in text and discourse. These two categories have the same mechanism in cohesion but the usage of ellipsis in text or discourse has a little bit confusing. References are used to infer the meaning referred to the referent while substitution and ellipsis are related to words and these two linguistic items have different characteristics in creating the text and in the case of interpretation. Usually, substitution is an association that finds inside the text or discourse. It is a type of technique used to avoid repetition

Conjunction

Conjunctions also called connectors and a very compulsory element to use or create the discourse or text. This cohesive device plays a very important role in making a strong bond in text. It beautifies the text and helps to understand the meaning of the text and discourse. This discourse marker has both

categories of grammatical and lexical aspects but most of his use is taken as a grammatical one. This category has a different characteristic from all the other categories. Conjunctions are further divided into four subcategories; temporal (time relation), causal (Cause and effect relation), adversative (unexpected relation) and additive (adding relation).

Lexical Cohesion

Lexical cohesion is defined as it is the term which is used to explain the meanings of words to make relations in text. Lexical cohesion is the name of the mechanism in which word choice is the main function to create text. Synonyms are the prominent ones in this regard. Collocation has the central position in arranging the words in a proper form to create a meaningful text with the appropriateness from all aspects (Halliday & Matthiessen, 2004). Lexical cohesion is described as a pattern of attaining cohesion by repeating the same word or phrase or using chains of synonymous words that come up toward the continuity of lexical meaning (Baker & Ellece, 2011). It can be explained in the easiest shape, Lexical cohesion takes place where the same word comes again and again and the same referent in both places. Synonyms are words that have different forms morphologically and closely related to each other in meanings and cannot be always replaced in discourse or text. They are not exactly the same and the items which are used on the bases of their appropriateness in sentences and their synonyms cannot be applied (Yule, 1996). Meronymy is a constituent part of the main class of something like a finger is the constituent part of the hand. Man, woman, boy and girl are the subordinates of human beings. (Ellece & Barker, 2011). Content words are also part of lexical cohesion; these are the main words which have meaning on their own without referring to other words in the text. These content words are also called open-class

words. Function words are those words which do not have meanings on their own, they just help to complete the sentences like articles, prepositions, auxiliaries and modal verbs. These words are also called closed-class words. They cannot change their form. In short, all the lexical devices are compulsory for the makeup of the text and any type of discourse.

Research Methodology & Theoretical Framework

The researchers used the corpus-based method to recognize the way to use lexical cohesion in the text of Charles Dickson's writing and to consider the textual pattern of meaning in his literary text. This study was conducted on the bases of both quantitative and qualitative methods to get the results statistically and in explanatory forms. Antconc software as a tool was used to study the text on corpus-based analysis. Charles Dickson's *Great expectations* were taken as machine readable data to check and explore in-depth knowledge about this study. The focus was given particularly on function words and collocations to lemmatize the study.

Sampling

The researchers have picked up only two lexical items to delimit this study. The first one is a preposition and the other one is collocation. Furthermore, regarding collocation, we have selected only three adjectives as collocation (little, poor and great).

Data Collection & Analysis

Data was collected by taking screenshots through AntConc computer software of functional words like articles prepositions and collocational words corresponding to the research objectives. Respective data were analyzed statistically with their frequency which has been illustrated with tables

separately in numeric form and theoretically as well also for explicit results.

In this study, the researchers utilize a corpus-based approach to focus on his functional and collocative words and the frequency of these words in his novel. Moreover, he will try to consider this novel which has been used as a model to analyze the style of language by Charles Dickens. This well-known piece of literature is an appropriate and challenging task to analyze. Why the researcher is going to choose this one because it has not been explored by many corpus linguists. The researcher will do to dig up an in-depth analysis to enhance the informative knowledge in his study. The researcher will go ahead with Corpus-based analytical results and description. Then he will focus on the specific words delimitating them in the text for analysis. Furthermore, he will present the collocative association in the text and how these relational words create textual meaning in the author's piece of literature.

Machine-readable data in the form of a collection of millions of words are given to a computer in corpus analysis. For instance, the statistical analysis of different authors' works or a single author can be compared with texts by another class of authors which differentiates the style of the author, but it has been done by the stylisticians in previous years. The aim of corpora with the collection of millions of words in the form of texts regarding a particular author's work like Charles Dickens or different authors of the 19th century is to describe the discourse. The most important characteristics are examined through the framework of corpus regarding language: first, the language is seen from a social perspective with its form and meaning, and the second is the lexical representation of language. The main concern of corpus linguistics is to study the repeated words and their typical use in language, which words have an association with others and where they occur in the text. As a result, corpus

linguistics depends on the proof of language usage in the form of collected and examined data in corpora.

The first step to examining corpora is corpus stylistics and this is the expansion of corpora among linguists. The second prime aspect is to look theoretical framework of the corpus in which the meaning and form are studied on behalf of its association. In this linguistic feature, the focus is on the deviant or flexible grammar with local classifications of representation and it raises from linguistic standards that take to the formation of aesthetic outcomes. The third probability of the machine-readable technique is the corpus annotation in which a specific linguistic trait is by proceeding a corpus and organizing a complete and comprehensive analysis of the characteristic as it happens in this corpus. In general, "many corpus linguists are actively engaged in issues of language theory, and many generative grammarians have shown an increasing concern for the data upon which their theories are based, even though data collection remains at best a marginal concern in modern generative theory." In defining descriptive adequacy, abstract grammar plays a very important role in making sentences well-formed. Explanatory adequacy comes at the top after getting the descriptive properties and after applying the abstract principles to understand the meaning beyond language and then it will be called universal grammar.

The job of corpus linguists is to check the complications and variations that are found in language and in their debates they give priority to descriptive adequacy instead of explanatory adequacy against the job of a generative grammarian. In a sense, corpora are the best mediums in their performance to check linguistic hypotheses regarding their falsifiability, completeness, simplicity, strength, and objectivity. It means corpora do have not such contribution to generative grammar. As it is mentioned that generative

grammar goes on to analyze the textual function of language whereas in corpora both sides of the language, descriptive and communicative tools of language are studied.

A Pattern of Function words

Every literary writer has his own style of writing whether it is a linguistic pattern or contextual pattern and this thing makes the writer unique from others' work. A few decades earlier, when corpora were not introduced, analysts chose stylistics to study the style of writing but stylistics only showed the deviated language in literature. On the other hand, the corpus-based technique has brought revolutionary steps in the history of linguistic analysis not only in literary writing but also in other subjects like discourse analysis, research and etc. The researcher has also tried to get absolute results in analyzing the usage of linguistic items in Dickens's work through a corpus-based approach. Two tables have been drawn to illustrate the collected data with their highest frequencies. In the first table, the list of prepositional words has been shown and the second table consists of the list of collocational words. This is the first step in this research and then the second step is to describe the textual meanings of this pattern. The textual meaning pattern process is used to investigate the style of writing.

The researcher sent the text to AntConc software after transferring it from PDF to Word then he studied first about the word list of the text. It is clear visibility in the above screenshot that how many total numbers of words and how many types of words are in this novel. There are 189736 words in this text which can be seen on this screen in the form of "word token" and very close to the word token, there are another thing "word types" so, this text has 10731 types of words. Through this software, the researcher got the absolute results of this text. It is a very interesting and satisfactory medium to study the text in a way for the researcher that the results are in

the statistical or numeric form. This one was the general information about the text through this corpus-based study. Then researcher went for the in-depth detailed study regarding his own area of interest which already he mentioned in the form of objectives

Firstly, the researcher checked prepositions in this text and wrote the prepositional word "to" in this software to identify how many other words it comes to make the relations and give the meanings. He also investigates the frequency of prepositional words. After analyzing, the researcher came to know that the prepositional word "to" has the highest frequency 5152 times in the text used by Charles Dickens. He uses this word with different other words like articles, verbs and adjectives etc. The author has used "to" with the article "a" most of the time, with main verbs, with determiners in the form of adjectives. The relationship of "to" with the word "be" in the form of "to be". So, this is the first statistical result which has been taken through Antconc about the prepositional word.

The prepositional word "though" has been shown with his relation with other words. This word comes with determiners, possessive pronouns with pre-determiner pronouns and sometimes with adjectives in the form of determiners. This word occurs 126 times with the lowest frequency in this text used by the author. Other prepositional words have been shown with their frequencies accordingly in table no. 01 below.

Table 1.

Rank	Frequency	Prepositional words
01	5152	To
02	4436	Of
03	3029	In
04	1759	With
05	1637	At

06	1421	On
07	1390	For
08	797	By
09	635	Up
10	580	From
11	492	Into
12	370	Down
13	366	Upon
14	309	Before
15	293	After
16	263	Over
17	186	Off
18	157	Without
19	126	Through

The illustrated prepositions with their positions and the relationships with other words show how these prepositions have been used in a different style to create the textual meanings in an effective way to inspire and entertain the reader.

Collocations of Little, Poor and Great

In the next step, the researcher analyzes collocational words in this text. Here are some words with their association in the form of collocation. Dicken uses peculiar collocations in his writing by making a combination of different parts of speech. The researcher selected the three adjectives as collocational words from the text Great Expectations. The repetition of these words with high and low frequencies shows feelings of Dickens. Frequencies of these words little, great and poor describe the particular style in making a new way of a literary text with a peculiar compound of words. The below screenshot of Antconc 3.5.8-based corpus analysis of these collocations will help to give the statistical results of collocations regarding these adjectives. The frequencies and ranks of these collocational words are in the table below

Table 2.

Rank	Frequency	Collocational words
1	371	Little
2	199	Great
3	77	Poor

After getting the results through this procedure, the researcher met with the sensible approach of Charles Dicken of creating associations by using an adjective with a noun in a different style. This collocational association of words compared with the style of other writers then it will be a totally different style of collocation with textual meanings. Basically, authors use this tool to create their text meaningfully to convey their philosophical thoughts or their hidden themes. Corpora are unable to give the textual meanings, its job is to show the frequency of the words and tell us the collocation of the words which the writer built up by using the acceptable channels of the rule of language. Charles Dickens' work also was also taken to explore how he also creates a different style of writing to make the meanings. Now, it is the approach of the analysts of a researcher to understand this association and infer the textual meanings of the text.

Conclusions

The primary purpose of the respective research was to discover the style of using lexical items deployed by Dickens. The research highlights significant results that corpus-based analysis through Antconc 3.5.8 computer software proved very fruitful to comprehend the text from different perspectives. It can help to explore new horizons for the research to investigate the more in-depth association and relation among the various words and through this channel, readers and language analysts can understand the textual meanings of the text. Lexical items effectively play a very vital role in academic writing because these items make the strong bond in the text to make it cohesive. The major objective of lexical cohesion is to make capable readers understand the rationale-based connections in the text. The facts of this study revealed that the rule helps researchers who are interested in improving their academic writing and want to enhance their knowledge to comprehend different discourses. More research can be adopted to compare the frequency and position of lexical items in literary texts. To conclude, it is believed that corpus-based analysis of the texts on large scale may help the learners to gain more detailed information regarding lexical cohesion and their usage in different styles of writing.

References

- Baker, P. & Ellece, S. (2011). *Key Terms in Discourse Analysis*. London & New York, NY: Continuum International.
- Biber, D., & Susan, C. (1999). *Lexical bundles in conversation and academic prose*. In *Out of corpora*, edited by H. Hasselgard, & S Oksefjell, 181-190. Amsterdam-Atlanta GA: Rodopi.
- Crystal, D. (2008). *A Dictionary of Phonetics and Linguistics*. London: Blackwell.
- Halliday, M.A.K. (1994). *An Introduction to Functional Grammar (2nd edition)*. London: Edward Arnold.
- Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. & Matthiessen, C. (2004). *An Introduction to Functional Grammar (3rd. ed.)*. London: Hodder Arnold
- Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stubbs, M., & Halbe, D. (2012). *Corpus Linguistics: Overview*. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*, 1377-1379. Oxford: Blackwell.