

An Investigation of Hedges and Boosters in Pakistani Opinion Articles: A Corpus-based Study

Syeda Fizza Batool

MPhil Scholar, Air University, Islamabad

Haleema Majeed

MPhil Scholar, Air University, Islamabad

Tehseen Zahra

Assistant Professor, Air University, Islamabad

Abstract

Hedges and boosters are very important in writing as these expressions indicate the writers' standpoint. Previous studies (Abbas & Shehzad, 2018; Zafar, 2018) show that a little work has been done in Pakistan on hedges and boosters; especially in newspaper opinion articles. The current study aims at investigating the most frequently occurring hedges and boosters in Pakistani newspaper opinion articles and to analyze writers' stance of using hedges and boosters in writing opinion articles. The data was collected from 5 Pakistani English newspapers; The Dawn, The Pakistan Observer, The Express Tribune, The Nation and The Daily Times. 50 opinion articles were randomly taken from these newspapers and comprises a corpus of 47927 words. The articles chosen were written by different authors and these were published in November, 2018. *MetaPak* (Abbas, Shehzad, & Ghalib, 2017) software was used to extract the hedges and boosters from the corpus. The frequencies of these hedges and boosters were represented through graphs. The researchers analyzed and discussed these hedges and boosters in order to analyze writers' standpoints. The software showed the total frequency of *would*, *could* and *should* to be 143 times, 60 times and 61 times respectively in the corpus. After the manual analysis of the collocations of these words, it was revealed that the occurrence of *could* and *would* as a hedge is 58 each and of *should* as a booster is also 58. The findings of the study showed that most frequently occurring hedge markers were *would* and *could* and the most frequently occurring booster was *should*. After the analysis of the data, it was found that the use of hedge marker by the writer showed his stance of being cautious and careful in his writings. The writers, in many cases, do not want to promise themselves to the statements; rather they are more uncertain in their words and opinions they make. This finding showed that the Pakistani writers' do not speak explicitly but keep themselves at a distance in projecting their stance. On the other hand, the findings concluded that in the use of boosters there is certainty in writers' voice. It also showed that the writers' is voice seemed confident and assertive. The study reveals that the employment of boosters in writings is because the writer wants his readers to be convinced about their opinions and suggestions.

Keywords: hedges, boosters, opinion articles, METAPAK, writer's stance

1. Introduction

An opinion article is a discourse composed by somebody who is autonomous of the newspaper ("More information about opinion articles," 2005). An opinion article is a piece of writing, published in a newspaper or magazine that basically mirrors the writer's standpoint about the subject. The opinion articles are contributed by individuals from outside the paper

and speak to their own perspectives. Opinion articles are public discourse that communicate with a mass group of onlookers and assume a vital job in the deciding and moving of public opinion. Consequently, to have the capacity to viably express their thoughts, writers take care in the utilization of their procedures in persuading the groups of onlookers. Authors of opinion articles utilize distinctive hedges and boosters so as to demonstrate their position and stance about the idea. Hedges indicate the reluctance of the author to exhibit propositional data genuinely and unquestionably (Hyland, 2004). Hedges can also be used as a way to avoid from taking full responsibility for the statement. Boosters are considered as highlights that express the author's solid certainty for a claim (Holmes, 1982) and confirmation and attestation of a proposition unquestionably (Abdi, Rizi, & Tavakoli, 2010). Boosters are expressions which use adverbs to boost different words and intensifiers, which enhance a term. To hedge intends to waffle on an issue, to abstain from submitting oneself. Boosters exhibit certainty of the author about his viewpoint. According to Mary M. Talbot (2010), hedging and boosting devices are modal elements; element that modify the force of a statement, either weakening it or intensifying. These components change the power of an announcement, either debilitating it or escalating. Hedges and boosters influence the tone and style of the composition. Hedges and boosters are basic features for authors to clear up their epistemic position and position identified with the writer–reader communication (Cao & Hu, 2011). So, it can be said that hedges and boosters are important stance makers that reveal standpoints of the writer on certain content.

The current study is an investigation of hedges and boosters in Pakistani opinion articles in the newspapers using corpus analysis tool. A corpus is a collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as a starting-point of linguistic description or as a means of verifying hypotheses about a language (Crystal, 1991). Corpus linguistics deals with the principles and practice of using corpora in language study. A computer corpus is a large body of machine-readable texts (Crystal, 1992). Corpus linguistics is 'the study of language based on examples of real-life language use' (McEnery & Wilson, 1996). So, Corpus analysis is technique to analyze a corpus which is electronically stored 'real-life' language sample. Corpus linguistics includes the utilization of Computers to quickly search and analyze databases of real language. These databases are called corpora (the plural of Latin corpus) and they can involve any principled accumulation of composed or transcribed spoken language. Instances of well-known corpora are the British National Corpus (BNC) and the American National Corpus.

Many studies (Abbas & Shehzad, 2018; Zafar, 2018) have been conducted on the use of hedges and boosters in different genres. But there is no significant work done on the use of hedges and boosters in opinion articles by the writers and this specific research is also not yet conducted in the Pakistani context. The current study investigates the use of hedges and boosters in Pakistani newspapers opinion articles. It also examines the writers' stance of using hedges and boosters. This study will provide an opportunity to the academic discourse learners to enhance and broaden the understanding of use of hedges and boosters and how these makers can be used to express any stance more clearly and deeply.

The current study is significant in a number of ways. It will help the novice authors in writing opinion articles more effectively by using the interpersonal features. It will also assist them on the use hedges and boosters and in showing their stance of how to lessen the authority of your voice and how to make your voice more confident and convince the reader. As the use of hedges and boosters culturally varies and also is conventional to particular area and genre. This study will be helpful to know how the Pakistani writers use these features to

express their stance while writing and what are the conventions being followed in the context. The current study will also be very helpful for the academic discourse world as it will give them a direction of using hedges and boosters while giving any sort of opinion. It will also be supportive for the academic discourse world as it is a corpus-based study and will guide the novice researcher about the corpus and analyzing it. Furthermore, it will also give how to assert your voice when a novice writer is writing an opinion article.

2. Literature Review

Opinion articles are the part of professional genre used by the professional and expert writers in the respective field. The expert writers use hedges and boosters to show their stance in writing. Lakoff (1973) was the first to introduce the term hedge. He defined it as ‘words whose job it is to make things more or less fuzzy’ (p. 195). Hedges and boosters can reflect different aspects of writing and delivering the message. Hedges can lower the stance of the writer and booster makes the stance firmer and stronger. Stance is said to be the personal voice of the author which is socially and culturally influenced (Hyland, 2005).

Hyland (2005) came up with a model known as Hyland’s taxonomy of meta-discourse. This model analyzes two types of categories that are textual features and interpersonal features. Textual features include logical connections, frame markers, endophoric markers, evidential and code glosses whereas interpersonal features include hedges, boosters, attitude markers, self-mentions and engagement-markers. Interpersonal features according to the model involve the reader in the text. This model is very useful in analyzing the interpersonal features of any text or genre.

Abbas, Shehzad, and Ghalib (2017) carried out a corpus-based research about hedges and boosters in the final investigation report of joint investigation team of the panama case. He analyzed hedges and boosters from the perspective of metadiscoursal markers. By applying Hyland’s model and using the *MetaPak* tool he finds out that the use of boosters was more prominent than the use of hedges. The total boosters employed in the selected corpus of JIT were 47 with a frequency of occurrence 360 times and the hedges used were 53 with and occurrence frequency of 327 times. Although the hedges used were greater in number, but the frequency of boosters employed in the corpus were greater in number.

Many studies have been conducted to show the use of hedges and boosters and their features. They can be studied and analyzed in different instances. Zafar (2018) studied the metadiscursive features of hedges and boosters in research article abstracts of English and Education. She found that the most common occurring boosters in the abstracts of both the disciplines are *found*, *showed*, and *demonstrated* in a chronological order. She also found that modal auxiliaries like *may* and *could* use as hedges have the highest frequency found in the articles.

Trajkova (2011) investigates the functional and linguistic form of hedges. She took English and Macedonian newspaper for the analysis. She also studies how hedges can be used in attaining and constructing the persuasion. She kept in view the cultural differences that could be employed in the use of hedges in any writing. She came up with the findings that the most frequently used hedges in editorials are the modal verbs. This study is also very important as it propose some indications regarding cultural differences of different areas and writers in a way a writer writes, and reader perceives a written text of a specific genre. This study tells that as the use of interpersonal features like hedges and boosters are based on the

cross-cultural differences of writing. So, it can be said that the use and application of these textual features are also conventional keeping in view the difference and contrast of the culture and territory.

The use of hedges and boosters in English academic articles contrasts according to the discipline in which they are used. The choices that writers make seem to be constrained by the discourse characteristics and rhetorical styles of each discipline and reflect and represent the nature of different disciplinary individualities (Takimoto, 2015).

Takimoto (2015) further said in his research that as the discipline of humanities and social sciences is more subjective, the use of hedges and boosters in these disciplinary articles is more than that of the natural sciences. Natural sciences tend to be more objective and fact based that requires objectivity so the use of hedges and boosters in such articles is less than that of the other disciplinary articles. This research shows that hedges and boosters and their use also depend on the discipline and the individual characteristics and properties a discipline possess. Same is the case with the genres. As every genre holds a different set of characteristics, similarly, the writer uses different ways to show their stance to project their voice.

Aib and Zahra (2017) wrote an article about the attitude markers in the Pakistani English book reviews. It was a corpus-based study that analysed 100 book reviews which make 76000 words in total. This study was about the attitude markers like the personal mention 'I' and other modes of attitudes that presents writers' personal point of view. They found that Pakistani English book reviews employ attitude in two ways. The first is they assert their stance by using the personal pronoun I. Secondly; they also employed other modes in the book reviews. The study mentioned these as the 'six categories' in their research. Based on these results they concluded that writing book reviews is a two way process. The current study aims to answer the following research questions:

1. What are the most frequently occurring hedges and boosters employed in Pakistani English newspaper opinion articles?
2. How do the hedges and boosters show the stance of the writers in Pakistani English newspaper opinion articles?

3. Research Methodology

The data for the current study was collected from the popular Pakistani English newspapers. Five leading Pakistani English newspapers which were selected for the extraction of data are, *The Dawn*, *The Pakistan Observer*, *The Express Tribune*, *The Nation* and *The Daily Times*. After selecting the newspapers, 50 opinion articles which were published in November, 2018 and written by different authors were taken. Then with the help of METAPAK Software the hedges and boosters were separated. The data was then manually cleaned by the researcher. Hedges and boosters were cleaned by looking on different collocates and eliminated those collocates where the selected word was not functioned as a hedge or a booster. The frequencies of these hedges and boosters were then plotted on excel sheet and table of frequencies were made. Then these hedges and boosters were analyzed in order to understand writer's standpoints. The rationale for this research is to find out the different hedges and boosters in Pakistani newspaper opinion articles and to analyze writer's stance of using hedges and boosters in writing opinion article.

3.1. Sampling Technique and Size

The sample for the current research was selected through convenient sampling technique. The data was collected randomly from the available opinion articles of newspapers. This provided the basic information quickly and efficiently. The researcher has selected sample size of 50 opinion articles from Pakistani English newspapers to build the corpus and to have a better representation and generalization of the results obtained. The corpus selected for the current study comprises the size of 47927 words. The topics of opinion articles were random, some were related to current affairs, and some were related to history and so on.

3.2. Data Collection: Research Tool

Qualitative data in the form of newspaper opinion articles were used. As the current study is going to explore the use of hedges and boosters in Pakistani English newspaper opinion articles, so, exploratory research design is followed. *MetaPak* software was used as a research tool (Abbas, Shehzad, & Ghalib, 2017). This research tool is very helpful in extracting and analyzing the values and frequencies of the *textual* and *interpersonal* features as proposed by Hyland (2005).

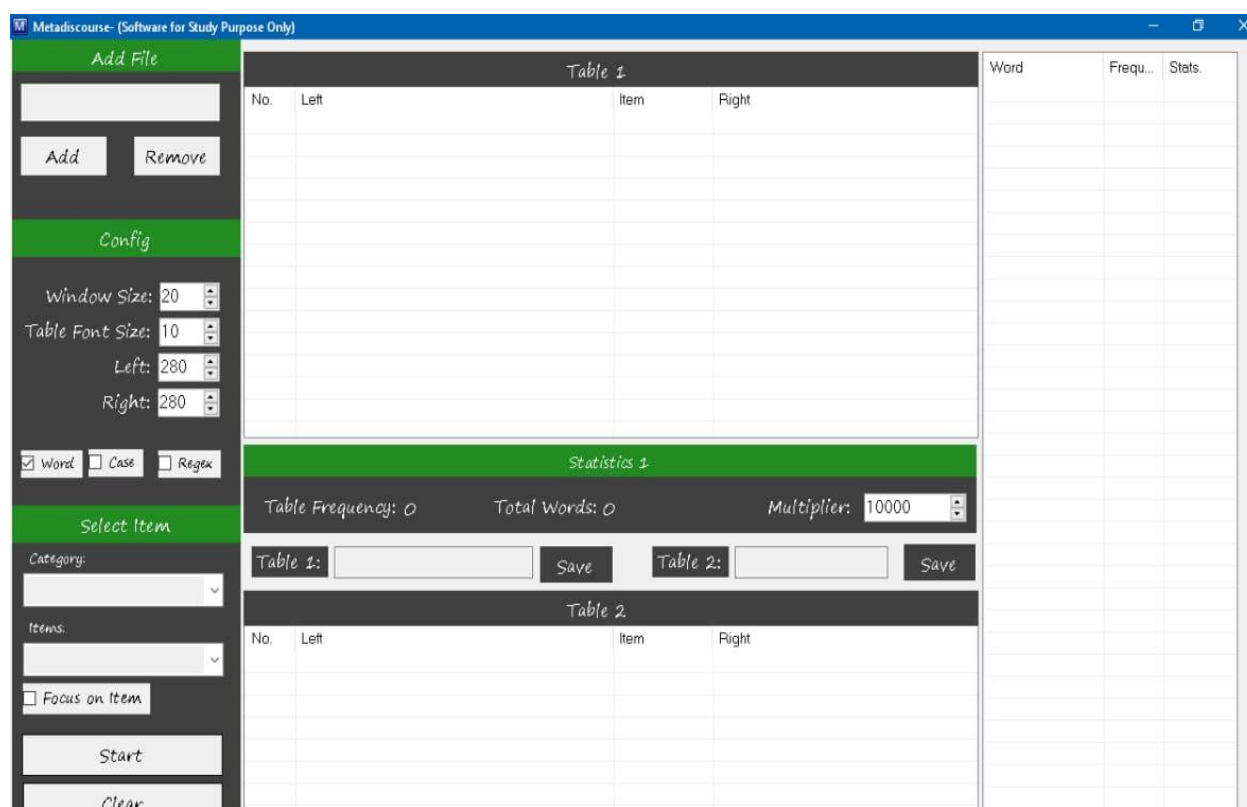


Figure 1: MetaPak tool used for the analysis

The corpus-based analysis was conducted in the current study as it derived results from the analysis of corpus of hedges and boosters in opinion articles of Pakistani English Newspapers by following qualitative approach. For quantitative analysis, the researcher used the frequencies obtained from *MetaPak* software. The identity of the writers was kept confidential. The sampling was random but the newspapers published November 2018 onward were chosen for the analysis.

3.3. MetaPak

MetaPak is a corpus tool for metadiscourse analysis of any text. The tool is made by keeping in view the Hyland's model (2005), *Taxonomy of Metadiscourse*. *MetaPak* is helpful in finding the frequencies, concordances and normalized value. It fulfills the gap found in the previous tools like AntConc and WordSmith.

3.4. Theoretical Framework

The current study is based on the famous Hyland's (2005) model *Taxonomy of Metadiscourse*. The model consists of two categories: textual features and interpersonal features. According to the model, *textual category* consists of five further types namely, Logical connectives, Frame marker, Endophoric markers, Evidentials and Code glosses. The *interpersonal features* are those that involve and engage the reader in the text. Interpersonal category also has five types, namely, Hedges, Boosters, Attitude markers, Self-mentions and Engagement markers. The current study analyzes only two interpersonal features: hedges and boosters. Hedges are those features that reflect the full commitment of the writer to the sentence; for example, *may*, *might*, *could*, *would*, *perhaps*, *some*, *possible* and boosters are those that show writers certainty or emphasize force; for example, *definitely*, *must*, *in fact*.

4. Findings and Discussion

This section deals with the findings and discussion of the data. It focuses on the qualitative and quantitative findings of the research. The findings of hedges and boosters are discussed in two different sections.

4.1. Use of Hedges in Opinion Articles

Below is the figure 2, which represents the occurrence of 58 hedges in the selected corpus of 47927 words. 58 hedges collectively occur with a frequency of 457. *Could* and *would* show the same occurrence during the study of the corpus i.e. 58 times and in total they comprised a percentage of 25.38. *May* also shows occurrence which is 52 times. Analyzing the three of them separately will constitute a very significant finding i.e. 36.76%. This significant finding is quite interesting and noteworthy as it makes 11% of one fourth of the total result. This is very interesting to note that the highest frequently occurring hedges are the modal auxiliaries, which show a writer's low tone and careful attitude towards the sentence and his words. The rest 63.24% is distributed among the other 56 hedges found. Besides, the most commonly occurring hedges i.e., *could* and *would*; *may* is also used 52 times. Furthermore, *may*, *might* and *maybe* together makes a sum of 75 of the total occurrences. This depicts that not only one form can be used as a hedge but different forms could also be implied as they show, somehow a different stance of writer. Including further, different forms of verbs are observed during the study. For instance, the occurrence of lemmas *claim*, *argue* and *appear* are found to be 31, 12 and 16 respectively. Surprisingly, hedges such as *sometimes*, *indicates*, and *usually* occurs the least in the corpus with a frequency of 1, 2 and 3 times respectively.

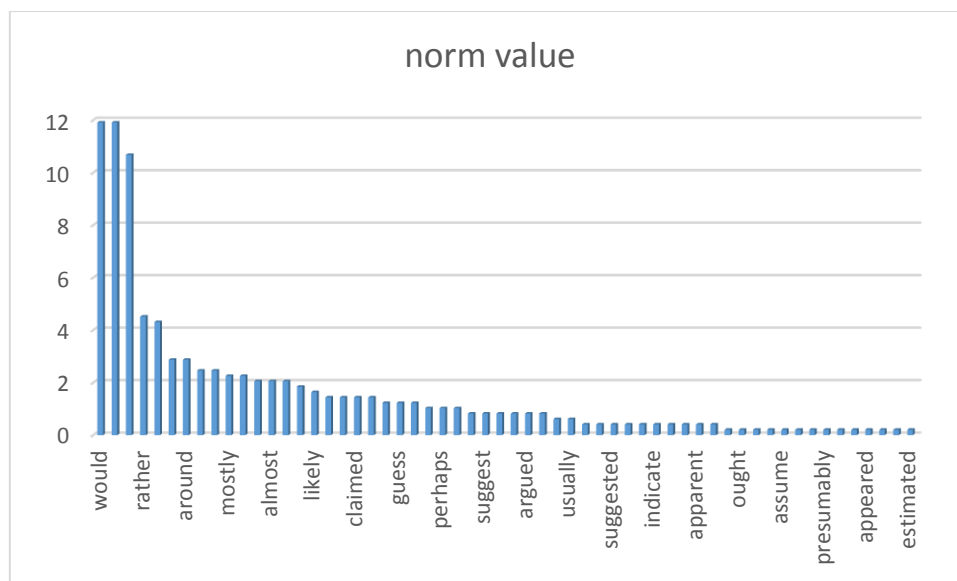


Figure 2: Hedges in the opinion article corpus

4.1.1. Examples of the Hedges from the Corpus

As it is evident from the findings that the most commonly used hedge markers in the opinion article corpus are the modal verbs *would* and *could*. Besides these two, other modal verbs are also used. Now, let us examine how these are employed as a hedge and what stance they possibly could show. Writers use modal verbs or other modal expressions when they need to project their voice about an opinion or attitude with a light and soften tone. All modal expressions writers' practice is about their view of the world. Considering the examples taken from the corpus (shown below) about *would*, it shows that it is used when a writer want to show his views about the future or to predict what might be a result of something in a coming time or it might also be a some kind of suggestion or a vision of the writer. Writers' also use this modal verb when want to show some sort of possibility and want to make some suggestions about something.

- *Sindh energy minister should be speedily issued. The installation of one or more solar panel production plants in Pakistan **would** be a profitable venture. Pakistan may have delayed exploiting Thar coal far too long.*
- *.....mental stunting kids are born with are chiefly because of mothers' malnourishment during their pregnancies. Under such circumstances, one **would** be forced to think that there isn't enough food available to sustain the entire population of 7.5 billion.*
- *Non-Muslims **would** be able to progress and contribute to the Pakistani economy much better than they do presently. It **would** uplift the Pakistani economy in addition to ensuring that Pakistan fulfils its obligations under international law. Some of our.....*

Besides the modal verbs' usage as hedges, there are other hedges that indicate a softer tone in a more polite way. These could be *probable*, *probably*, *suggest*, *claim*, *argue*, *almost*, *appear*, *sometimes* and *indicate* etc. Analysing the usage and the stance of *claim* used by the writers, it could be said that when it is used it shows that writer is very much careful in making statements. He prefer not to say anything strong that he could stand with and prove it, therefore, at some point he use *claim* as to show that it is not his but others' stance of something. Such statements could be true or not, but it depends on the reader to believe it or not.

Following examples from the corpus regarding the hedge *claim* are mentioned below:

-advertising agencies would, naturally, have us believe that it definitely is so. In justifying their existence, these advertising companies **claim** that commercials help afford to the consumer the right to choose. The critics, on their part, argue that commercials.....
- The US also needs to abandon its rhetoric of Pakistan playing a duplicitous role and give credence to her **claims** of having taken indiscriminate action against all the terrorist entities based on her soil. Pakistan has suffered the most.....
-white is beautiful. Woe unto those damsels who are born with dusky complexions! Advertisers, unashamedly, promote products that are **claimed** to perform the miracle of whitening the skin virtually overnight. One is left wondering about the moral right of.....

4.2. Use of Boosters in Opinion Articles

Figure 2 shows the occurrence of 57 boosters 480 times in the selected corpus. The frequently occurring boosters with the minimum occurrence of 14 are used 257 times is almost 53.54%. The other is distributed among 43 Boosters. The most frequent expression of assertion is *should* with the occurrence of 58 times which is 12.08% of the total occurrence of 57 Boosters. The other most frequently occurring boosters are *must*, *never*, *fact* with the occurrence of 28, 24 and 22 respectively. Moreover, the use of *certainly* and *actually* are also found relatively frequent with the occurrence of 14 and 10 respectively in selected corpus. *Evidently*, *undeniable*, *clearly* and *of course* are among the boosters which are least occurring in the selected corpus of opinion articles.

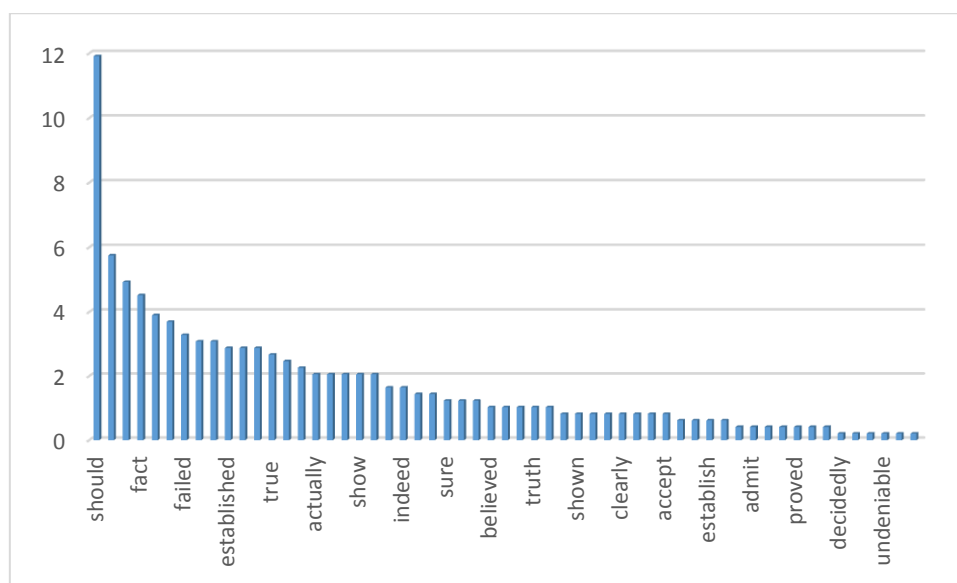


Figure 3: Boosters in opinion article corpus

4.2.1. Examples of the Boosters from the Corpus

As findings of the current study shows that the most frequently occurring booster is *should* with the frequency of 58. Along with *should* other boosters are also used such as *must*, *actually* etc. Now, let us analyze how these are used as boosters and what stance they possibly could show. Writers usually use *should* to give advice and make suggestions through their writings. The negative form of *should* i.e. *should not* is usually used in formal contexts

or when writer or speaker wants to emphasize something and forcefully project his voice (Should. n.d.). The examples taken from the corpus (shown below) about *should* indicate that writer put emphasis on his stance and he is firm on his standpoint. Also, *should* is used most commonly to talk about what is the ideal or best thing to do in a situation. Analysing the usage and the stance of *claim* used by the writers, it could be said that author is forcefully conveying his stance through the use of *should not*. Negative form of *should* indicates forceful stance of the writer. Author used *should* to display his position about the situation. Writer used the form *should be* for the best way that can be adopted in a situation. Writer also uses *should be* in the way of suggesting.

1.that is worse than killings and exploits religious passion to openly mock, divide and undermine the state? Imran Khan **should** not ignore what history's verdict on him might be. We generally know what the essential issues are. Similarly, we.....
2.a bulk of the food that is laid to waste to be consumed. Finally, the amount of food wasted **should** be consumed through regulated and efficient ways. In lieu of letting people fall prey to hunger-related diseases, the tones.....

Besides the modal verbs usage as a booster, there are other boosters that indicate a firm and strong tone of the writer. The other boosting device in the corpus was *must* with the occurrence of 28 times. *Must* expresses strong obligation and necessity but in the examples below, writer used *must* to show strong recommendation. Through the analysis of the usage and stance it could be said that author has employed strictly recommending tone while writing. Writer used *must* to show suggestion. It shows that writers have adopted recommending and suggesting tone and position about the issue.

Following examples from the corpus can be seen to clear that the writer used *must* in such a way or not.

1.encourage and support efforts by Pakistan to fight terrorism and restoration of peace in the region. The US particularly **must** acknowledge ground realities and make her peace initiatives compatible to them. Any approach contrary to the realities is bound.....
2.after Jinnah's death, when he resigned and left Pakistan for good. His letter makes for extremely sad reading. We **must** freely admit that we in Pakistan have failed to live up to the various promises we made to the.....

An interesting finding according to researcher was the use of booster *actually* to show sarcasm. As a discourse marker, *actually* is often used in speaking. It is used to get somebody's attention, to introduce a new topic or to say something that somebody may not like. It is also used in speaking to emphasize a fact, to show a contrast between what is true and what somebody believes, and to show surprise about this contrast and to correct somebody in a polite way. *Actually*. (n.d). The booster *actually* used by the writer sarcastically in the following example. Writer expresses his ironic tone about the topic through the use of *actually*.

3.in 90 days once the PTI came into power? Well, guess what? Not happening. If you are someone who **actually** believed this could happen, the joke's on you. But moving on? Remember the Prime Minister's appeal to the overseas.....
4. judged on the same standards. Despite this sameness, many in the party still harbor fantastic illusions about themselves. Many **actually** believe in their own fanciful rhetoric. But fantastic illusions and fanciful rhetoric is no match for the power of.....

Answer to the first research question is quite simple and quantitative. After analyzing the corpus and using the *MetaPak* tool for the identification of boosters and hedges, it is found that the total interpersonal markers; hedges and boosters are 937 with the former comprising 457 and later 480. This result shows that there is a very significant usage of hedges and boosters in the Pakistani English opinion articles. The most frequently used hedges in the opinion articles are the modal verbs and among the modal verbs; *would* and *could* are the most used. They comprise a total of 36.76% among the hedges which is a very substantial and momentous result. In addition to modal verbs, the other commonly used hedges are found to be *rather* followed by *seems*, *claims* and *around* with a frequency of 22, 21, 14 and 14. These mentioned hedges are commonly used which makes a sum of 52.29% of the total hedges used in the corpus. Evaluating the boosters, it is disclosed that the most repeatedly existence is of the booster marker *should* with a percentage of comprising 12.08% of the total alone. *Should* is followed by *must* and *never* with a frequency of 28 and 24 respectively. The most frequently occurring boosters are found to be 20 with a minimum frequency of 10 and making 71.45% of the total. This shows the significance of the boosters and its frequency of employment in the opinion articles.

The answer to the second research question is very interesting and fascinating as Hyland (2005) refers stance as the personal voice of the writer. Hedges and boosters are actually the two elements or the key dimensions through which a writer show his voice and either disguise his view in the form of careful use of hedges or project his words with clearly defined straightforward words. The findings revealed that how the stance of the writer could be employed by the use of hedges and boosters. By using the hedge marker, the writer becomes more careful about his writing and the choice of words. It shows that writer does not want to say something explicitly about any particular topic. By the use of hedge in opinion article, writer makes his voice soft and lenient. Sometimes writers also use hedges in the writing when they are not sure about something; they express doubt and uncertainty in their work and writing. The study of the selected corpus shows that the writers in many cases don't want to commit themselves to the statements they make; rather they show a sort of uncertainty in their sayings and opinions. Opinions by using hedging markers are not powerful and strong as the boosters are.

Contrary to this, boosters are quite different in their use and manner of their tone. A writer uses boosters to show his stance of being very confident and blunt about his words. By using boosters, writer makes his statements firmer and stronger in its nature. A strong emphasis is put in the statements by the use of boosting makers such as *should*, *must*, *most*, *actually*, *always*. Writers employ boosters in their writings to convince the readers about their opinions and suggestions. Form the selected corpus it is concluded that the Pakistani writers use a number of boosters in their writings to more it more solid, rich and assertive. In some instance these boosters are applied to make firm suggestions as a suggestion is very important and significant. All these ways reflect the mind and stance of the writer how they engage their writer in reading. The most instant stance where a writer might want to use booster is where he is sure about something and want to deliver a fair amount of assurance.

5. Conclusion

A corpus-based approach was used to analyze and investigate the employment of interpersonal markers i.e. hedges and boosters in the Pakistani English opinion articles. Findings of the current study show the diversity in the usage of the hedges and boosters by the Pakistani opinion article writers. A slight difference in the frequency of the both hedges

and boosters was discovered during the study. 457 hedges and 480 boosters were employed by the Pakistani authors in giving their opinion and personal attitude in different situations and to the different issues. The current study shows that Pakistani writers use hedges and boosters to show their stance and personal belief about something. They can show their commitment and reserved nature using certain hedging markers. Same can be concluded for the boosters. As the name indicates, boosters are for boosting. Authors use boosting markers to show how confident they are to their words and statements. However, a correct usage of a booster can persuade and convince the reader to his statements. It is also noted during the study that the use of modal verbs as hedges and boosters is more significant as compared to the rest of the hedging and boosting markers.

In short, there is a frequent use of hedges and boosters in the Pakistani opinion articles. These markers reflect and depicts different stance of the writer in a different context. A writer can either lower his tone by the use of less persuasive marker hedge or can boost up and assert his statement by the use of booster.

References

- Abbas, A., Shehzad, W., & Ghalib, H. (2017). MetaPak: An exclusive corpus tool for metadiscourse analysis, presented at Metadiscourse across Genres, North Cyprus, 2017. North Cyprus: Middle-East Technical University.
- Abbas, A., & Shehzad, W. (2018). A study of hedges and boosters in the final investigation report of joint investigation team of panama case-Pakistan. *Corporum: Journal of Corpus Linguistics*, 1(1), 1–14
- Abdi, R., Rizi, M. T., & Tavakoli, M. (2010). The cooperative principle in discourse communities and genres: A framework for the use of metadiscourse. *Journal of Pragmatics*, 42(6), 1669–1679. doi:10.1016/j.pragma.2009.11.001
- Aib, I., & Zahra, T. (2018). Metadiscourse practices: A corpus-based study of attitude markers in Pakistani English book reviews. *Corporum: Journal of Corpus Linguistics*, 1(1), 15–27.
- Cao, F & Hu, G. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English- and Chinese-medium journals. *Journal of Pragmatics* 43(11), 2795–2809. doi: 10.1016/j.pragma.2011.04.007
- Crystal, D. (1991). *A dictionary of linguistics and phonetics*. Oxford, England: Blackwell.
- Crystal, D. (1992). *An encyclopedic dictionary of language and languages*. Oxford, England: Blackwell.
- Holmes, J. (1982). Expressing doubt and certainty in English. *Regional Language Centre Journal*, 13(2), 9–28. doi: 10.1177/003368828201300202
- Hyland, K. (2004). Graduates' gratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes*, 23(3), 303–324. doi: 10.1016/S0889-4906(03)00051-6
- Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. London, England: Continuum.
- Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. doi: 10.1177/1461445605050365
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4), 458–508. doi: 10.1007/BF00262952
- McEnery, A. & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh, Scotland: Edinburgh University Press.
- More information about opinion articles. (2005, December 29). *The New York Times*. Retrieved from <https://www.nytimes.com/2005/12/29/opinion/more-information-about-opinion-articles.html>
- Should modal verb - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com. (n.d.). Retrieved from <https://www.oxfordlearnersdictionaries.com/definition/english/should>
- Takimoto, M. (2015). A corpus-based analysis of hedges and boosters in English academic Articles. *Indonesian Journal of Applied Linguistics*, 5(1), 95–105. doi: 10.17509/ijal.v5i1.836
- Talbot, M. (2010). *Language and gender* (2nd ed.). Cambridge, England: Polity.
- Trajkova, Z. P. (2011). Toning down statements in newspaper editorials. *The Journal for Languages and Literatures of the Faculty of Philosophy in Novi Sad*, 1(1), 71–84. doi: 10.19090/zjik.2011.1.71-84
- Zafar, S. (2018). An exploration of stance in research article abstracts through boosters and hedges. *Corporum: Journal of Corpus Linguistics*, 1(1), 28–47.

Attitude Markers as Metadiscourse in Pakistani English Newspaper Editorials: A Corpus-Based Study

Ali Raza Siddique

PhD Candidate, Department of Applied Linguistics, GC University, Faisalabad

Asim Mahmood

Professor, Department of Applied Linguistics, GC University, Faisalabad

Hafiz Muhammad Qasim

Lecturer, Department of Applied Linguistics, GC University, Faisalabad

Abstract

The primary goal of this study is to investigate the writers' choice of attitude markers to explore how these preferred markers show the writers' stances and styles of communication to their readers. The secondary goal is to study how attitude markers serve their functions in Pakistani English Newspaper Editorials. For the study, a total number of 1000 of editorials, 250 editorials from each of the four major Pakistani English Newspapers; i) Dawn News, ii) The News, iii) The Express Tribune and iv) The Frontier, were collected. Textinspector.com and Hyland's taxonomy of attitude markers (2005a) were used as tools to identify the attitude markers in the corpus data. This finding of the study showed that the Frontier's editorialists used attitude markers more frequently than the other three newspapers' editorialists. Dawn News editorialists were the second frequent users. In conclusion, the editorialists of The Frontier and Dawn News used a large number of attitude markers to show their stances and to communicate their opinions.

Keywords: attitude markers, metadiscourse, Pakistani English newspaper editorials, a corpus-based study

1. Introduction

There are many different ways by which humans communicate their attitudes. Attitude markers are one of them. According to Qadir and Shakir (2015, p. 34), attitude markers such as 'fortunately, happily and a number of other expressions represent addresser's stance in a particular situation or event at sentence level and discourse level'. In this research study, we make an attempt to explore how attitude markers have been used as 'interpersonal metadiscourse' in Pakistani English Newspaper Editorials. Attitude markers are, by definition, non-propositional items used by the writers to establish a personal connection with their readers. Interactional metadiscourse analysis of written data shows ways in which writers' attitude are expressed.

In the last few years, we have seen a revival of research interest in studying written discourse. A number of efforts have been made to investigate the ways by which a text is developed through what is called metadiscourse analysis (Halliday & Hasan 1976; Beaugrande & Dressler, 1981). Williams (1981) defines metadiscourse as "writing about writing" (p. 40). According to Vande Kopple (1985), metadiscourse is "communication about communication" (p. 83), and for Hyland (1998), it is "discourse about discourse" (p. 437). The overall goal of metadiscourse is to study how linguistic resources are used to communicate the writers' stance.

Attitude markers are pragmatic markers by implication and they have been referred to under various terms in the discourse literature. Halliday and Hassan (1976) call them ‘sentence connectives’; Van Dijk (1979) and Stubbs (1983) label them as ‘pragmatic connectives’; Quirk, Leech, and Svartic (1985) name them as ‘semantic conjuncts’; Schorup (1985) sees them as ‘discourse particles’; for Fraser (1988, 1990), these are ‘pragmatic markers’; Bazanella (1990) names them as ‘phatic expressions’; Redeker (1990, 1991) calls them as ‘discourse operators’; Erman (1987) views them as ‘pragmatic expressions’; Knott and Dale (1994) tag them as ‘cue phrases’; Ariel (1994) describes them as ‘pragmatic operator’; Östman (1995) names them ‘pragmatic particles’; Blakemore (2002) and Schiffrin (1987) call them ‘discourse markers’, and Hyland (1998, 2005a) calls them ‘metadiscourse’. All these terms have specific meanings and functions for these respective researchers. In this study, we make an attempt to study those pragmatic markers that function as attitude markers to investigate how these by writers to communicate to their readers to their attitudes toward contents. The present study considers the use of attitude markers as part of interpersonal metadiscourse. Both quantitative and qualitative approaches have been used to analyze data.

2. Literature Review

The past studies contributed in the development of theoretical framework. For example, De Bryun (1998), attitude markers ‘serve as a means by which the user of the language makes obvious what his feelings, emotions, or views are about the propositional content of the utterance being made.’ He further considers attitude markers as signals that express humans’ emotions. Milne (2008) is of the viewpoint that “attitude markers neither perform the functions of connective words nor take part to communicate conceptual meaning of the text” (p. 13).

Oskouei (2011), in her broad study on metadiscourse, explored the interactional variation between English and Persian discourses. She, specially, worked on the use of ‘certainty and ‘uncertainty’ markers in Magazine editorials. She studied the distribution of attitude markers in the editorials using both qualitatively and quantitatively. She concluded that the use of interactional metadiscourse features by British and Persian editorialist reasonably varied from one another due to their respective cultural backgrounds and the British editorialists favored the use of uncertainty markers while the Iranian editorialists favored the use of certainty markers.

Kindiki’s analysis (2009) of discourse in Kiitharka’s language spoken in, Bantu, Keyna showed that discourse/speech modifiers, discourse particles, pragmatic particles and discourse operator pragmatic markers can also function as attitude markers. Blagojević (2009) conducted a research study on the use of ‘attitude markers’ in discipline of academic research articles. He studied Serbian and English research articles of various disciplines such as sociology, social psychology and philosophy. He discussed the distribution of attitude markers in the following ways: a) Adverbs and adverbial phrases functioning as sentence adverbials – disjuncts, b) Verb-modifying adverbs functioning as subjuncts – intensifiers, c) Adjectives functioning as subjective complements in sentences with expletive ‘it’, d) Adjectives functioning as prenominal modifiers, e) Modal verbs expressing obligation, f) Nouns of specific semantic content. The findings of the study indicated that Serbain research articles compared to English research articles more readily expressed their attitudes than their English colleagues.

Negahdari (2009) studied the distribution of six types of attitude markers: a) Adverbial phrases and adverbs functioning as sentence adverbials – disjuncts, b) Verb-modifying adverbs functioning as subjuncts – intensifiers, c) Adjectives functioning as subjective complement in sentences with expletive ‘it’, d) Adjectives functioning as pronominal modifiers, e) Modal verbs expressing obligation, and f) Nouns of specific semantic content. Milne (2008) investigated ‘deontic verbs’. Deontic verbs are used when the addressees are expected to oblige the addressers’ intentions. Some of the deontic verbs are *have to*, *must*, *it is advised to*, and *ought*. Generally, deontic statements express the importance of actions in addressers’ perspectives.

2.1. Listing of Models for Developing a New Model on Attitude Markers

Table 1 provides a summary of various types of attitudinal markers as identified by researchers.

Table 1: Attitudinal markers

		Evaluation	
		Appraisal	
Attitudinal Markers: (Oskouei, 2011)	Markers of Value	Expressions of Obligation Expressions of Attitude	Deontic Verbs Attitudinal Adverbs Attitudinal Phrases have to, must, it is advised that, it is necessary, should remarkably, fortunately, unfortunately, happily even more surprising, It is interesting, it is worth saying, it is only common sense that
	Negation Expressing Counter-Expectancy	Negation Adversative	they do not feel responsible but, although, not only ... but also
Attitude Markers: (Schiffrin, 1987 & Fraser, 1990)	Verb	I rejoice, I regret, apologize, congratulate, appreciate, deplore, regret, detest, welcome, thank	
	Adverbs Interjections Whole Idioms	unfortunately, luckily, happily	
Attitude Markers: (Blagojević, 2009)	Adverbs and adverbial phrases functioning as sentence adverbials	Disjuncts	Paradoxically, Clearly, Predictably, Critically
	Verb-modifying adverbs functioning as subjuncts	Intensifiers	fully, calmly, partly, individually, separately, collectively, slightly, specifically, relatively, separately, strongly, directly, essentially, primarily, discursive, dynamically, statically, clearly, actually, obliquely, punctually, violently, grammatically, nonviolently, precisely, metonymically, dynamically, fully, statically, literally, congruently, metaphorically, critically, grammatically, critically, linguistically, clearly, directly, symbolically, uniquely, typically, selectively, technically, miserably, highly, contextually, probably, socially, generically, strongly, normally, necessarily, entirely
	Adjectives functioning as subjective complement in sentences with expletive ‘it’ Adjectives functioning as pronominal modifiers Modal verbs expressing obligation Nouns of specific semantic content		

Attitude Markers: (Halliday & Hassan, 1976)	Addition	and, moreover, also, in addition, besides, furthermore
	Contrast	however, but, in contrast, the former – the latter, on the one hand – on the other hand, actually, while, nevertheless
	Similarity	similarly, such, equally, the same
	Exemplification	for instance, for example, in other words
	Chronology	second, first, then, thereafter, afterwards, meanwhile, next later, at the same time, finally, ultimately, at last
Attitude Markers: (Milne, 2008)	Causality	consequently, so, therefore, accordingly, thus, although, hence, because of, since, as a result
	Attitude	naturally, of course, obviously, unfortunately, fortunately, admittedly, certainly
	Summary	to sum up, to summarize, in brief, in conclusion
	Deontic Verbs	have to
	Attitudinal Adverbs	remarkably, unfortunately
	Attitudinal Adjectives	It is surprising, It is absurd
	Cognitive Verbs	I think, I feel

The study had three research questions. These are as follows:

1. How frequently attitude markers are used in Pakistani English Newspaper Editorials?
2. What are their functions?
3. How similar are the editorials of Dawn News, The Express Tribune, The Frontier, and The News in their use of attitude markers?

3. Methodology

The present study considered attitude markers as a sub-category of interactional metadiscourse. Both quantitative and qualitative methods were used to study the data. The quantitative study focused on obtaining the frequencies of attitude markers as used in the newspaper editorials and comparing the frequency counts across the four newspapers. The qualitative study focused on initially studying how attitude markers were used as metadiscourse and later on developing a classification system of attitude markers as well as identifying the propositional content in the Pakistani English Newspaper Editorials. The data consisted of 1000 editorials (250 from each of the four newspapers: *Dawn News*, *The News*, *The Frontier* and *The Express Tribune*). In order to ensure that diachronic changes did not affect the selected editorials, newspaper editorials that were published during the months of March and April 2016 only were included in the dataset and were analyzed.

3.1. Proposed Model for this Study

In order to cover the qualitative component of the study, a new model called ‘interactional category’ was developed and used. Please see table 2 for details regarding the proposed model:

Table 2: Proposed model for this study

Interpersonal Category	Attitudinal Markers	Expressions of Obligation
		Expressions of Attitude
		Negation expressing counter-expectancy

3.2. Formation of List of Attitude Markers

In the proposed model, attitude markers were codified as a sub-category of ‘interactional’ in ‘interpersonal metadiscourse’. For codification, attitude markers were taken from both textinspector.com and Hyland’s (2005a) book *Metadiscourse: Exploring Interaction in Writing*. After refining the final lists, both lists were merged and all duplicate markers were removed. As it can be seen in table 3, there were a total of 72 merged markers.

Table 3: Formation of final lists of attitude markers

Categories	Sub-Categories	Textinspector.com	Ken Hyland’s Book	Duplicate Markers	Merged Markers
Interpersonal Metadiscourse	Attitude markers	26	64	08	72

3.3. Proposed Attitude Markers as Sub-Category of Interactional Metadiscourse

For analyzing the data, this study has identified a new set of attitude markers classified under what I prefer to call ‘interactional metadiscourse’. See the list below:

Table 4: List of proposed attitude markers as interactional metadiscourse

Attitude Markers				
!	Desirable	Fortunate	Prefer	Unbelievably
Admittedly	Desirably	Fortunately	Preferable	Understandable
Agree	Disagree	have to	Preferably	Understandably
Agrees	Disagreed	Hopeful	Preferred	Unexpected
Amazed	Disagrees	Hopefully	Remarkable	Unexpectedly
Amazing	Disappointed	Important	Remarkably	Unfortunate
Amazingly	Disappointing	Importantly	Shocked	Unfortunately
Appropriate	Disappointingly	Inappropriate	Shocking	Unusual
Appropriately	Dramatic	Inappropriately	Shockingly	Unusually
Astonished	Dramatically	Interest	Striking	Usual
Astonishing	Essential	Interesting	Strikingly	
Astonishingly	Essentially	Interestingly	Surprised	
Correctly	Even	Must	Surprising	
Curious	Expected	Ought	Surprisingly	
Curiously	Expectedly	Pleased	Unbelievable	

3.4. Development of Corpus

Four sets of corpora were retrieved from Pakistani English Newspaper Editorials. Each corpus went through multiple stages as described below.

3.4.1. Data Procurement

The editorials were retrieved from the internet sites of the respective newspapers:

- 1) www.dawn.com/authors/2677/editorials
- 2) www.tribune.com.pk/author/79/editorials
- 3) www.thefrontierpost.com/category/33/editorials
- 4) www.thenews.com.pk/print/category/editorials

3.4.2. Retrieval of Corpus

For data retrieval, the data were retrieved in the following stages. Firstly, the national editorials were retrieved manually in the word files from online sources as mentioned above. Second stage involved recording metadata in excel sheets under the following categories: files number, word types, token types, source of data, retrieval information from online, title of data, published date, and names of writers of the editorials. In the third stage, the word files were renamed for identification. In fourth stage, the publication date, the authors' name and titles of editorials were removed. In the fifth stage, word files were converted into notepad files. In the sixth stage, all editorials were gathered in a single folder, named: 'compiled up files 1 to 1000'.

3.4.3. Corpora Size and its Distribution

The selected editorials were examined under the proposed model of attitude markers. The corpora size is given in table 5 below.

Table 5: Corpora size

Newspapers	Token Words	Word Types
Dawn News	103,596	10,053
The Express Tribune	93,048	9,737
The Frontier	172,878	14,812
The News	103,860	9,633
Total Corpora Length	473,382	44,235

For the analysis purpose, the identified attitude markers were divided into three types: a) Expressions of Obligation, b) Expressions of Attitude and c) Expressions of Negation and counter-expectancy. For doing the analysis, this study developed a corpus expression of attitude features as proposed in table 6. The significance of the expression opened new avenues for the researchers. The attitude markers turned into corpus expression and it was used to count the features of attitude markers in frequency counts at once. See table 6 below:

Table 6: Developed expressions of metadiscourse features

Expression for Attitude Markers	One-word expression	amazingly admittedly appropriately agrees agree amazing appropriate amazed astonishing astonishing astonished correctly curiously disappointing curious disagree Desirably desirable disappointed disagreed disappointingly disagrees dramatically dramatic essential even essentially expected fortunately expectedly fortunate hopefully important hopeful importantly interest inappropriate inappropriately interesting prefer interestingly pleased Preferable preferably preferred must ought remarkable remarkably surprisingly shocked shocking shockingly striking strikingly surprised surprising unfortunate unfortunately unusually understandably unbelievable unbelievably understandable unexpected unexpectedly unusual usual
	Two-word expression	have to even x

4. Results and Discussion

The Quantitative findings with regard to the use of propositional and non-propositional attitude markers in Pakistani English Newspaper Editorials (e.g. (for example, *Dawn News*, *The News*, *The Express Tribune* and *The Frontier*) are presented in the following tables.

Table 7: Results of *The News* corpus

Interactional Category	Sub-Categories	The News		
		Propositional	Non-propositional	Sum
	Attitude Markers	66	334	400

Table 7 shows that the News Corpus editorialists used attitude markers both propositionally and non-propositionally. A further analysis of the attitude markers revealed that *The News* editors used attitude markers non-propositionally in higher frequency. Table 8 contains findings from data *Dawn News*.

Table 8: Results of *Dawn News* corpus

Interactional Category	Sub-Categories	Dawn News		
		Propositional	Non-propositional	Sum
	Attitude Markers	30	761	791

Table 8 shows that the editorialists employed attitude markers propositionally and non-propositionally. A further analysis indicated that *Dawn News* editorialists used attitude markers non-propositionally in higher frequency. Table 9 contains findings from data on *The Express Tribune*.

Table 9: Results of *The Express Tribune* corpus

Interactional Category	Sub-Categories	The Express Tribune		
		Propositional	Non-propositional	Sum
	Attitude Markers	79	331	410

Table 9 shows that the editorialists at *The Express Tribune* used attitude markers propositionally and non-propositionally; however, further analysis indicated that they have used non-propositional attitude markers in higher frequency. Table 10 contains the frequency counts of attitude markers as identified in the corpus *The Frontier*.

Table 10: Results of *The Frontier* corpus

Interactional Category	Sub-Categories	The Frontier		
		Propositional	Non-propositional	Sum
	Attitude Markers	23	791	815

Table 10 shows that the editorialists used attitude markers both propositionally and non-propositionally; further analysis, however, showed that *The Frontier* employed non-propositional attitude markers in higher frequency.

Table 11 contains the combined results with regard to the numbers of occurrences attitude markers in the Pakistani English newspaper Editorials (e.g. *The News*, *Dawn News*, *The Express Tribune*, and *the Frontier*).

Table 11: Results of all corpora

Category	No. of Newspapers	Attitude Markers		Total Markers
		Propositional	Non-propositional	
Interactional Metadiscourse	Dawn News	30	761	791
	The Express Tribune	79	331	410
	The Frontier	23	791	814
	The News	66	334	400
	Total	198	2217	2415

The overall results show that all newspaper editorials used attitude markers non-propositionally than propositionally. As far as their comparison is concerned, *The Frontier's* editorialists had used higher number of attitude markers than any other newspaper editorialists (e.g. *The News*, *Dawn News*, and *The Express Tribune*). The results of MFs of each corpus are given below.

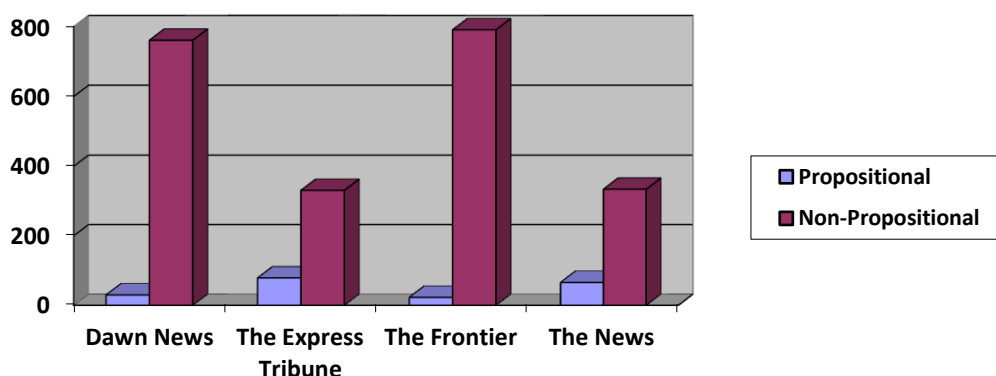


Figure 1: Results of MFs of each corpus

4.1. Attitudinal Markers

Attitude markers show a way of expressing or commenting in speaking or writing. The integral feature of this research is to study attitude markers. This study focused on non-propositional contents. Attitude markers are metadiscourse features. The logic behind the use of attitude markers is to show the attitude of the writers toward the propositional content and to “invite others to endorse their feelings, tastes and normative assessments they are announcing. Thus, declarations of attitude are dialogically directed towards aligning the addressee into a community of shared value and belief” (Martin & White, 2005, p. 95). As far as the current study is concerned, the definition of ‘attitudinal markers’ given by Hunston and Thompson (2000, p. 21) has been taken to do the analysis in the form of markers of value: ‘expressions of obligation, expressions of attitude, and expression of negative expressing counter-expectancy’ (Oskouei, 2011).

4.1.1. Expressions of Obligation

Expressions of obligation are helping verbs and impersonal expression made up of deontic verbs. In this study, the use of ‘expressions of obligation’ is subcategorized under Attitude markers similar to Milne (2008). This category has those statements of obligation which are used to address third party and are discussed through expressions of deontic such as have to, must, it is advised to, and ought as can be seen below.

1. Unhappily, democratic debate inside Pakistan far too often appears to be a private competition between political parties. In truth, democracy *ought* to be about the people and their evolving choices (The Dawn News file no. 138).
2. Also, European nations *must* be reminded of their pledge to help countries neighboring Syria that house refugees under great economic stress (The Dawn News file no. 73).
3. Germany will also *have to* develop a much more comprehensive and forceful foreign policy for Northern Africa and the Middle East (The Frontier file no. 72).

In the above-mentioned examples (1) to (3), ‘ought, must and have to’ have been used to indicate an obligation and to anticipate a suggestion about the circumstances in which the action is done; hence, it is regarded as ‘expressions of obligation’.

4.1.2. Expressions of Attitude

Expressions of attitude can be done through the use of attitudinal adverbs as shown in the following examples:

1. More *remarkably*, in attendance were avowed anti-democrats like Hafiz Saeed, whose interest in parliamentary democracy is nil (Dawn News file no. 218).
 - a. In example (4), *remarkably* has been used by the writer to highlight explicitly the interest of Hafez Saeed as well as to communicate the writer’s attitude.
2. It is obvious how tough the task is, especially given that it is not as yet certain how much leeway Bilawal Bhutto Zardari will be allowed by his father and other mentors, some of whom he could so *happily* do without (Dawn News file no. 203).
3. At the rate he’s going, there would soon be no person in the land left to judge him. *Fortunately*, the American legal system doesn’t work that way (The Frontier file no. 75).

In the above example (5), *happily* indicates that the writer feels that it will be better for Bilawal Bhutto Zardari to leave some of his mentors aside. Similarly, in example (6) *fortunately* has been used by the author to explicitly show how the American legal system works. Attitude has been expressed through attitude adverbs. Therefore, they are considered as interactional features. That is why; expressions of attitude are subcategorized under attitude markers. Through such markers, the writer has also shown his stance.

4.1.3. Negation Expressing Counter-Expectancy

In the previous studies, *negation markers* have not been regarded as metadiscourse features. Although, these can be a feature of *persuasive writing* as proposed by Martin and White (2008), in our opinion by using negation in writing, a writer can expose covertly that

the rejection is given after using alternative positive positions. The following examples are given below:

1. Cameron's allies *do not* dismiss the possibility of a 'reconciliation reshuffle' after the vote, though they emphasise that Cameron is not taking victory for granted -- a wise position in these volatile times (The Frontier file no. 117).
 - a. In above example (7), 'Cameron's allies do not dismiss the possibility' indicates the writer's view about Cameron's allies who are not willing to discharge the opportunity of reconciliation.
2. Khursheed said the opposition had a lot of reservations over Nawaz Sharif's speech that he delivered in the National Assembly as he did not answer the key questions of the opposition. He said the prime minister's family was *never* included in the 22 richest families (The News file no. 114).

In above example (8), the sentence 'He said the prime minister's family was never included in the 22 richest families' implies that Khursheed has a different viewpoint. Further, attitude marker is expressing the attitude as a stance of the author or the writer toward the propositional in the form of conveying surprises, agreement, giving importance in comparison with commenting status of information, i.e. relevance, undeniable, truth and etc. This subcategory is categorized under interactional metadiscourse as a major category in this study.

5. Conclusion

The goal of the study was to investigate how editorial writers effectively communicate their stances in their writings. The findings indicate that The Frontier editors used more attitude markers than the other editors. The findings also indicate that the excessive use of attitude markers enables the writers to communicate to their readers more effectively. The conclusion of the present study is made at theoretical and empirical foundations. To contribute in theoretical foundation, the purpose of this research has reduced the entangled aspects in the existing studies. Considering a new entity 'negation expressing counter-expectancy' has been first time used to show the writer's attitude. However, these new categories were introduced first time in Pakistan to cover this area. Secondly, the use of personal pronouns is actually propositional. Though, as the importance of personal pronoun function is concerned it engages readers. That is why; it has been taken into interpersonal metadiscourse. Only setting aside the one category pronouns, the personal pronouns employed in choosing text are taken in various classifications because of the nature of inclusion. When first person pronouns are taken with attitudinal verbs, such as *we think* and *I believe*, being supposed to be attitudinal expressions. To contribute in an empirical foundation, the most frequent use of sub-categories in the chosen corpora reveals the attitude of editorialists who engage their readers or audience interpersonally. The editorialists guide the readers personally with readers or audience. This study gives some important suggestions that the proposed categorization in this study could be used in comparison of native and non-native languages, texts and spoken languages as per cultural differences. This study may be conducted on the other newspaper of Pakistan and across other countries.

References

- Ariel, M. (1993). Pragmatic operators. In R. Asher (Ed.), *Encyclopedia of language and linguistics*. Oxford, England: Pergamon Press.
- Bazanella, C. (1990). Phatic connectives as intonational cues in contemporary spoken Italian. *Journal of Pragmatics*, 14(4), 629–647. doi: 10.1016/0378-2166(90)90034-B
- Beaugrande de, R., & Dressler W. U. (1981). *Introduction to text linguistics*. London, England: Longman Linguistics Library.
- Blagojević, S. (2009). Expressing attitudes in academic research articles written by English and Serbian authors. *FactaUniversitatis, Series: Linguistics and Literature*, 7(1), 63–73. doi: 811.111 : 811.163.41] : 001.891
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge, England: Cambridge University Press.
- De Bryun, P. S. (1998). Pragmatic markers: the neglected part of speech. *South African Journal of Linguistics*. 16(4), 127–135. doi: 10.1080/10118063.1998.9724149
- Erman, B. (1987). *Pragmatic expressions in English : A study of you know, you see and I mean in face-to-face conversation* (PhD dissertation). Almqvist & Wiksell International, Stockholm, Sweden. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-80642>
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1/4), 19–33.
- Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14(3), 383–398. doi: 10.1016/0378-2166(90)90096-V
- Halliday, M. A., & Hasan, R. (1976). *Cohesion in English*. London, England: Routledge.
- Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in text: Authorial stance and the construction of discourse*. London, England: Oxford University Press.
- Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30(4), 437–455. doi: 10.1016/s0378-2166(98)00009-5
- Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. London, England: Continuum.
- Kindiki, S. K. (2009). *Pragmatic functions of attitude markers in Kĩtharaka* (Master's thesis, University of KwaZulu-Natal, Durban, South Africa). Retrieved from <https://researchspace.ukzn.ac.za/handle/10413/281>
- Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1), 35–62. doi: 10.1080/01638539409544883
- Martin, J. R., & White, P. R. (2005). *The language of evaluation: Appraisal in English*. London, England: Palgrave Macmillan.
- Milne, E. D. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40(1), 95–113. doi: 10.1016/j.pragma.2007.10.003
- Negahdari, A. (2009). *Attitude markers in English and Persian*. Paper presented at University of Khorasan Islamic Azad, Iran. Retrieved from <http://conference.khuisf.ac.ir/DorsaPax/userfiles/file/pazhohesh/1-5.pdf>
- Oskouei, L. K. (2011). *Interactional variation in English and Persian: A comparative analysis of metadiscourse features in magazine editorials* (Unpublished doctoral dissertation). University of East Anglia, Norwich, England.

- Qadir, I. A. A, & Shakir, M. (2015). Attitude markers in the writing of EFL learners: A pragma-stylistic study. *مجلة أبحاث البصيرة للعلوم الإنسانية*, 40(3), 30–44. doi: 10.33762/0694-040-003-010
- Quirk, R., Leech, G., & Svartic, J. (1985). *A comprehensive grammar of English language*. London, England: Longman.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14(3), 367–381. doi: 10.1515/ling.1991.29.6.1139
- Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29(6), 1139–1172. doi: 10.1515/ling.1991.29.6.1139
- Schiffrin, D. (1987). *Discourse markers*. Cambridge, England: Cambridge University Press.
- Schourup, L. C. (1985). *Common discourse particles in English conversation*. London, England: Routledge.
- Stubbs, M. (1983). *Discourse analysis*. Oxford, England: Basil Blackwell.
- Van Dijk, T. A. (1979). Pragmatic connectives. *Journal of pragmatics*, 3(5), 447–456. doi: 10.1016/0378-2166(79)90019-5
- Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36(1), 82–93. doi: 10.2307/357609
- Williams, J. M. (1981). *Style: Lessons in clarity and grace*. Boston, MA: Scott Foresman.
- Östman, J. O. (1995). Pragmatic particles twenty years after. In *Organization in discourse* (pp. 95-108). Turku, Finland: University of Turku.

A Corpus-based Study of Reporting Verbs in Citation Texts Using Natural Language Processing

Imran Ihsan

Assistant Professor, Air University, Islamabad

Sarah Imran

English Language Teacher, Lahore Grammar School, Islamabad

Osama Ahmed

MS Scholar, Deakin University, Melbourne, Australia

Muhammad Abdul Qadir

Capital University of Science and Technology, Islamabad

Abstract

In scientific literary writings, authors often cite other researches to formulate their opinions and findings. The selection of the reporting verb for such purpose plays an important role in their citations. Reporting verbs may exhibit variety of strengths when used in different contexts and scenarios. Therefore, a compilation of reporting verbs used by authors in various contexts and its formulation in the form of a dataset can provide a basis for corpus-based analysis of citations and its reasons. Sentiment analysis techniques can categorize a citation into *Positive*, *Negative* or *Neutral* sentiments. Natural Language Processing techniques can automatically tag verbs used in a citation with high accuracy. This paper is a sentiment-based study, conducted to formulate a citations' reporting verb corpus, by categorizing the citation texts from a selected dataset into three sentiments. Using NLP techniques, reporting verbs are extracted from these citation texts and their frequencies are calculated. The study also describes the analysis of extracted verbs in each sentiment.

Keywords: reporting verb, citation analysis, sentiment analysis, NLP, scientometrics

1. Introduction

A reference to a published source or even an unpublished one is known as citation. Ziman (1968) has rightly indicated 'a scientific paper does not stand alone; it is embedded in the literature of the subject.' Similarly, Narin (1976) defines 'a reference is the acknowledgement that one document gives to another; a citation is the acknowledgement that one document receives from another.' Malin (1968) says, 'a citation implies a relationship between a part or the whole of the cited document and a part or the whole of the citing document.' In research papers, authors refer to the research of other authors to report on their own findings. For such purpose, reporting verbs are used such as *suggests* or *argues* etc. Reporting verbs differ in terms of their strength (assertiveness) and the context in which they are used. Same reporting verb can have two different meanings if used in a positive and in a negative context respectively. Therefore, it is important to understand the context in which a reporting verb is used within a citation.

In scientific writing, positive credits and negative criticisms can be identified using sentiment analysis techniques by finding the sentiment polarity of a citation context (Xu et

al., 2015). There are various sentiment analysis techniques including keyword spotting, lexical affinity, statistical methods, and concept-level techniques (Banker & Patel, 2016). Citation sentiment analysis, an application of sentiment analysis, determines the sentiment polarity of a citation context towards the cited paper. In citation sentiment analysis, a citation can be categorized as in one of three sentiment polarity classes that are *Positive*, *Negative* or *Neutral*. Athar and Teufel (2012) have performed experiments using citation sentiment analysis by tagging a citation with positive, negative or neutral intentions in computational linguistic publications. A publicly available dataset houses around 8700 citation texts along with each citation's sentiment polarity. An example is shown in Figure 1 that depicts three different citation texts with *Positive*, *Negative* or *Neutral* sentiment polarities.

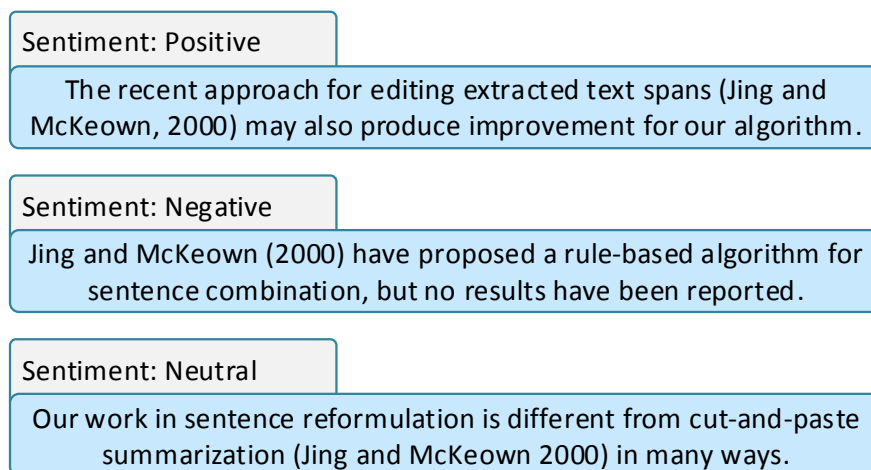


Figure 1: Citation texts with ‘positive’, ‘negative’ or ‘neutral’ sentiment polarity

In order to have a corpus-based study of reporting verbs in citation texts, sentiment polarity of a citation texts plays a vital role. Therefore, a mechanism must be formulated that can extract reporting verbs from citation texts along with the sentiment polarity in which that verb is used. Natural language processing techniques can help identifying the verbs within a citation text. In corpus linguistics, part-of-speech tagging (POS tagging or PoS tagging or POST), also called grammatical tagging or word-category disambiguation, is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. A simplified form of this is commonly taught to school-age children, in the identification of words as nouns, verbs, adjectives, adverbs, etc. Part-of-Speech tagging is possible using Natural Language Processing (NLP) Techniques. Using these techniques, verbs and reporting verbs can be identified within a citation text. Combining the sentiment polarity of a citation and use of reporting verb in a context can provide a basis of complex problems such as finding the cognitive relationship between a citing and cited paper or semantic enrichment of a citation graph.

This research paper focuses on development and analysis of a reporting verb corpus, by extracting the verbs from citation texts using NLP techniques, classifying them in Positive, Negative and Neutral classes based on the sentiment polarity of the citation text in which that verb is used. Paper organization: Section 2 explains the natural language processing methods and techniques to tag part-of-speech words and Section outline the NLP toolkits available for this purpose. Section 4 defines the methodology adopted to develop the corpus. Section 5 describes the experiments performed, results tabulated, and the analysis performed. Section 6 explains concludes the paper with future directions.

2. Literature Review

A Part-Of-Speech Tagger reads a text in a defined language tokenizes the text in words and assigns parts of speech to each word such as noun, verb, adjective etc. There are two distinct techniques that are used in Part-Of-Speech Tagging. One is defined as ‘Supervised’ tagging and the other is ‘Unsupervised’ (Adhvaryu, & Balani, 2015). ‘Supervised’ tagging used a pre-trained corpus whereas ‘Unsupervised’ tagging requires training of the corpus. Each technique has further classifications into ‘Rule Based’, ‘Statistical Method’ and ‘Transformation’ based techniques. In the rule-based technique, sentences in each text are tagged using manually designed rules. However, statistical method uses corpus-based techniques such as Hidden Markov Model (HMM), Maximum Entropy Markov Model (MEMM) and Conditional Random Fields (CRF). Lastly, transformation-based technique tags unannotated corpus by machine learning rules and feature.

2.1. Rule Based Method

Rule based method required a large set of manually designed rules that are disambiguated using morpheme ordering and contextual information (Kumar & Josan, 2010). Using these linguistic rules, the technique assigns each word in a sentence with an appropriate tag. Some of these rules are defined as *verb identification rule*, *noun identification rule*, *pronoun identification rule*, and *adjective identification rule* (Garg, Goyal, & Preet, 2012). Development of such rules is a complex and a time-consuming process.

2.2. Statistical Method

Using the probability of occurrences of words, statistical method tags the words and is most widely used as corpus-based technique. It also looks for the context of a sentence and because of that two similar words are identified as two different based on their context. Thus, if the first word is identified as a *noun* then second word will also be a *noun*, *verb* or an *adjective*. However, it is heavily subjected to the probability distribution. Statistical methods define three distinctive algorithms that can be used independently.

First is HMM – Hidden Markov Model (Youzhi, 2009; Jurafsky & Martin, 2000), second is MEMM – Maximum Entropy Markov Model (Lafferty, McCallum, & Pereira, 2001) and the last is CRF – Conditional Random Fields (Zhang, Huang, & Liang, 2009). HMM – Hidden Markov Model adapts a generative model and makes use of Markov assumption to define the probability of a tag dependent upon a small and fixed number of previous tags. MEMM – Maximum Entropy Markov Model uses a conditional probabilistic sequence model and is based in next state classifiers. However, like other independently trained next state classifiers, it is also a victim of label bias problem (Lafferty, McCallum, & Pereira, 2001). CRF – Conditional Random Fields is a sequence model problem with all the positives of MEMM with a resolved label bias problem. CRF uses a single exponential model by defining a joint probability for the complete sequence of labels in a selected sequence.

2.3. Transformation Based Method

Transformation based method (Bach, Cuong, Ha, & Binh, 2008) used a step-by-step method to improve the tagging. It initially defines a basic tag, that is improved with each iteration and the algorithm stops when no further transformation makes a tag optimal. The algorithm is trained using pre-annotated corpus. Each transformation works in two segments:

a *rewrite rule*, and a *triggering environment*. For example, one can apply a transformation using a rewrite rule known as *change tag from verb to noun* and then applying triggering environment that is *the preceding word is a determiner*. It means that the transformation can only be pertained to the word next to a determiner. Thus, when a transformation is applied, the tag applied on a word is changed from *verb* to *noun*.

In Transformation Based Method, initially uses a five-step approach for learning algorithm and then applies POS tagging in three stages. These stages are *label every word with most-likely tag*, *examine all possible transformation and select the most improved one* and *retagging*. These three stages are repeated until the terminating condition is reached.

2.4. NLP Toolkits

Natural language processing comes in the domain of artificial intelligence that deals with human interaction with the machine. Several tools and toolkits are available for natural language processing. These toolkits use diverse programming languages. Some toolkits are proprietary, and some are open source and easy to use such as NLTK, Apertium, Stanford NLP etc. The Table 1 shows a few natural language processing toolkits with the description of unique features defined in each toolkit.

Table 1: Tabulation of NLP Toolkits

Toolkit	Language	Open Source	Description
GATE (Cunningham, Wilks, & Gaizauskas, 1996)	JAVA		Sentence Splitter, POS Tagger
Ellogon (Petasis, Karkaletsis, Paliouras, Androutsopoulos, & Spyropoulos, 2002)	C, C++		TIPSTER
MontyLingua (Ling, 2006)	Python, Java		POS Tagger
IceNLP (Loftsson & Rögnvaldsson, 2007)	Java		Icelandic Language Only
NLTK (Bird & Loper, 2004)	Python	Yes	Human Natural Language Data
Apertinum (Brandt, Loftsson, Sigurbórsson, & Tyers, 2011)	C++	Yes	Spanish Language Only
Carabao (Vadim & Garr, 2013)	Visual C++		Natural Language Entity Morphology
PSI Toolkit (Graliński, Jassem, & Junczys-Dowmunt, 2013)	C++		Natural Language Processing

In *GATE Toolkit*, GATE is acronym of ‘General Architecture for Text Engineering’ and is an open source toolkit written in JAVA. The toolkit provides three basic elements: a database to store text and schema, a Graphical User Interface and a collection of wrappers.

Ellogon Toolkit is specially designed for engineering systems to improve information retrieval and extraction. The toolkit also provides a multilingual Graphical User Interface. *Monty Lingua Toolkit* is developed by MIT Media Lab, the toolkit uses Python language process all aspects of English language text. It defines its own POS Tagger known as Monty Tagger. *IceNLP Toolkit* is an open source toolkit designed for processing Icelandic texts and consists of POS Tagger, Module Pre-Processor and Finite State Parser. Its word error rate is 50.6%. *NLTK Toolkit* is written in Python and provides ready-to-use linguistics using statistical NLP. It also provides various built-in corpora for testing and classifiers such as Naïve Bayes, Maximum Entropy and Binary Tree. *Apertium Toolkit* is basically designed for Spanish language; it has integrated other languages such as Danish, French and Italian. It uses POS Tagger, Lexical Selection, Lexical Transfer, Structural Transfer, Morphological Generator. *Carabao Language Toolkit* can capture any form of natural language entity with complex morphology. The extracted sequences in Carabao can be viewed as a regular expression. *PSI Toolkit* is designed to process natural language processing texts. There are three types of processors used in this toolkit: *Readers*, *Annotators* and *Writers*.

Based on the feature of each toolkit, most widely used toolkits are Apertium, and Sanford NLTK. These toolkits are also open source and are available free of cost. After further study of these open source toolkits, we selected Stanford NLP toolkit based on our familiarity with programming language *Python* and the richness of features like tokenization, sentence splitting, POS tagging etc. It also has the potential to perform complex NLP tasks like *sentiment extraction*, *translation*, *spell checking*, and *word prediction* etc.

3. Research Methodology

Many studies have been conducted in past to analyze the reporting verbs in doctoral theses, students' assignments, research articles, and journals. Our research focuses upon development of a reporting verb corpus used by authors and researchers while citing other documents based on their sentiment polarity. Using Hyland's (2000) framework as a guideline in document analysis, we have formulated a methodology as shown in Figure 2.

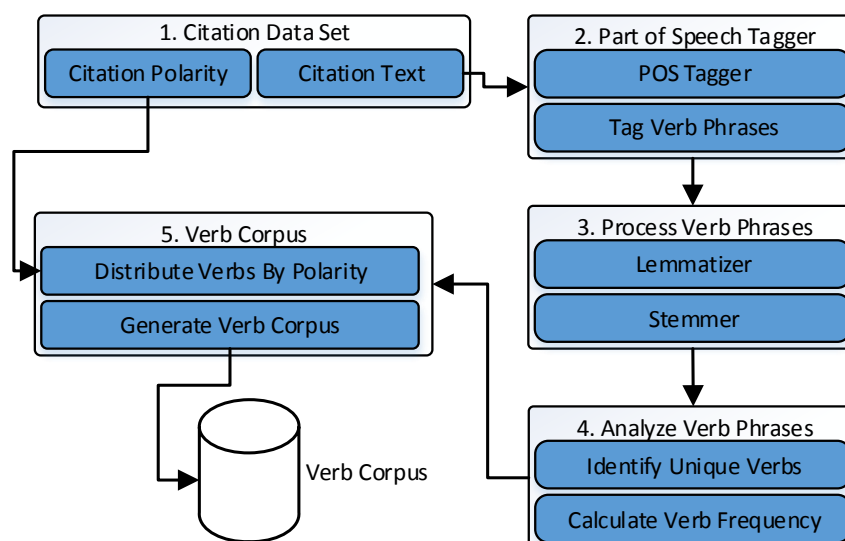


Figure 2: Methodology derived from Hyland's (2000) framework

3.1. Citation Data Set

A variety of datasets are publicly available for content-based citation analysis. In our experiments, we have used the most common dataset, ACL Anthology Network (AAN) (Radev, Muthukrishnan, Qazvinian, & Abu-Jbara, 2013). It is a comprehensive, manually designed dataset developed using papers published by ACL and Computational Linguistics journals over a period of four decades. Athar (2014) annotated the data of more than 8700 citation sentences with *Citing Paper ID*, *Cited Paper ID*, *Citation Text* and *Sentiment Polarity*. This new annotated dataset contains citation texts, annotated with *Positive*, *Negative* and *Neutral* sentiment polarity. Therefore, these two elements, citation text and citation polarity become our input as a selected dataset.

3.2. Part of Speech Tagger

POS Tagger marks a word in a text (corpus) as conforming to a part of speech, established using both the definition and the context. Context of a word defines the relationship of word with its adjacent and related words in a phrase, sentence, or paragraph. A simplified version of this is common practice to teach school-age children for the identification of words as *nouns*, *verbs*, *adjectives*, *adverbs*, etc. There are almost 34 tags for various part-of-speech used by POS Tagger. These tags are distributed among Preposition, Adjective, Noun, Verb, Adverb, Prepositional Phrase etc. However, in our experiment, we are only concerned with Verb phrases present in our selected dataset. Therefore, each citation text of AAN dataset was read by POS Tagger and the verbs were tagged in 6 different verb phrases. These verb phrase tags and their meaning are shown in Table 2.

Table 2: Verb Tags and Their Meanings

No.	Verb Tag	Meaning
1	VB	Base Form
2	VBD	Past Tense
3	VBG	Gerund or Present Participle
4	VBN	Past Participle
5	VBP	Non-3 rd Person Singular Present
6	VBZ	3 rd Person Singular Present

3.3. Process Verb Phrases

After the selection of dataset, each citation text is pre-processed before analyzing. Two common methods to prepare the text are: stemming and lemmatization. Both processes work in harmony and reduce the inflectional forms of each word into a common base or root as shown in Table 3. These two methods perform different tasks. Following example explains how these processes work. A complete set of 8700 citation texts with tagged verb phrases is pre-processed in a similar fashion with stemming and lemmatization alternatively.

Table 3: Stemming vs. Lemmatization

Form	Stemming		Lemmatization	
	Suffix	Stem	Morphological Info	Lemma
studies	-es	studi	Third person, singular number, present tense of verb <i>study</i>	study
studying	-ing	study	Gerund of the verb <i>study</i>	study

3.4. Analyze Verb Phrases

Once the tagging is complete, a simple algorithm finds the unique verbs and calculates their frequency (No of times a unique verb phrase used in the complete dataset). If a verb phrase has high frequency, it means that it is commonly used verb while citing, thus carries more weight in the corpus.

3.5. Corpus Generation

The selected dataset contains citation texts available in all 3 sentiment polarities that are *Positive*, *Negative* and *Neutral*. Therefore, in order to generate corpus of reporting verbs, all the extracted verbs are analyzed against the sentiment polarity of their parent citation texts. Verbs appearing in citation texts with *Positive* or *Negative* sentiment are more assertive in nature, rather than the verbs that appear in citation texts with *Neutral* sentiment. After finding the distribution of each verb in all three sentiments, the verb corpus is populated with the findings.

4. Experiments and Results

ACL Anthology Network AAN dataset houses 8700 citation texts. The verbs were extracted through the techniques of POS Tagging, Lemmatization and Stemming and then they were placed in the excel sheet along with their frequencies. The unique verb tokens taken out from citations came to be 389 with the verb ‘use’ on the top with frequency of 3425. The second most common verb is ‘train’ with 1285 frequency and ‘tag’ as third most common verb with frequency of 792. Figure 3 shows a word cloud of 389 verbs whereas Figure 4 shows the verbs with frequency of greater than 200.



Figure 3: High frequency verbs - word cloud

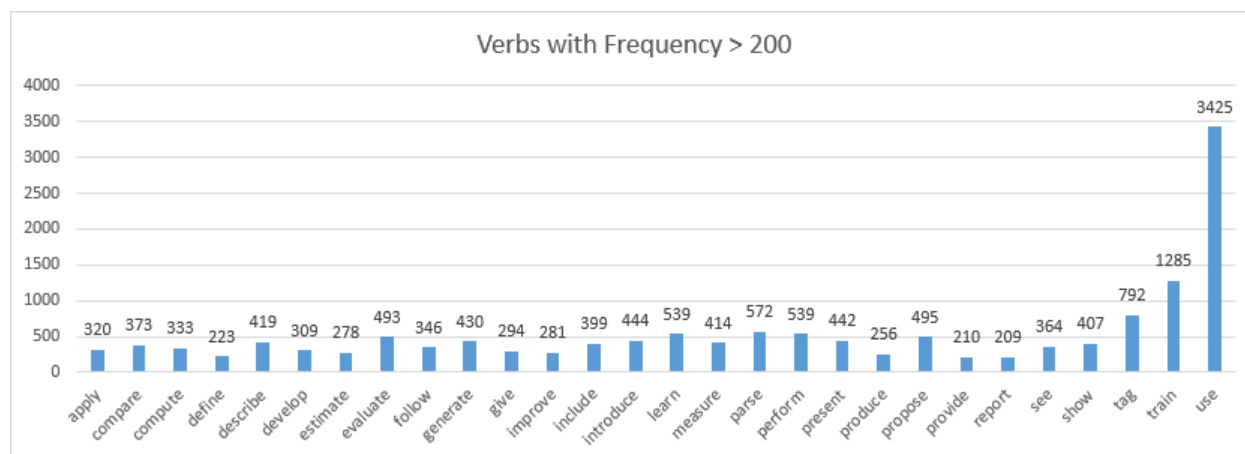


Figure 4: Verbs with frequency greater than 200

In order to find the accuracy of our experiment, 3 random sets of 100 citation texts were selected. These three sets were annotated for verbs manually by experts in English Language. After careful comparison of manual and automated extracted verbs, a confusion matrix was generated for *True Positive*, *True Negative*, *False Positive* and *False Negative*. The findings are shown in Table 4 and the average accuracy using the matrix came to be 88.25%.

Table 4: Confusion matrix

	Positive	Negative
True	303	0
False	18	22

After finding the frequency of reporting verbs in complete dataset, next step is analyzing the distribution of these verbs in different sentiments. Using the sentiment polarity information available in selected dataset, each verb frequency was checked in citation texts with ‘Positive’, ‘Negative’ and ‘Neutral’ sentiments. The dataset contains 829 citation texts with ‘Positive’ sentiment, 280 with ‘Negative’ sentiment and 7627 with ‘Neutral’ sentiment. In order to find the frequency percentage of reporting verbs in their respective sentiment, formula shown in Equation 1 is used and the resultant distribution for top 20 verbs is shown in Figure 5.

$$\frac{\sum \text{VerbOccurencesinSelectedSentiment} \times 100}{\sum \text{NoofVerbOccurencesinAll 3 Sentiments}}$$

(Equation1)

From the resultant frequency percentage distribution, it is evident most of the verbs exhibit correlation in all 3 sentiments but some of the verbs show deviant behavior, with higher number of percentages in *Positive* or *Negative* sentiment but lower in *Neutral* sentiment as compared to other verbs. A few of such deviant verbs are ‘Achieve’, ‘Improve’, ‘Perform’ and ‘Show’. Therefore, such verbs carry different weight when used in different sentiments. In the end, we have assigned each verb a weight in the resultant reporting verb corpus based on their usage in different sentiments.

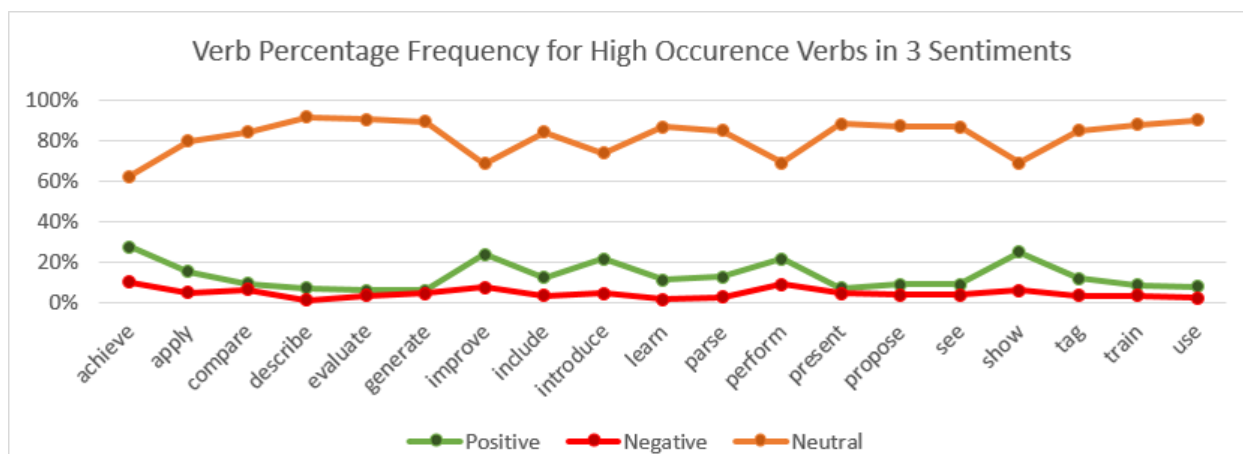


Figure 5: Verbs percentage frequency for high occurrence verbs in 3 sentiments

5. Conclusion

We started our research work to identify a certain number of reporting verbs that are used in academic and scientific writings. In this paper, we used Natural Language Processing techniques and toolkits to tag part-of-speech ‘verbs’ from 8700 citation texts from AAN Dataset. We also used sentiment analysis to distribute and analyze the extracted verbs into ‘*Positive*’, ‘*Negative*’ and ‘*Neutral*’ sentiments. Using the frequency of occurrence of each verb in all three sentiment polarity contexts, a reporting verb corpus was formulated. However, there are possibilities to further analyze these verbs using English Language Verb Classes and Alternations such as Verbs with Predicate Behavior. Only after the mapping of our reporting verb corpus with English Language Verb Classes, it is possible to automatically infer the motivation of a citation. In future, a mapping dictionary will be developed for automatic extraction of a cognitive relationship between the citing and the cited paper.

References

- Adhvaryu, N., & Balani, P. (2015). Survey: Part-of-speech tagging in NLP. *International Journal of Research in Advent Technology*, 102–107.
- Athar, A. (2014). *Sentiment analysis of scientific citations* (856). Retrieved from University of Cambridge Computer Laboratory website: <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-856.pdf>
- Athar, A., & Teufel, S. (2012). Context-enhanced citation sentiment detection. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 597–601). Montreal, Canada: Association for Computational Linguistics. Retrieved from <https://www.aclweb.org/anthology/N12-1073>
- Bach, N. X., Cuong, L. A., Ha, N. V., & Binh, N. N. (2008). Transformation rule learning without rule templates: A case study in part of speech tagging. In *2008 International Conference on Advanced Language Processing and Web Information Technology* (pp. 9–14). Dalian Liaoning, China: IEEE.
- Banker, S., & Patel, R. (2016). A brief review of sentiment analysis methods. *International Journal of Information Sciences and Techniques (IJIST)*, 6(1/2), 89–95. doi:10.5121/ijist.2016.6210
- Bird, S., & Loper, E. (2004). NLTK: The natural language toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions* (pp. 214–217). Barcelona, Spain: Association for Computational Linguistics. doi: 10.3115/1118108.1118117
- Brandt, M.D., Loftsson H., Sigurþórsson H., & Tyers F.M. (2011, May). Apertium-IceNLP: A rule-based Icelandic to English machine translation system. In *Proceedings of Annual Conference of the European Association of Machine Translation* (pp. 217–224). [Presentation]. Leuven, Belgium: European Association for Machine Translation. Retrieved from <http://www.mt-archive.info/EAMT-2011-TOC.htm>
- Cunningham, H., Wilks, Y., & Gaizauskas, R. J. (1996). GATE - A general architecture for text engineering. In *Proceedings of 16th International Conference on Computational Linguistics* (pp. 1057–1060). Copenhagen, Denmark: Center for Sprogteknologi.
- Garg, N., Goyal, V. & Preet, S. (2012). Rule based Hindi part of speech tagger. In *Proceedings of COLING 2012: Demonstration Papers* (pp. 163–174). Mumbai, India.
- Graliński, F., Jassem, K., & Junczys-Dowmunt, M. (2013). PSI-toolkit: A natural language processing pipeline. In A. Przepiórkowski, M. Piasecki, K. Jassem, & P. Fuglewicz (Eds.), *Computational linguistics: Applications* (pp. 27–39). Berlin, Germany: Springer.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Delhi, India: Pearson Education, Inc.
- Kumar, D., & Josan, G. S. (2010). Part of speech taggers for morphologically rich Indian languages: A survey. *International Journal of Computer Applications*, 6(5), 1–9.
- Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In C. E. Brodley, & A. P. Danyluk (Eds.), *Proceedings of the Eighteenth International Conference on Machine Learning (ICML'01)* (pp. 282–289). San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Ling, M. H. T. (2006). An anthological review of research utilizing MontyLingua, a python-based end-to-end text processor. *The Python Papers*, 1(1), 5–13. Retrieved from [arXiv:cs/0611113](https://arxiv.org/abs/cs/0611113)

- Loftsson, H., & Rögnvaldsson, E. (2007). IceNLP: A natural language processing toolkit for Icelandic. In *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association* (pp. 1533–1536). Antwerp, Belgium.
- Malin, M. V. (1968). The science citation index: A new concept in indexing. *Library trends*, 16, 374–387.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ: Computer Horizons, Inc.
- Petasis, G., Karkaletsis, V., Paliouras, G., Androutopoulos, I., & Spyropoulos, C. D. (2002). Ellogon: A new text engineering platform. *CoRR*, 0205017, 1–7.
- Radev D. R., Muthukrishnan P., Qazvinian V., & Abu-Jbara, A. (2013). The ACL anthology network corpus. *Language Resources and Evaluation*, 47(4), 919–944. doi: 10.1007/s10579-012-9211-2
- Vadim, B., & Garr, B. (2013). Machine Translation: The Enterprise Point of View. In A. Neustein, & J. Markowitz (Eds), *Where Humans Meet Machines – Innovative Solutions for Knotty Natural-Language Problems*. New York, NY: Springer. doi: 10.1007/978-1-4614-6934-6_3
- Xu, J., Zhang, Y., Wu, Y., Wang, J., Dong, X., & Xu, H. (2015). Citation sentiment analysis in clinical trial papers. *AMIA ... Annual Symposium proceedings. AMIA Symposium, 2015*, 1334–1341.
- Youzhi, Z. (2009). Research and implementation of part-of-speech tagging based on Hidden Markov Model. *2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications (PACIIA)*, 2, 26–29. doi:10.1109/PACIIA.2009.5406648
- Zhang, X., Huang, H. & Liang, Z. (2009). The application of CRFs in part-of-speech tagging. *2009 International Conference on Intelligent Human-Machine Systems and Cybernetics* (pp. 347–350). Hangzhou, China: IEEE. doi: 10.1109/IHMSC.2009.210
- Ziman, J. M. (1968). *Public knowledge: An essay concerning the social dimension of science*. Cambridge, England: Cambridge University Press.

Study of Gendered Language in Blogs: A Corpus-Based Analysis of Fixed Phrases

Muhammad Yousaf

Lecturer, National University of Modern Languages, Islamabad

Ijaz Ali Khan

Lecturer, Government College Lahore, Sawabi

Azhar Habib

Lecturer, National University of Modern Languages, Islamabad

Abstract

The study explored fixed phrase fragments from the corpus of Pakistani blogs written by male and female bloggers in order to compare and contrast the language used by them. The fixed structures were analyzed and categorized with the help of the framework provided by Biber and Conrad (1999) and later adapted by Yousaf (2019) for Pakistani context. The results proved that both male and female writers use fixed phrases and rely heavily on them. Similarly, the study revealed that male and female writers used different percentages of fixed phrases belonging to different categories. Male bloggers used the highest percentage of Prepositional Phrase Fragments which is actually in line with the international norms as many studies highlight a high percentage of ‘Prepositional Phrase Fragments’ in different genres (see studies such as Hyland, 2008; Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004). On the other hand, if we look at the highest percentage of fixed structures by female bloggers, Noun Phrase Fragments are on the top with 22.41% of the total which is the second highest category in the blogs by male bloggers, although the difference is not very high yet the highest percentage of these structures in blogs by female bloggers is making it highly noticeable. The second highest category by female bloggers is Anticipatory It, That/There/To Fragments which is 18.96%. This category shows the stance of the writers whether they want to put their viewpoint directly or they keep the actor in the background. In this sense, female bloggers are more unbiased and present their view without focusing on the actor rather, they foregrounding the act. Hence, it can be concluded that fixed phrases in blogs have variation because of the gender of the bloggers.

Keywords: fixed phrases, gendered language, phrase fragments.

1. Introduction

The study of fixed structures has attracted many linguists and researchers. Fixed structures are word combinations which follow each other in a more recurrent manner which may not be expected by chance. Allen (2009) argues that these structures of language signify the register specific nature of diction and perform a variety of discourse functions. The same is indorsed by Haswell and Tedesco (1991), who states that the application of these fixed expressions may articulate the proficiency level of the language users. In non-native English speaking contexts such as Pakistan, the language learners need to know the functions and structures of these fixed expressions in order to improve their language skills. Furthermore, there are advantages in exploring these structures to better help the writers acquire the specific rhetorical practices of using language appropriately (Hyland, 2008).

The studies conducted mostly on the use of lexical bundles in academic writing are Biber and Conrad (1999), Biber, Conrad, and Cortes (2004), and Biber and Barbieri (2007). The researchers delimited their studies to the use of lexical bundles in a wide range of spoken and written university registers. Although corpus-based investigations of natural language data have established the existence of frequently recurring lexical chunks in texts (Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004), the major controversy was to define these structures in a manner fit for comprehension. This very phenomenon resulted in a number of taxonomies: ‘lexical bundles’ (Biber & Conrad, 1999), ‘prefabs’ or ‘lexical phrases’ (Nattinger & DeCarrico, 1992), ‘formulaic sequences’ (Schmitt & Carter, 2004), ‘sentence stems and ‘clusters’ (Scott, 1997).

The common point seems to be that lexical bundles are text-generated; with any combination of words occurring together rather frequently in a given text (Biber, 2006). Similar to this, the fixed phrases or structures may also be identified as chunks signifying gendered biases. This has been mentioned by Tannen (1994), who is of the view that gender and discourse have an indirect relationship. Ochs (1992) has also endorsed this view by saying that ways of speaking are associated with stances of women or men in a particular situation. The researchers, therefore, took more interest in not only identifying differences between the speech of male and female but also underlining the role of language in shaping social identities. Keeping the above discussion in mind, it can be said that majority of the work to view fixity has been focused on the academic genres, whereas, lesser focus has been given to news blogs. It is important to view the fixity of language, prevailing in academic genres, in news blogs too so that it can be found out how far the language users rely on prefabricated structures and fixed phrases in non-academic language use, such as news blogs. Moreover, the comparison of the fixed language used by male and female bloggers would add an additional dimension of gendered language to the study. Hence, this study is an attempt to compare language of Pakistani male and female bloggers with reference to use of fixed phrases. Moreover, the study also undertakes to show similarity/difference of fixed phrases in news blogs and academic language by comparing the results with some of the studies in academic discourse.

This paper investigates the use of ‘fixed phrases’ in blog writing on the basis of gender. Fixed phrases are also known as natural or fixed occurrences of lexical items, sometimes in the form of collocations, clusters or in the form of lexical bundles. The aim was to compare the use of these phrases by male and female bloggers in a corpus of written news blogs. For this purpose, the occurrence and distribution of different fixed structures across the texts were investigated. The corpus consists of 17000 words for which different blogs were collected and they were analyzed with the help of a computer program AntConc, 3.2.1. (Anthony, 2007). The findings of the study may be particularly useful for translators, News analysts and News editors.

2. Literature Review

Language plays an important role in human life for it serves many functions, for instance, to give information, to deliver messages, to express feelings, to win over people to do something, to make them believe, to express biases in terms of gender, caste and group as well as to share one’s inner emotions. Similarly, the language used in media does not only tell us about the behavior of media but also language as a tool of communication. The gender biases and differences are therefore reflected in the form of fixed structures, other than the themes selected by both male and female blog writers.

The corpus linguistic studies conducted in this context were mostly on the use of lexical bundles in academic writing (Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004; & Biber & Barbieri, 2007). These studies are generally about the use of these chunks of language in a wide range of spoken and written discourses. Reviewing the related literature reveals an evident gap in the previous studies. Although there are many studies on fixed structures; however, studies on these chunks in terms of gender are rather rare. Several works have been made to decide about the relationship between words, grammar and their natural co-occurrence; nevertheless, very few have been conducted to investigate the relationship between textual structures and gender-oriented positions, as an off-shoot of feminism, in freelance writing.

It has been discussed by Schiffrin (2001) in detail by saying that feminism as a thought gave rise to the problems of gender-specific discourse. The issue has also been taken into account by Thorne and Henley (1975) and by Lakoff (1975) with different themes like language and sex, difference and dominance, as well as male and female Language. The researchers got particularly interested in identifying differences between men and women's positions as individuals. Hence, they attempted to find out the role of language in creating a social identity. West and Zimmerman (1975) conducted a study on the gendered conversation. Brown (1980) and Tannen (1990) worked on the quality of language used by both male and female writers.

There are some writers like Goffman (1976), Tannen (1994), and Ochs (1992) whose works are predominantly concerned with gender and language relationship. Very recently, the study conducted by Yousaf and Khan (2019) has identified stance features in a corpus of blogs written by male and female writers. Majority of the studies, mentioned above, relied more on spoken discourse. The current is an attempt to explore the fixity of language in the blogs written by male and female bloggers. The analysis was done through AntConc 3.2.1 (Anthony, 2007), a software tool, available online.

The language of news is quite different from business and academic language. So the language used by any news media, including news blogs, has dominating aspect communication and information. According to Crystal (1969 as cited in Westin, 2002) whatever is part of the newspaper or written by a journalist cannot always be expected to be linguistically homogenous. This means that the function of the language of news media is to provide information and highlight power relations by using innovative language to serve the purpose. This makes it important to find out how far fixity of language prevails in news blogs and how far do male and female bloggers observe homogeneity/heterogeneity in the use of fixed phrases. The current study aims to answer the following research questions:

1. What are the different structural types of fixed phrases used in blogs by Pakistani male and female writers?
2. How far are the blogs by male and female writers similar or different in terms of fixed phrases?
3. Which structural types of fixed phrases are dominant in male and female blog writers?

3. Research Methodology

For the current study, the researchers compiled a corpus consisting of twenty news blogs, ten from male and ten from female bloggers each. Although the selection of the blogs was done randomly yet it was ensured that the blogs were written by Pakistani bloggers on

social and political issues so that the variation of language because of other factors, such as professional jargon and register, was minimum. Secondly, it was also ensured that the blogs selected were written by different bloggers and no two blogs by the same author were included in the corpus. It was done to keep balance in the frequency of fixed phrases, so that excessive use of fixed phrases by one author could not affect the results of the study.

As mentioned earlier AntConc 3.2.1 (Anthony, 2007) corpus analysis software was used for analyzing the current corpus. The software provides a feature of N-Gram which facilitates the selection of a number of words in a string of word and also helps in finding out the number of texts in which the string is present. This feature enabled the researchers to find out 3 and 4-word strings in the corpus which later on were analyzed and categorized. Moreover, it is to be highlighted that the minimum range for the N-Gram was kept to ensure that a fixed structure was not only used by one writer rather it was a feature of the blogs.

3.1. Theoretical Model

The study used Biber and Conrad (1999) framework for the categorization of word strings in different categories as per their structure. The model helps in dividing the structure in categories and then into different sub-categories, but this study only opts for the broader categories and the sub-categories have not been used. The adapted version of the model used by Yousaf (2019), inspired by Biber and Conrad (1999), was used which provided categories of phrase fragments as used in Pakistani context. It is also pertinent to mention here that the framework was used for academic discourse but it is appropriate to categorize different structures on the basis of their structural categories. The main contours of the model are as follows:

Table 1: Structural Categories

S. No.	Categories
1.	Noun Phrase Fragments
2.	Verb Phrase Fragments
3.	Noun + Verb Phrase Fragments
4.	Prepositional Phrase Fragments
5.	Verb/Adjective/Noun + to/that Phrase Fragments
6.	Adverbial/Adjectival Phrase Fragments
7.	Anticipatory It, That/There/To Fragments
8.	Other Expressions

4. Results and Discussion

The N-Grams found with the help of corpus analyzing software were divided into the categories provided by Yousaf (2019), and the detailed results and discussions are as follows:

4.1. Noun Phrase Fragments

Table 2 shows that there are 27 Three and Four-Word structures which are actually Noun Phrase Fragments. The difference of noun phrase fragments in male and female bloggers is very little as we can see that male bloggers use fourteen noun phrase fragments out of which thirteen are three-word strings and only one is a 4-Word string. It means the phraseology used by male Pakistani bloggers is mostly dependent on 3-Word strings. On the

other hand, if we look at female bloggers, there are 13 noun phrase fragments in total. Eleven of the thirteen are 3-word strings, whereas, only two are 4-word structures. This again shows that Pakistani Female bloggers too rely on 3-word strings. This shows that the frequency of a string of words is inversely proportional to the length of the string. Finally, if we consider the percentage of the noun phrase fragments in the selected blogs; male bloggers use 20.29% of the total fixed structures as noun phrase fragments. On the other hand, female bloggers have a higher percentage as they use 22.41% of the fixed structures as noun phrase fragments. If we compare these findings with academic discourse, the results show that the subjects of humanities rely more upon noun phrase fragments (Yousaf, 2019; Yousaf & Shehezad, 2018; Hyland, 2008). This means that the bloggers use more argument building language than relying on the facts and figures which, although having a little variation by male and female bloggers, gives a special feature to the blogs written in the Pakistani context.

Table 2: Noun Phrase Fragments

S. No.	Male Bloggers	S. No.	Female Bloggers
1	muhammad sultan shah	1	the blame game
2	part of the	2	the name of
3	the fact that	3	days left in
4	the most important	4	front of her
5	issues related to	5	me it was
6	moments of joy	6	mother in law
7	the end of	7	people are not
8	the middle of	8	the whole way
9	the society and	9	a part of
10	the world they	10	her mother in
11	a part of	11	nothing wrong with
12	middle of the	12	her mother in law
13	their own style	13	most of their time
14	the middle of the		the name of
Frequency	20.29%		22.41%

4.2. Verb Phrase Fragments

Table 3 shows that there are 17 verb phrase structures. The difference of verb phrase fragments in male and female bloggers is very high as compared to noun phrase fragments as we can see that male bloggers use eleven verb phrase fragments which are all three-word strings and there is no 4-Word string. It means the phraseology used by male Pakistani bloggers is totally dependent on 3-Word strings when they employ verb phrase fragments in particular. On the other hand, if we look at female bloggers, there are six verb phrase fragments in total and all of them are 3-Word strings. This again shows that Pakistani Female bloggers too rely on 3-word strings. Finally, if we consider the percentage of the verb phrase fragments in the selected blogs; male bloggers use 15.94% of the total fixed structures as verb phrase fragments. On the other hand, female bloggers have comparatively a very low percentage as they use 10.34% of the fixed structures which are verb phrase fragments. This means that female bloggers rely more on argument building and less on action word as compared to male bloggers, whereas male bloggers rely more on action word and phrases to make their writing more emphatic.

Table 3: Verb Phrase Fragments

S. No.	Male Bloggers	S. No.	Female Bloggers
1	needs to be	1	is going to
2	seem to be	2	going to be
3	be considered as	3	has given us
4	be taken to	4	was a sight
5	be used to	5	are going through
6	been considered as	6	are the best
7	is no doubt		
8	may well be		
9	share in the		
10	should not be		
11	used to have		
Frequency	15.94%		10.34%

4.3. Noun (/Pronoun) + Verb Phrase Fragments

Table 4 shows that there are 15 noun/pronoun + verb phrase structures. The difference of these fragments in male and female bloggers is not very high as we can see that male bloggers use seven noun/pronoun + verb phrase fragments which are all three-word strings and there is no 4-Word string. It means the phraseology used by male Pakistani bloggers heavily relies on 3-Word strings when they employ noun/pronoun + verb phrase fragments. On the other hand, if we look at female bloggers there are eight noun/pronoun + verb phrase fragments in total and all of them are 3-Word strings. This again shows that Pakistani Female bloggers too rely on 3-word strings. Finally, if we consider the percentage of the noun/pronoun + verb phrase structures in the selected blogs, male bloggers use 10.14% of the total fixed structures as noun/pronoun + verb phrase structures. On the other hand, female bloggers have comparatively very low percentage as they use 13.79% of the fixed structures which are noun/pronoun + verb phrase fragments. This means that the female bloggers use more noun/pronoun + verb phrase structures as compared to male bloggers, whereas, male bloggers rely more on action words and phrases to make their writing more emphatic. The lexical items show that the female bloggers use more “I” and “WE” pronoun to show their presence in the blogs, whereas male bloggers include “YOU” and “HE” in their blogs to engage the audience.

Table 4: Noun (/Pronoun) + Verb Phrase Fragments

S. No.	Male Bloggers	S. No.	Female Bloggers
1	he further explained	1	i am not
2	i am not	2	we have to
3	it was a	3	i am the
4	we used to	4	i have mentioned
5	you can not	5	this was the
6	you have a	6	the situation is
7	how women are	7	no one can
		8	how it is
Frequency	10.14%		13.79%

4.4. Prepositional Phrase Fragments

Table 5 shows that there are 26 Prepositional Phrase Fragments in the corpus. The difference of these fragments in male and female bloggers is reasonably high as we can see that male bloggers use seventeen Prepositional Phrase Fragments, fourteen of which are 3-Word strings, two are 4-Word strings and one is a 5-word string. It means the phraseology used by male Pakistani bloggers is not totally dependent on 3-Word strings when they employ Prepositional Phrase Fragments. On the other hand, if we look at female bloggers, there are ten Prepositional Phrase Fragments in total, eight of which are 3-Word strings and the remaining two are 4-word strings. This again shows that Pakistani Female bloggers do not only rely on 3-word strings. Finally, if we consider the percentage of the Prepositional Phrase structures in the selected blogs, male bloggers use 24.63% of the total fixed structures as Prepositional phrase fragments. On the other hand, female bloggers have comparatively a very low percentage as they use 17.24% of the fixed structures which are Prepositional Phrase Fragments. This means that female bloggers use less Prepositional Phrase structures as compared to male bloggers. This shows that male bloggers rely more on prepositional fragments and phrases which is similar to the use of phrasal structures in academic writing (Yousaf, 2019; Yousaf & Shehzad, 2018; Hyland, 2008; Kwary, Ratri, & Artha, 2017). The lexical items show that the male bloggers use a variety of prepositional heads as compared to the female bloggers which may reflect their linguistic competence as well. In these structures, some of the structures are apparently adverbial fragments in their use but structurally they are prepositional fragments, one such example is ‘in front of the’.

Table 5: Prepositional Phrase Fragments

S. No.	Male Bloggers	S. No.	Female Bloggers
1	in our society	1	in front of
2	in our lives	2	in my heart
3	in the education	3	in the name
4	in the middle	4	of the wrong
5	in their own	5	of their time
6	of the fact	6	out of it
7	of the house	7	out of which
8	of the world	8	by the time
9	of pakistan is	9	in front of her
10	over the world	10	in the name of
11	at the time		
12	from all corners		
13	over the country		
14	up in the		
15	in the middle of		
16	of the fact that		
17	in the middle of the		
Frequency	24.63%		17.24%

4.5. Verb/Adjective/Noun + to/that Phrase Fragments

The table 6 below shows that both male and female bloggers do not use these phrasal structures in their writing.

Table 6: Verb/Adjective/Noun + to/that Phrase Fragments

Male Bloggers	Female Bloggers
00.00	00.00

4.6. Adverbial/Adjectival Phrase Fragments

Table 7 shows that there are 18 Adverbial/Adjectival Phrase Fragments in the corpus. The difference of these fragments in male and female bloggers is not high as we can see that male bloggers use ten Adverbial/Adjectival Phrase Fragments, eight of which are 3-Word strings, and two are 4-Word strings. It means the phraseology used by male Pakistani bloggers are both 3 and 4-Word strings but the ratio of 3-Word strings is very high when they employ Adverbial/Adjectival Phrase Fragments. On the other hand, if we look at female bloggers, there are seven Adverbial/Adjectival Phrase Fragments in total, six of which are 3-Word strings and the remaining one is a 4-word string. This again shows that Pakistani Female bloggers do not only rely on 3-word strings. Finally, if we consider the percentage of the Adverbial/Adjectival Phrase Fragments in the selected blogs, male bloggers use 14.49% of the total fixed structures as Adverbial/Adjectival Phrase Fragments. On the other hand, female bloggers have comparatively a very low percentage as they use 12.06% of the fixed structures which are Adverbial/Adjectival Phrase Fragments but the difference is very low in terms of percentage. The lexical items show that the male bloggers use a variety of Adverbial/Adjectival Phrase Fragments as compared to the female bloggers which may reflect their linguistic competence as well.

Table 7: Adverbial/Adjectival Phrase Fragments

S. No.	Male Bloggers	S. No.	Female Bloggers
1	all over the	1	already on the
2	throughout the world	2	most of the
3	a lot of	3	most of their
4	all of us	4	much so that
5	because there is	5	more and more
6	due to the	6	so much so
7	during that time	7	so much so that
8	even in the		
9	all over the world		
10	all over the country		
Frequency	14.49%		12.06%

4.7. Anticipatory It, That/There/To Fragments

Table 8 shows that there are 28 Anticipatory It, That/There/To Fragments in the corpus of Pakistani blogs. The difference of these fragments in male and female bloggers is reasonably high as we can see that male bloggers use seven Anticipatory It, That/There/To Fragments, all of which are 3-Word strings. It means the phraseology used by male Pakistani bloggers is totally dependent on 3-Word strings when they employ Anticipatory It, That/There/To Fragments. On the other hand, if we look at female bloggers, there are eleven Anticipatory It, That/There/To Fragments, and all of them are 3-Word strings. This again shows that Pakistani Female bloggers only rely on 3-word strings when they use Anticipatory It, That/There/To Fragments. Finally, if we consider the percentage of the structures in the selected blogs, male bloggers use 11.11% of the total fixed structures as Prepositional phrase

fragments. On the other hand, female bloggers have comparatively a very high percentage as they use 18.96% of the fixed structures which are Anticipatory It, That/There/To Fragments. This means that female bloggers use more Anticipatory It, That/There/To Fragments as compared to male bloggers. These structures are used to show the absence of the writer from the text and make the text unbiased (Yousaf, 2019; Hyland, 2008). This shows that female bloggers use more Anticipatory It, That/There/To Fragments and phrases which is similar to the use of phrasal structures in academic writing while dealing with science subjects where reporting and description is higher than that of argument building (Yousaf, 2019; Yousaf & Shehzad, 2018; Hyland, 2008; Kwary, Ratri, & Artha, 2017). This means female bloggers rely more on description and reporting and they want to keep the source in the background.

Table 8: Anticipatory It, That/There/To Fragments

S. No.	Male Bloggers	S. No.	Female Bloggers
1	there is a	1	it was the
2	there is no	2	that i am
3	that women are	3	there is no
4	that you are	4	it was a
5	that you can	5	that she has
6	there was no	6	that we have
7	to be taken	7	there is nothing
8		8	there was a
9		9	to live in
10		10	to enjoy the
11		11	to make a
Frequency	11.11%		18.96%

4.8. Other Expressions

Table 9 shows that there are only 6 Other Expression Fragments in the corpus. There is no difference between these fragments in male and female bloggers as we can see that both male and female bloggers use three Other Expressions each, and all are 3-Word strings. The majority of the Other Expressions are conjunction fragments with only one exception of a Clause used by male bloggers. ‘I was thinking’ is a complete clause that is used in a majority in some of the genres. Finally, if we consider the percentage of the Other Expressions in the selected blogs, male bloggers use 4.34% of the total fixed structures as Other Expressions. On the other hand, female bloggers have comparatively a high percentage as they use 5.17% of the fixed structures which are Other Expressions. This means that male and female bloggers use almost the same percentage of these structures.

Table 9: Other expressions

S. No.	Male Bloggers	S. No.	Female Bloggers
1	and they were	1	and all other
2	and tried to	2	and his family
3	i was thinking	3	but there was
Frequency	4.34%		5.17%

4.9. Comparison of the Structural Categories

The table below shows the difference in use of different structural categories by male and female bloggers in their blogs. The highest percentage of ‘Prepositional Phrase

Fragments' is used by the male bloggers in the Pakistani context which is actually in line with the international norms as many studies (Biber & Conrad, 1999; Hyland, 2008, Biber, Conrad, & Cortes, 2004; Kwary, Ratri, & Artha, 2017; Yousaf, 2019) highlight the high percentage of 'Prepositional Phrase Fragments' in different genres. This category is second highest in the blogs written by female bloggers and the difference is significant as male bloggers use 24.63% of their fixed structures as Prepositional Fragments, whereas female bloggers use 17.24% of their fixed structures as Prepositional Phrase Fragments. On the other hand, if we look at the highest percentage of fixed structures by female bloggers, they use 22.41% of the structures as Noun Phrase Fragments which is the second highest category in the blogs by male bloggers, although the difference is not very high but still the highest percentage of these structures in blogs by female bloggers is making it significant. The second highest category by female bloggers is Anticipatory It, That/There/To Fragments which is 18.96% and in the blogs by male writers it is 11.11%. As stated earlier, this category shows the stance of the writers whether they want to put their viewpoint directly or they keep the actor in the background. In this sense, female bloggers are more unbiased and present their view without focusing on the actor, rather they foregrounding the act. Similarly, apart from 'Verb/Adjective/Noun + to/that Phrase Fragments', more or less all the categories have a difference of percentages which show that the language of blogs in Pakistani context is gendered language when it is viewed from the lens of fixed phrases.

Table 10: Comparison of Structural Categories

S. No.	Structural Categories	Percentage	
		Male Bloggers	Female Bloggers
1	Noun Phrase Fragments	20.29%	22.41%
2	Verb Phrase Fragments	15.94%	10.34%
3	Noun + Verb Phrase Fragments	10.14%	13.79%
4	Prepositional Phrase Fragments	24.63%	17.24%
5	Verb/Adjective/Noun + to/that Phrase Fragments	00%	00%
6	Adverbial/Adjectival Phrase Fragments	14.49%	12.06%
7	Anticipatory It, That/There/To Fragments	11.11%	18.96%
8	Other Expressions	04.34%	05.17%

5. Conclusion

The study explored fixed phrases from the corpus of Pakistani blogs written by male and female bloggers. The main aim of the study was to compare the language of male and female bloggers with reference to fixed phrase fragments. For this purpose, three research questions were formulated. The first question was about the use of different structural types of fixed phrases present in the corpus. The fixed structures were analyzed and categorized with the help of framework adapted by Yousaf (2019) and the results confirmed that apart

from ‘Verb/Adjective/Noun + to/that Phrase Fragments’ all other structural categories are present in the blogs both by male and female writers. The second question was focused on the similarity and difference of the blogs with respect to fixed phrases. The results proved that blogs both by male and female writers use the fixed phrase and rely heavily on them. The third question attempted to explore the dominant categories of fixed phrases. The study proved that male and female writers used different percentages of fixed phrases belonging to different categories as male bloggers used the highest percentage of ‘Prepositional Phrase Fragments’ in the Pakistani context which is similar to Hyland (2008), Kwary, Ratri, and Artha (2017) and Yousaf (2019) highlighting the high percentage of ‘Prepositional Phrase Fragments’ in different genres. This category is second highest in the blogs written by female bloggers and the difference is significant as male bloggers use 24.63% of their fixed structures as Prepositional Fragments whereas female bloggers use 17.24% of their fixed structures as prepositional Phrase Fragments. On the other hand, the highest percentage of fixed structures by female bloggers, they use 22.41% of the structures as Noun Phrase Fragments which is the second highest category in the blogs by male bloggers, although the difference is not very high but still the highest percentage of these structures in blogs by female bloggers is making it significant to highlight the difference of language use. The second highest category by female bloggers is Anticipatory It, That/There/To Fragments which is 18.96% and in the blogs by male writers it is 11.11%. As stated earlier, this category shows the stance of the writer whether they want to put their viewpoint directly or they keep the actor in the background. In this sense, female bloggers are more unbiased and present their view without focusing on the actor; instead they are foregrounding the act. Hence, it can be concluded that although blog writing has its own phraseology yet there are differences because of multiple factors, the gender of the bloggers being one of them. The study suggests that phraseology of the blogs should be explored to find out variations in it due to category of blogs (Sports, Science, Education, Politics, Social issues etc.).

References

- Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 1, 105–127.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286. doi:10.1016/j.esp.2006.08.003
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–189). Amsterdam, Netherlands: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. doi: 10.1093/applin/25.3.371
- Brown, P. (1980). How and why women are more polite: Some evidence from a Mayan community. In McConnell-Ginet, S., Borker, R. & Furman, N. (Eds.), *Women and language in literature and society*, (pp. 111–135). New York, NY: Praeger Publishers.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. doi: 10.1016/j.esp.2003.12.001
- Goffman, E. (1976). Gender display. In *Gender advertisements* (pp. 1–9). London, England: Palgrave Macmillan.
- Haswell, R. H., & Tedesco, J. E. (1991). *Gender and the evaluation of writing*. Paper presented at the Annual Meeting of National Council of Teachers of English, Seattle, WA.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. doi: 10.1016/j.esp.2007.06.001
- Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, 7(1), 131–140. doi: 10.17509/ijal.v7i1.6866
- Lakoff, G. (1975) Hedges: A study in meaning criteria and the logic of fuzzy concepts. In *Contemporary research in philosophical logic and linguistic semantics* (pp. 221–271). Dordrecht, Netherlands: Springer.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. New York, NY: Oxford University Press.
- Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (p. 335–358). Cambridge, England: Cambridge University Press.
- Schiffrin, D. (2001). Language and public memorial: ‘America’s concentration camps’. *Discourse & Society*, 12(4), 505–534. doi: 10.1177/0957926501012004005
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). Amsterdam, Netherlands: John Benjamins Publishing.
- Scott, M. (1997). PC analysis of key words—And key key words. *System*, 25(2), 233–245. doi: 10.1016/S0346-251X(97)00011-0
- Tannen, D. (1990). Rethinking power and solidarity in gender and dominance. *Annual Meeting of the Berkeley Linguistics Society*, 16(1), 519–529. doi: 10.3765/bls.v16i1.3433
- Tannen, D. (1994). *Gender and discourse*. New York, NY: Oxford University Press.

- Tannen, D. (1994). *Talking from 9 to 5: How women's and men's conversational styles affect who gets heard, who gets credit, and what gets done at work*. New York, NY: William Morrow and Company, Inc.
- Thorne, B., & Henley, N. (1975). Difference and dominance: An overview of language, gender, and society. In B. Thorne & N. Henley (Eds.), *Language and sex: Difference and dominance* (pp. 5-42). Massachusetts, MA: Newbury House Publishers.
- West, C., & Zimmerman, D. H. (1975). Small insults: A study of interruptions in cross-sex conversations between unacquainted persons. In B. Thorne, N. Henley, & C. Kramarae (Eds.), *Language, gender, and society* (pp. 102–117). Massachusetts, MA: Newbury House Publishers.
- Westin, I. (2002). *Language Change in English Newspaper Editorials*. Amsterdam, Netherlands: Rodopi.
- Yousaf, M. (2019). *A Corpus-Based Study of lexical Bundles as Building Blocks of Academic Discourse*. Unpublished Thesis. Air University, Islamabad. Pakistan.
- Yousaf, M. & Khan, I. A. (2019). Gender-Based Study of Stance Features in Blogs through Corpus Techniques. *Corporum*, 2(1), 17–30.
- Yousaf, M., & Shehzad, W. (2018). Prevalence of Prefabricated Structures in Academic Discourse: A Corpus-Based Study. *International Journal of English Linguistics*, 8(5), 297–306.

Effectiveness of Corpus in Teaching English Synonyms

Amina Shahzadi

PhD Scholar, Lecturer, University of Education, Multan

Ayesha Asghar

PhD Scholar, Lecturer, University of Agriculture, Faisalabad

Saira Javed

PhD Scholar, Lecturer, International Islamic University, Islamabad

Abstract

Synonymy, an intricate linguistics feature, is considered one of the challenging facets in teaching English as a second language. It is a tough task for English teachers to teach every synonym pair. Therefore, enabling English language Learners (ELLs) to use the existing corpus data could help them to learn synonyms by exploring and analysing the collocational and colligation patterns of synonyms in a corpus to achieve accuracy in language usage. This research aims to explore how naturally occurring discourse of corpus can be used to teach different functions of synonyms effectively. Sketch Engine (SkE) tool was used to analyse the collocation, concordance, word sketches and sketch difference of synonyms *arrive* and *reach* in British National Corpus (BNC). Moreover, the Longman Dictionary of Contemporary English (2014) was used to study different senses of synonyms *arrive* and *reach*. It is found that there are more meanings associated with *reach* than *arrive* suggesting wide occurrence and usage of *reach*. SkE also revealed that the frequency of *reach* is much higher than *arrive* in BNC. It is found that there are some similarities in grammatical categories of collocates and syntactic patterns of both the synonyms in BNC as synonyms *reach* and *arrive* occur mostly with the subject on the right and preposition on the left. However, the frequency of some of the grammatical features is different for both the synonyms and there are considerable differences in the syntactic patterns of their colligation. The results of this study reveal that analysis of a naturally occurring corpus can be used as an effective strategy to distinguish and teach the synonyms *arrive* and *reach* in ESL classroom. The study recommends the use of corpus-based analysis in language teaching which may clarify synonym misuse and enhance the vocabulary of language learners.

Keywords: *arrive*, BNC, corpus, collocation, ESL, language learning, *reach*, synonyms

1. Introduction

English as a lingua franca consists of millions of words and a number of these words called synonyms share similar senses of meanings and other semantic features. The concept of synonyms is important in English and English language learners need to learn synonyms to achieve accuracy and fluency in the English language. Teaching and learning synonyms is a complex and challenging task as it is difficult to learn every synonym and all synonyms cannot be used interchangeably in every context.

The corpus-based approach of language analysis is more reliable as it is based on empirical data to assist language teachers and researchers to find differences in the use of language. Albader (2001) affirms the suitability of corpus analysis to understand the

similarities and differences between apparently similar and interchangeable words. The corpus-based analysis helps to identify more specific criteria and suggestions for the usage of these words. Therefore, this research study aims to investigate synonym *reach* and *arrive* in British National Corpus by analyzing the major similarities and differences between these two synonyms according to their sense/meanings, connotations, collocations and grammatical patterns.

2. Literature Review

Synonymy is an important yet a complex concept in the field of lexicography and lexical semantics. In semantics, a synonym is used to refer to a relationship of similarity and sameness of meanings of different words (Jackson & Amvela, 2000). Synonyms are classified as strict or absolute and loose synonyms. In strict synonymy, two words can be used interchangeably and the substitution of one word with the other does not change the meanings and style of the message. Strict synonyms are not common in a language. In loose synonyms, there is a significant overlap in meanings of two words and they cannot be used to substitute each other in some contexts. Loose synonyms are used frequently in a language (Biber, Conrad & Reppen, 1998).

English is rich in synonyms enabling its speakers “to convey meanings more precisely and effectively” (Liu & Espino 2012, p. 198). Synonymy is a challenging area in English Language teaching and learning because of variations in their meaning and usage. Thesaurus is not enough to study lexicology and synonyms; therefore, using a corpus to identify and extract synonyms has its implications for English language teaching.

Dialect and the formality of synonymous word context, connotation and grammatical patterns are used as a criterion to distinguish between synonyms in the English language (Phoocharoensil, 2010). Speakers of different dialects of English use different words to refer to the same person, thing, or concept e.g. football and soccer, aerial and antenna, aero-plane and airplane are some of the pairs of British-American synonyms. The style or formality of the context in which synonyms occur also helps to distinguish synonymous words. One of the synonyms in a pair is used in a more formal context than the other e.g. the words ‘ammunition’, ‘comprehend’, ‘lavatory’ and ‘intoxicated’ tend to be more formal than ‘weapon’, ‘understand’, ‘toilet’ and ‘drunk’ respectively. The connotation is another important factor for synonym use because emotive meanings they express like “to smile or speak in a very unkind way that shows you have no respect for someone or something” (LDOCE, 2009, p. 1994). Similarly, a grammatical pattern is also important for using a synonym. Usually, words having same meanings do not have same syntactic pattern for instance able and capable both have same meaning but able requires infinitival phrase and capable needs prepositional phrase.

In last few decades the idea of corpus semantics has resulted in significant corpus-based studies on synonyms (Church, Gale, Hanks, & Hindle, 1991; Gries and Otani, 2010; Partington, 1998). Miller and Charles (1991) analyzed the Brown Corpus and concluded that if the two words are substitutable in the same linguistic context, they are considered synonymous in their meanings. Gilquin (2003) conducted a research to analyze the differences between get and have. Different studies have been conducted to investigate the synonym pairs e.g. ‘quake’, ‘quiver’; ‘big’, ‘large’, ‘great’; ‘actually’, ‘genuinely’, ‘really’, ‘truly’ and ‘principal’; ‘primary’, ‘chief’ and ‘main’ and ‘major’ respectively (Atkins &

Levin, 1995, Biber, Conrad, & Reppen, 1998, Liu & Espino, 2012, Liu, 2010). Castello (2014) also used corpora to study synonyms *strong* and *powerful*. He analyzed the spoken components of the Bank of English (BoE) to compare *strong* and *powerful* and found some differences in the effect these synonyms have on words in close proximity. Similarly, Gu (2017) in his corpus-based study of synonyms *obtain* and *gain* used online tools Sketch Engine, BNC Web and Just the Word. It was concluded that the synonyms *gain* and *obtain* has some differences in genre, collocation, colligation and semantic prosody.

Synonyms are analyzed by investigating their collocation and colligation patterns. Collocation is an important and challenging concept in second language teaching and learning. Collocation, a term coined by Palmer (1933), can be defined as “lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in direct syntactic relation with each other” (Heid & Gouws, 2006, p. 980). The collocation properties of lexical items help to determine their meanings (Sinclair, 1991). The words that co-occur are in proximity but not necessarily adjacent to each other and do not occur in any fixed order. According to Hoey (2005), collocations should not be confused with lexical co-occurrence as collocations are “psychological association(s) between words up to four words apart and are evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution” (p. 5).

Firth (1957) used the term collocation in theoretical linguistics but he relied on intuition for his research on the use of collocation. Corpus linguists are of the view that studying the co-occurrence in a corpus is an authentic way to identify and analyze collocates of a given word. According to Sinclair et al. (2004), to investigate the two lexical items as collocates, it is important to know the total size of the text, frequency of occurrence of both the lexical items and the number of times they co-occur in the text. Firth (1968) also introduced the term colligation to refer to grammatical categories and grammatical relations distinguishing it from collocation which emphasizes lexical inter-relations. A word’s colligations help to study its grammatical functions.

It is challenging and time consuming for lexicographers to differentiate between synonyms by using traditional methods (Kamiński, 2017). However, vocabulary studies have been revolutionized with the advent of computer technology and corpus-based evidence which facilitate “an empirical basis for determining vocabulary behavior, instead of relying on appeals to intuition or tradition” (Schmitt, 2010, p. 12). English Language Learners have confusions regarding synonyms. Only dictionaries are used to assist ELLs in learning synonyms. Therefore, the current study aims to use empirical evidence to investigate the patterns of language relating to synonyms *arrive* and *reach* to teach synonyms effectively. This paper aims to:

1. differentiate the meanings of synonyms *reach* and *arrive*
2. explore their lexical information such as connotations and collocations
3. investigate the syntactic information such as grammatical categories and grammatical patterns of *reach* and *arrive*
4. compare and contrast Oxford Learners’ Dictionary information with concordance line

3. Research Methodology

In order to explore *arrive* and *reach* synonyms, the quantitative and qualitative analysis was conducted with the help of the software toolbox Sketch Engine (SkE). It investigated *arrive* and *reach* use in naturally occurring corpus of BNC. SkE toolbox is used

in different fields e.g. lexicography, translation and language teaching as it provides easy access to many ready to use corpora. It provides concordance lines showing keywords in context helps to define lexical and structural information about the keyword. It can be used to perform different functions. In this study SkE is used to study and analyse:

1. Frequency/tokens of *arrive* and *reach*
2. Collocation
3. Word sketch
4. Sketch differences
5. Syntactic Patterns/Colligation

Furthermore, grammatical information from the Longman Dictionary of Contemporary English (2014) was consulted to identify the different meanings and senses associated with the synonym pair i.e. *reach* and *arrive*.

3.1. Data Collection

Entries from Oxford Learner's dictionary and concordance lines from BNC were used as data to study the synonyms *arrive* and *reach* as natural authentic English based on language corpora. BNC is a monolingual, synchronic and general type of corpus. BNC is a 100 million words corpus which was designed to represent the British English for later part of 20th century. It consists of 90% written and 10% spoken texts collected from wide range of disciplines during 1960 to 1990.

4. Analysis and Discussion

In this section, a detailed analysis of the synonyms *reach* and *arrive* is presented in order to reveal the different meanings of these synonyms. Their lexical information such as connotations and collocations is also analysed. Moreover, the syntactic information such as grammatical categories and grammatical patterns of *reach* and *arrive* are also investigated.

4.1. Meanings/Senses associated with *reach* and *arrive*

Longman dictionary of Contemporary English (2014) was consulted to identify different senses associated with *reach* and *arrive*.

4.1.1. Senses associated with *arrive*

Seven senses are found for verb *arrive* in Longman dictionary of Contemporary English as given below:

- i. **Get somewhere:** give me a call to let me know that you have *arrived* London safely.
- ii. **Be delivered:** (come) the card *arrived* on my birthday.
- iii. **Happen:** when her wedding day *arrived*, she was very nervous.
- iv. **Something new:** since computer *arrived*, my job has become easy.
- v. **Be born:** their baby *arrived* after mid-night.
- vi. **Arrive at a decision/solution/compromise:** after much consideration, we *arrived* at a decision.
- vii. **Success:** when he saw his name painted on the door he knew he had *arrived*.

4.1.2. Senses associated with reach

Following senses are found to be associated with *reach*:

- i. **Development:** I have *reached* the point where I can earn good salary.
- ii. **Rate/amount:** wind speed *reaching* up to 180mph.
- iii. **Agree:** the talks will continue until a conclusion was *reached*.
- iv. **Reach a target:** we hope to *reach* our one million target by next year.
- v. **Touch:** she *reached* into her bag and produced a business card.
- vi. **Length/height:** the phone lead isn't long enough to *reach* the bedroom.
- vii. **Arrive:** we *reached* Islamabad early in the morning.
- viii. **Speak to somebody:** I can probably *reach* him on his mobile.
- ix. **Be seen/ heard:** cable TV *reaches* a huge audience.
- x. **Information:** the news *reached* us in Lahore.
- xi. **Communicate:** I just can't seem to *reach* Momina anymore.
- xii. **Reach for the stars:** to convince Momina is just like to *reach* for the stars.

It is found that there are more senses associated with *reach* as compared to *arrive*. Moreover, it is also noted that *reach* has the meaning/sense of *arrive* associated with it but *arrive* has no sense of *reach* associated with it in Longman Dictionary of Contemporary English. This suggests that *reach* should have wide occurrences and usage as compared to *arrive* as it has more meanings associated with it. For further analysis of synonyms *arrive* and *reach*, BNC was explored with the help of SkE.

4.2. Frequencies of *arrive* and *reach* in BNC

Concordance helps to compare the frequencies of different words. Using the concordance hit through SkE enabled to make a comparison of frequency of occurrence of synonyms *arrive* and *reach* in British National Corpus. Table 1 reveals that the frequency of *reach* is much higher than *arrive* in BNC.

Table 1: Frequency of *arrive* and *reach* in BNC

	<i>Reach</i>	<i>Arrive</i>
Total	23,975	13,631
Per million	213.51	121.40

4.2.1. Frequencies of *arrive* and *reach* in different text types

A further comparison of the frequency of usage of the synonym pair was explored in different text types of British National Corpus. The analysis as shown in table 2 highlights that the occurrence of both the synonyms i.e. *arrive* and *reach* in written books and periodicals is significantly higher than in any other text types. Moreover, frequency of *reach* is greater than *arrive* in almost all text types which strengthens the senses found in dictionary as well.

Table 2: Comparison of frequency of *arrive/reach* in different text types of BNC

Text Type	<i>Arrive/Freq.</i>	<i>Reach/Freq.</i>
Spoken context governed	36 (3.27p/mil)	494 (4.40p/mil)
Spoken demographic	127 (1.13 p/mil)	134 (1.2 p/mil)
Written-books and periodicals	12034 (107.17 p/mil)	21,601 (192.37p/mil)
Written Misc.	852 (7.59p/mil)	1502 (13.40p/mil)
Written to be spoken	251 (2.24 p/mil)	243 (2.61p/mil)

(POS filter = all (adj., adv., conj., det., n, part., prep., pronoun, verb))

The detailed comparison of frequencies of *arrives* and *reach* in different written texts is highlighted in table 3.

Table 3: Frequency of *arrive* and *reach* in different written text types

Text Type	<i>Arrive</i>	<i>Reach</i>
Imaginative	3693(32.90 per mil)	6339 (56.50 per mil)
Informative applied science	538(4.79 per mil)	1,676 (14.90 per mil)
Informative: arts	4622(41.20 per mil)	1,117 (9.90 per mil)
Information: belief & thought	464(4.31 per mil)	690 (6.14 per mil)
Informative: commerce & finance	709(6.31 per mil)	1,358 (12.09 per mil)
Informative: leisure	2433(21.67 per mil)	4,200 (37.40 per mil)
Informative: natural and pure sciences	13631(121.40 per mil)	1,019 (9.07 per mil)
Information: social sciences	1266(11.30 per mil)	2,519 (22.40 per mil)
Informative: world affairs	2821(25.10 per mil)	4,428 (39.40 per mil)

The table 3 further explains the different categories of written text types where the near synonyms *reach* and *arrive* are used. It is clearly noted that in BNC *reach* occurred more in imaginative texts 6339 (56.50 per mil) and synonyms *arrive* is found more in arts related texts 4622 (41.20 per mil) performing informative function. Synonym *reach* is also found more in informative texts related to leisure 4200 and world affairs 4428. However, the frequency of synonym *arrive* is 2433 in leisure related text and 2821 in world affairs which is significantly less as compared to frequency of *reach* in these texts.

4.3. Collocation

The right and left collocates of *arrive* are mostly prepositions and nouns and sometimes it also collocates with adjectives and adverbs in BNC. The following table 4 illustrates the top 50 right side collocates of *arrive*.

Table 4: First 50 most frequent right side collocates of arrive

Sr. #	Collocates	Frequency	Log dice	Sr.#	Collocates	Frequency	Log dice
1	At	3,645	7.898	26	Post	27	5.414
2	Late	188	7.568	27	Hospital	32	5.414
3	Scene	103	7.401	28	Today	42	5.389
4	Home	243	7.098	29	Moscow	20	5.328
5	Safely	43	6.526	30	Hotel	25	5.306
6	Early	123	6.475	31	From	503	5.281
7	Back	292	6.462	32	On	838	5.276
8	Minutes	78	6.379	33	York	26	5.228
9	Airport	39	6.350	34	Office	34	5.210
10	London	108	6.266	35	Hong	18	5.172
11	England	78	6.199	36	Next	59	5.131
12	Britain	70	6.103	37	Soon	34	5.131
13	Paris	38	6.009	38	Kong	17	5.108
14	Later	96	6.004	39	Doorstep	15	5.108
15	Here	137	5.965	40	Hour	25	5.087
16	Morning	57	5.836	41	Ten	32	5.085
17	Destination	25	5.801	42	Afternoon	22	5.076

18	Shortly	28	5.775	43	America	23	5.068
19	Yesterday	50	5.735	44	Desk	18	5.065
20	Before	140	5.620	45	Unannounced	14	5.063
21	In	2,533	5.526	46	New	36	5.063
22	Heathrow	20	5.524	47	Country	40	5.049
23	Station	30	5.501	48	Dover	14	5.026
24	House	64	5.483	49	Evening	25	5.001
25	After	141	5.420	50	September	23	4.982

The main categories of the right collocate of *arrive* as shown in the table 4 may be classified into four groups:

1. Noun: scene, destination, home, minute, morning, airport, London, England, Britain, Paris, Heathrow, station, house, post, hospital, Moscow, Hotel, New York, office, Hong Kong, doorstep, hour, ten, afternoon, America, desk, country, door, evening, September
2. Preposition: at, in, on, after, from
3. Adverb: Late, safely, early, Later, here, shortly, yesterday, before, today, next, soon, afternoon
4. Adjective: back, unannounced, new

Table 5 illustrates the left collocates of *arrive*. As shown in table 4.5 the main categories of the left collocate of *arrive* may be classified into five groups:

1. Noun: ambulance, guests, Police, train, letter, moment, due, noun, day, refugees, morning, nights, they, who, time, minutes, troops, visitors, we, hour, hours, doctor, she, began, team, he, mail, afternoon, evening, bus
2. Preposition: when, before, until, due, since, after, by
3. Adverb: finally, soon, eventually, due, recently, later
4. Adjective: newly, first, staying, just
5. Conjunction: since

Table 5: First 50 most frequent left side collocates of *arrive*

Sr. #	Collocates	Frequency	Log dice	Sr.#	Collocates	Frequency	Log dice
1	When	704	8.765	26	who	392	5.973
2	When	1,365	7.927	27	time	312	5.969
3	Finally	129	7.594	28	minutes	56	5.901
4	Before	415	7.188	29	troops	33	5.892
5	Ambulance	66	7.174	30	By	71	5.890
6	Guests	73	7.165	31	visitors	32	5.885
7	Soon	133	7.099	32	we	500	5.877
8	Until	212	7.093	33	hour	43	5.869
9	Police	135	6.974	34	staying	29	5.864
10	Newly	51	6.687	35	hours	56	5.863
11	Train	66	6.681	36	doctor	35	5.861
12	Letter	80	6.655	37	We	175	5.857
13	Eventually	60	6.557	38	recently	43	5.815
14	Moment	94	6.501	39	just	208	5.732
15	Had	1,149	6.445	40	later	79	5.723
16	Due	72	6.368	41	she	420	5.698
17	Since	136	6.361	42	began	54	5.690
18	First	297	6.288	43	team	48	5.680
19	Waiting	52	6.226	44	he	734	5.659
20	Day	153	6.201	45	They	149	5.649

21	Refugees	34	6.183	46	mail	23	5.576
22	Morning	67	6.069	47	afternoon	31	5.571
23	After	218	6.049	48	Before	27	5.546
24	Nights	32	6.020	49	evening	36	5.27
25	They	690	6.013	50	bus	26	5.523

This means that the categories of left and right collocates of *arrive* are same as nouns, prepositions, adverb and adjectives are used as right and left collocates of *arrive*. However, there are a greater number of adverbs used as right collocates of *arrive*. Table 6 and 7 show the top 50 right and left side collocates of *reach*.

Table 6: First 50 most frequent right side collocates of reach

Sr. #	Collocates	Frequency	Log dice	Sr.#	Collocates	Frequency	Log dice
1	Agreement	523	8.900	26	pocket	84	6.666
2	Conclusion	303	8.453	27	proportions	78	6.650
3	Peak	255	8.302	28	touch	96	6.644
4	Stage	326	8.086	29	summit	79	6.629
5	Age	292	7.782	30	maturity	71	6.523
6	Point	398	7.705	31	across	127	6.452
7	Final	219	7.538	32	high	150	6.449
8	Decision	203	7.364	33	road	110	6.449
9	Level	240	7.306	34	her	832	6.442
10	Top	229	7.299	35	bottom	77	6.347
11	Out	1,042	7.287	36	per	192	6.345
12	Hand	263	7.242	37	\$	120	6.341
13	Climax	112	7.220	38	touched	65	6.314
14	Conclusions	115	7.178	39	million	115	6.287
15	End	290	7.081	40	significance	65	6.221
16	Target	116	6.984	41	compromise	59	6.217
17	Heights	92	6.932	42	before	231	6.191
18	Where	430	6.912	43	cent	138	6.181
19	Destination	88	6.848	44	his	872	6.137
20	Its	625	6.847	45	ground	84	6.136
21	Door	163	6.829	46	goal	63	6.116
22	Height	92	6.773	47	inside	76	6.101
23	Maximum	95	6.753	48	gate	55	6.078
24	semi-finals	78	6.710	49	consensus	53	6.078
25	Levels	114	6.709	50	highest	59	6.073

The main categories of the right collocate of *reach* as shown in table 6 may be classified into three groups

1. Nouns: agreement, conclusion, peak, stage, age, point, decision, top, hand, climax, conclusions, end, target, heights, destination, door, height, semi-finals, levels, pocket, proportions, summit, maturity, road, bottom, significance, compromise, cent, ground, goal, gate, consensus
2. Prepositions: out, across, per, before, inside
3. Adjectives: maximum, million, highest

Table 7: First 50 most frequent left side of reach

Sr. #	Collocates	Frequency	Log dice	Sr.#	Collocates	Frequency	Log dice
1	Agreement	470	8.746	26	She	705	6.390
2	Until	617	8.367	27	She	318	6.377
3	When	519	8.068	28	Once	133	6.351
4	Before	679	7.746	29	We	712	6.334
5	Within	426	7.713	30	Never	182	6.327
6	Beyond	196	7.546	31	Far	150	6.318
7	Easy	180	7.288	32	He	1,169	6.299
8	When	846	7.158	33	News	82	6.285
9	Failed	134	7.052	34	Soon	100	6.281
10	Finally	121	6.943	35	They	253	6.267
11	Conclusion	106	6.937	36	Yet	116	6.266
12	Had	1,585	6.875	37	As	171	6.235
13	By	175	6.854	38	Hand	123	6.146
14	Upper	97	6.811	39	agreements	53	6.031
15	They	1,230	6.805	40	Have	944	6.018
16	Conclusions	81	6.672	41	Once	62	6.012
17	Eventually	96	6.655	42	Point	120	5.975
18	Time	510	6.589	43	Before	56	5.970
19	Has	802	6.561	44	Unable	57	5.964
20	Compromise	74	6.544	45	Long	148	5.963
21	Decision	113	6.519	46	Now	261	5.961
22	Could	487	6.466	47	Expected	75	5.929
23	Can	653	6.429	48	Agreement	46	5.903
24	He	528	6.420	49	Out	397	5.895
25	Stage	101	6.395	50	been	512	5.88

In addition to above categories 3 possessive pronouns and 2 verbs touch and touched are also found in first 50 right collocates of *reach*.

The dominant categories of left collocate of *reach* may be categorized in 3 groups

1. Nouns: Agreement, conclusion, time, compromise, decision, stage, news, hand, agreements, point
2. Adverbs: when, finally, eventually, once, never, far, soon, yet, as, once, now
3. Prepositions: until, before, within, beyond, by, before, out

In addition to the above categories, pronouns such as they, he, she, we, he; adjectives such as easy, upper, unable, long and verbs such as has, have, had, can, Could, Been, failed, expected, are also dominant left collocates of *reach*.

It is noted that there are some similarities in grammatical categories of collocates of *arrive* and *reach* as prepositions, nouns and adjectives are used as right collocates of both the synonyms and noun, prepositions, and adverbs occur as left collocates of both. However, some differences are also found in the left and right collocations. Adverbs as right collocates of *arrive* and conjunctions and adjectives as left collocates of *arrive* are not found in the right and left collocation patterns of *reach* respectively.

4.4. Syntactic Patterns of *arrive* and *reach*

Word Sketch function of SkE was used to find the syntactic patterns of both the verbs *arrive* and *reach* in BNC corpus. 15 syntactic patterns of *arrive* are presented in table 8 and table 9 summarized the 16 syntactic patterns of *reach* as used in the BNC corpus.

Table 8: Syntactic patterns of *arrive*

Categories	Frequency	Score	Examples
Subject	4932	0.36	The <i>sea-mail letter</i> arrived this after-noon.
Object	1684	0.12	I was pleased to hear that <i>you</i> arrived safely home after your excursion to Southampton.
Modifier	2435	0.70	The waiter arrived <i>back</i> at the table.
And/or	225	0.20	The majority of holiday flights departs and <i>arrives</i> on schedule.
Pp_at-p	2783	0.20	It was late when they arrived <i>at</i> the old house.
Pp_in-p	2169	0.15	She arrived <i>in</i> central London.
Pp_on-p	529	0.04	The flowers arrived <i>on</i> the scene two years ago.
Pp_from-p	328	0.02	Jenny's <i>arriving</i> from London.
Pp_with-p	255	0.02	The waiter arrived <i>with</i> third bottle of valpolicella.
Pp_for-p	200	0.01	She arrived <i>for</i> the work the following Monday morning.
Pp_by-p	131	0.01	The information had arrived <i>by</i> the first post that morning.
Pp_as-p	69	0.01	She arrived <i>as</i> dusk was falling.
Arrive+Wh	67	0.00	They may well arrive <i>when</i> this is all over.
Inf-objects	346	0.03	I arrived <i>to</i> find the house empty.
arrive-adj	164	0.01	I arrived home very late that night.

Table 8 highlights the wide occurrence of *arrive* with preposition e.g. *at*, *in*, *on*, *from*, *with*, *for* etc.

Table 9: Syntactic patterns of *reach*

Categories	Frequency	Score	Examples
Subject	4994	2.90	Her <i>hand</i> reaches for the switch
Object	14285	6.20	In 1988 Guinness reached <i>agreement</i> with Financiere Agache to form a new company
Modifier	3048	0.50	The band finally reached the end of their last number
And/or	181	0.10	Anna smiled and <i>reached</i> for Carolina's hand
Pp_for-p	749	3.60	He reached <i>for</i> his glass and knocked it over
Pp_by-p	298	2.30	I liked my room, reached <i>by</i> an open staircase
Pp_in-p	274	0.60	Even when agreement was reached <i>in</i> principal with the British
Pp_into-p	233	5.00	Travis reached <i>into</i> his pocket for match
Pp_on-p	180	1.00	It may be reached <i>on</i> foot by using a track
Pp_to-p	137	0.50	He stood up, his vast frame reaching <i>to</i> the ceiling
Pp_with-p	118	0.70	agreement was reached <i>with</i> the landlord for his consent
Pp_at-p	88	0.80	The agreement, reached <i>at</i> a meeting of foreign ministers at Luxemburg
Reach+Wh	123	1.60	Eventually the agreement is reached <i>where</i> no further packing would be possible
Inf-objects	125	0.30	Lisabeth followed my eye-line and reached <i>to</i>

			<i>pull</i> down Binky’s creep
Reach+Adj-as noun	141	3.4	when he <i>reached</i> the Olympic <i>final</i> in Los Angeles
Reach-adj	143	0.40	his influence began to <i>reach further</i> into the bishopric

Tables 8 and 9 reveal that syntactic patterns of *arrive* and *reach* are similar which means they share similar syntactic behaviour. However, it was noted that the frequency of some of the grammatical features is different for both the synonyms as the frequencies of objects, modifiers and Wh- words of *reach* are higher than that of *arrive*.

4.5. Comparison of Collocation Patterns of *arrive* and *reach*

In order to compare the two synonyms i.e. *arrive* and *reach* according to their collocation context of both lexical and grammatical collocates, the Sketch difference function of SkE was used. Though the two verbs *reach* and *arrive* share number of syntactical patterns as it is shown in table 8 and 9, yet there is a considerable difference in the collocates of these syntactic patterns as it is shown in figure 1.



Figure 1: Comparison of collocation patterns of *reach/arrive* (differences as verbs)

Figure 1 shows the differences in the collocation pattern of synonyms *reach/arrive* as verbs. It is found that the collocation token of *reach* and *arrive* in the ‘and/or’ pattern is 181 and 225 respectively highlighting that ‘and/or’ are used with *arrive* more frequently than with *reach*. Words which frequently co-occur in ‘and/or’ with *reach* are pass, smile and try and are not used with *arrive* and the word say collocates with both *reach* and *arrive* equally which is 7 times. On the other hand, words which frequently collocate with verb *arrive* are do, make, take, leave, ask, depart and these words do not collocate with *reach*.

The collocation tokens for *reach* in the pattern of ‘subjects’ are 4994 and 4932 for *arrive* pointing out that there are equal numbers of words used as subject of *reach* and *arrive*. Though collocation tokens of *reach* and *arrive* are almost equal, there are more words that collocate with *arrive* than with *reach*. There are five words in subject pattern that collocate only with *reach* such as hand, temperature, total, unemployment and arm and there are more words that collocate only with *arrive* such as troop, help, crew, night, visitor, police,

ambulance, guest, waiter, cheque, coach, spring, refugee etc. and some others that can be used with both *reach* and *arrive* as verbs such as news, car, train, day, letter.

The collocation tokens for *reach* in the ‘object’ pattern are 14, 285 and collocation tokens of *arrive* in this category are 1684, indicating that there is a significant difference in collocation of *reach* and *arrive* with different ‘objects’. As it is shown in figure 1, there are more words that collocate as object of *reach* such as agreement, conclusion, stage, peak, age, level, point, end, top, height, climax, target etc. There are some words that only collocate with *arrive* such as Immigrant, hour, afternoon, morning, minute and some other words can be used with both *reach* and *arrive* such as decision and home.

The collocation tokens for *reach* and *arrive* in the ‘modifier’ pattern are 3048 and 2435 respectively which indicate that there is not much difference in the token of modified of both the verbs i.e. *reach* and *arrive*. There are some words that only collocates as modifier of *reach* such as inside, easily, almost, nearly, down, forward and the words that collocate with *arrive* in the modifier pattern are punctuality, promptly, shortly, duly, home, safely, newly, early, late. Whereas some words like recently, back, first, here can be used with both *reach* and *arrive*.

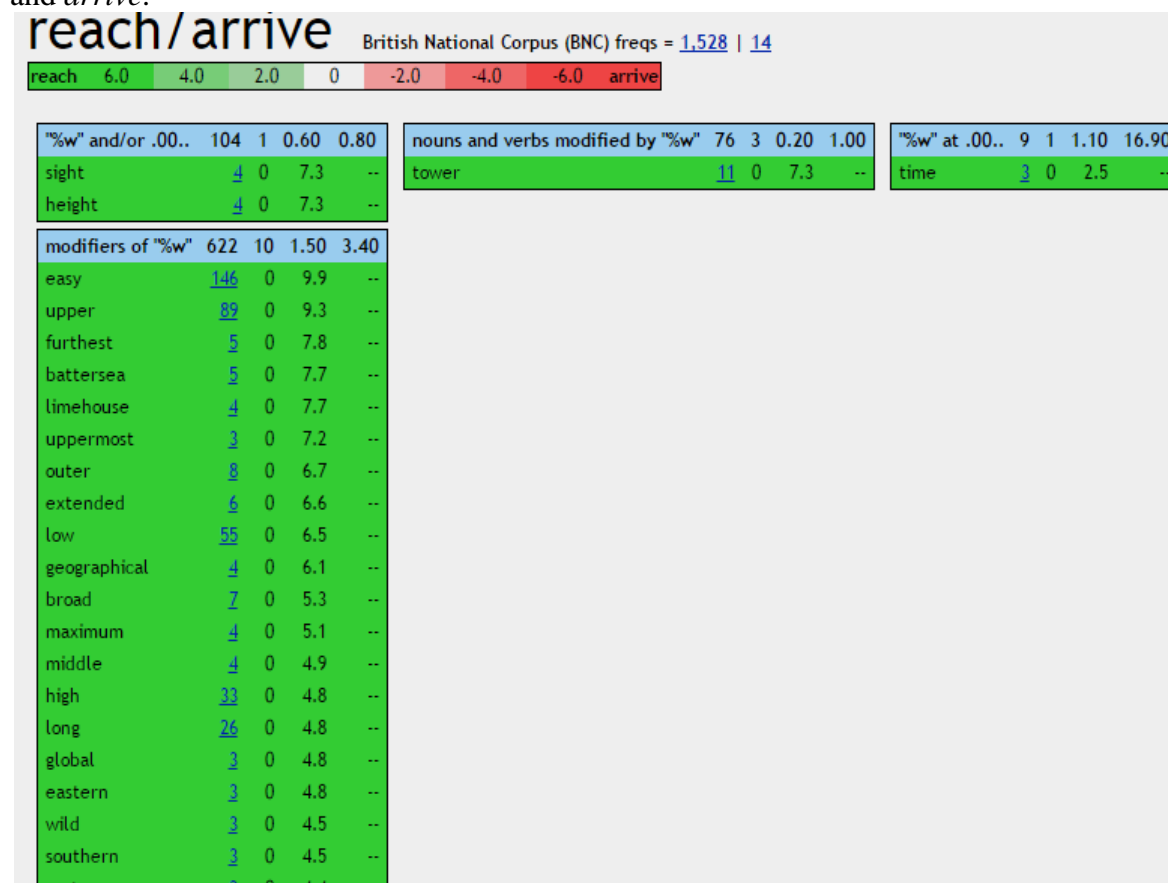


Figure 2: Comparison of collocation patterns of reach/arrive (differences as noun)

Figure 2 shows the differences in collocation pattern of *reach* and *arrive* as noun. There is significant difference between the collocations of *reach* and *arrive* in ‘and/or’ pattern as the collocation token for *reach* and *arrive* are 104 and 1 respectively. The collocation tokens for *reach* in the modifier pattern are 622 and 10 for *arrive*. The words that are used with *reach* in modifier pattern are uppermost, outer, extended, low, geographical, broad, high, long, wild, eastern etc.

4.6. Findings

The Longman Dictionary of Contemporary English (2014) shows that there are more meanings associated with *reach* than *arrive* which suggests that *reach* should have wide occurrences and usage as compared to *arrive*. Correspondingly, it is found that frequency of *reach* is much higher than *arrive* in BNC which means synonym *reach* is used more frequently than *arrive*. The occurrence of both the synonyms, i.e. *arrive* and *reach*, in written books and periodicals is significantly higher than in any other text types. However, frequency of *reach* is greater than *arrive* in almost all text types which strengthens the senses found in dictionary as well.

It is noted that there are some similarities in grammatical categories of collocates of *arrive* and *reach* as prepositions, nouns and adjectives are used as right collocates of both the synonyms and noun, prepositions, and adverbs occur as left collocates of both. However, some differences are also found in the left and right collocations. Adverbs as right collocates of *arrive* and conjunctions and adjectives as left collocates of *arrive* are not found in the right and left collocation patterns of *reach*.

Word sketch function of SkE used to analyse the syntactic pattern of the synonyms revealed similarities in syntactic patterns as both the synonyms occur mostly with the subject on the right and preposition on the left. Similar syntactic patterns of *arrive* and *reach* means they share similar syntactic behaviour. However, it was noted that the frequency of some of the grammatical features is different for both the synonyms. As the frequencies of objects, modifiers and Wh- words of *reach* are significantly higher than that of *arrive*. Though the two verbs *reach* and *arrive* share number of syntactical patterns yet there is a considerable difference in the collocates of these syntactic patterns. Sketch difference function of SkE was used to study the collocation patterns which revealed that though collocation tokens of *reach* and *arrive* in the subject pattern were almost equal but there were more words that collocate with *arrive* than with *reach*. The collocation tokens for *reach* in the ‘object’ pattern are 14285 as and collocation token of *arrive* which 1684. So, collocation token of *reach* in object patterns is significantly higher than that of *arrive*. This is because there are more meanings associated with *reach* and it is there has more usage in language. However, more words are used with *arrive* in ‘and/or’ pattern than with *reach*. There is a significant difference in collocation of *reach* and *arrive* with different ‘objects’ and there are more words that collocate as object of *reach*.

5. Conclusion

A corpus-based analysis of two verb synonyms *arrive* and *reach* reveal some differences in genre, colligation and collocation patterns in BNC. Synonym *reach* has wide occurrence in BNC as it has more senses associated with than *arrive*. There is a significant difference in collocation of *reach* and *arrive* with different objects. The findings of the study have its implications in English language teaching. It is a tough task for language teachers to teach every synonym pair to English Language Learners (ELLs). Therefore, enabling ELLs to use the existing corpus data can help them to learn synonyms by exploring and analysing the collocational and colligation patterns of synonyms in a corpus. The findings of the study recommend complementing dictionary entries with corpus-based analysis in English teaching for better understanding for frequently used synonym expressions. It might increase accuracy in English language learners.

References

- Albader, Y.B. (2001). *Comparing the usage and distribution of two near-synonymous words: Error vs. mistake*. (MA), University of Manchester.
- Atkins, B. T. S., & Levin, B. (1995). Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography*, 8(2), 85–114. doi: 10.1093/ijl/8.2.85
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, England: Cambridge University Press.
- Castello, D. (2014). *A corpus study of strong and powerful*. University of Birmingham, United Kingdom: unpublished assignment.
- Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Exploiting on-line resources to build a lexicon* (pp. 115–164). London, England: Psychology Press.
- Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In Philological Society (Eds.), *Studies in Linguistic Analysis* (pp. 1–32). Oxford, UK: Blackwell.
- Firth, J. R. (1968). A synopsis of linguistic theory. In F. R. Palmer (Ed.), *Selected Papers of J.R. Firth 1952-59* (pp. 168–205). London: Longmans.
- Gilquin, G. (2003). Causative ‘get’ and ‘have’: So close, so different. *Journal of English Linguistics*, 31(2), 125–148. doi: 10.1177/0075424203031002002
- Gries, S. Th., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34, 121–150.
- GU, B. (2017). Corpus-based study of two synonyms—obtain and gain. *Sino-US English Teaching*, 14(8), 511–522.
- Heid, U., & Gouws, R. H. (2006). A model for a multifunctional dictionary of collocations. In *Proceedings of the 12th EURALEX International Congress* (pp. 979-988). Torino, Italy: Edizioni dell’Orso. Retrieved from <https://euralex.org/publications/a-model-for-a-multifunctional-dictionary-of-collocations/>
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London, United Kingdom: Routledge.
- Jackson, H., & Amvela, E. Z. (2000). *Words, meaning and vocabulary: An introduction to modern English lexicology*. London, United Kingdom: Cassell.
- Kamiński, M. P. (2017). Visualisation of collocational preferences for near-synonym discrimination. *Lexikos*, 27(1), 237–251. doi: 10.5788/27-1-1401
- Liu, D. (2010). Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*, 15(1), 56–87. doi: 10.1075/ijcl.15.1.03liu
- Longman Dictionary of Contemporary English*. (2014). Pearson Longman; 6th Revised edition.
- Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1), 1–28. doi: 10.1080/01690969108406936
- Palmer, H. E. (1933). *Second interim report on English collocations, submitted to the Tenth Annual Conference of English Teachers, under the auspices of the Institute for Research in English Teaching, Dept. of Education, Tokyo, October 1933*. Kaitakusha, Tokyo.
- Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam, Netherlands: John Benjamins.
- Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, 3(10), 227–245.

- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Springer.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. New York, NY: Oxford University Press.
- Sinclair, J. M., Jones, S., & Daley, R. (2004). *English collocation studies: The OSTI report*. R. Krishnamurthy (Ed.). London, United Kingdom: Bloomsbury.

Exploring the Projection of the Colonizer/Colonized Binary through the Use of Adjectives in Selected Short Stories: A Corpus-based Study

Hira Fatima

MPhil Scholar, Air University, Islamabad

Sahar Gul

MPhil Scholar, Air University, Islamabad

Wasima Shehzad

Dean, Faculty of Social Sciences, Air University, Islamabad

Abstract

Postcolonial discourse has been a focus of research for the last two decades. Previous studies attempted to investigate several themes and patterns underlying these discourses using different methodologies (Chae, 2015; Budiyanto & Latifah, 2018; Laguarta Bueno, 2018; Salami & Tabari, 2018). Unlike previous studies, this study utilizes corpus tools to investigate the *us versus them* binary found in postcolonial texts. The study investigates how prejudice is reflected against the colonized in the selected short stories through the use of adjectives. It looks at the representation of prejudice in the short stories disseminating and reinforcing the binary of us and them. For this purpose, a corpus of 53861 words from eleven postcolonial short stories was formulated. Van Dijk's (1998) concept of the *Ideological Square* served as the theoretical framework for this study whereas Corpus Linguistics was the main methodological construct. The adjectives used in the corpus were analysed using WordSmith Tools 5.0. (Scott, 2008). From the adjectives, 19 were chosen for analysis and in those adjectives only 4 were positive while the remaining 15 were negative. The results show that the collocations of the negative adjectives with the colonized are dominant as compared to the positive adjectives which are more common with the colonizer. These results determine that in-group and out-group polarization is created through the use of adjectives.

Keywords: adjectives, applied linguistics, corpus linguistics, postcolonial studies, Us vs. Them binaries

1. Introduction

The present study is a corpus-based analysis of postcolonial short stories for exploring the adjectives. It analyses the *us versus them* binaries portrayed through the adjectives used in the texts. The purpose of exploring these binaries is to reveal the underlying ideas of colonizer and colonized in the light of Van Dijk's (1998) concept of the *Ideological Square*.

The binary opposition can be defined as a pair of 'terms that although opposed to one another are necessarily bound together as each other's condition of possibility' (Buchanan, 2010). Binary opposition in literature and culture studies are used in order to explore the relationships between people who belong to different groups, for instance, upper and lower class boundaries lead different groups of people to discrimination and prejudice. In literature,

binary opposition is a system which is used to explore the differences, such as class, cultural, and gender discrimination, between the individuals (Van Dijk, 1990). The postcolonial texts and the themes presented in them are also analysed in the light of binary oppositions (Said, 1978).

A remarkable prominence in the use of corpora can be seen in the field of linguistics. The corpus-based techniques allow the researchers to expand the size of their data which helps them in generalizing their findings in a wider context (Brezina, McEnery, & Wattam, 2015). The use of corpora in order to analyze the literary texts is a methodological approach which is mostly considered under the umbrella of Stylistics. Stylistics is often defined as the linguistic study of literature. In its general sense, it analyses the style of a language (Crystal & Davy, 2016). To look at the corpus from theoretical perspective, it helps the researchers to analyze the co-occurrence of words in the texts which are associated with different meanings (Sinclair, Jones, & Daley, 2004). Although the study of adjectives falls under the field of stylistics yet the focus of this research is on the use of adjectives in the postcolonial texts in order to highlight the portrayal of the binary oppositions. To execute the theme of *us versus them* in the texts, corpus is used as a tool.

Several studies have been conducted on the adjectives by different scholars. One such study was conducted on Bapsi Sidhwa's, who is an American–Pakistani novelist of Gujarati Parsi descent writing in English, use of adjective in her novels. The study focused on testing the hypothesis that female writers, and especially Sidhwa, use extensive adjectives in their writings. The study is conducted by applying corpus as a tool and the results verified the hypothesis (Mahmood, Mahmmod, & Nawaz, 2014).

The present study also deals with the adjectives used in the postcolonial texts. The short stories selected for this research are all written by postcolonial writers and come under the domain of postcolonial literature as they deal with the elements of colonialism, which is the behaviour of the colonizer and its effects on the colonized. Postcolonial literature includes the works that have some relations to the subjugating forces of imperialism and colonial expansion. This literature originated from the areas that once underwent colonization, for instance, Indian subcontinent, Nigeria, South Africa, and numerous parts of the Caribbean (Golden, 2015).

Eleven short stories; *The Yellow Wallpaper* by Gilman (2008), *A Rose for Emily* by Faulkner (2008), *Dead Men's Path* by Achebe (2008), *Recitatif* by Morrison (1983), *Once Upon a Time* by Gordimer (1991), *Our Dead, Your Dead* by Shamsie (2011), are selected. The eleven stories have several postcolonial elements embedded in them; however, the current study has tried to investigate the dichotomy of *us versus them* with the help of frequently used adjectives.

The current study is significant in the fields of postcolonial studies as well as Corpus Linguistics as it has taken the data i.e. short stories written by postcolonial writers and analysed it with methods and tools from Corpus Linguistics. There have been several studies on postcolonial discourse; however, those studies are mostly focused on literary analysis of texts rather than using corpus-based techniques (Chae, 2015; Budiyanto & Latifah, 2018; Laguarda Bueno, 2018; Salami & Tabari, 2018). Especially in Pakistan, Corpus Linguistics is rather a new field and therefore it is a less explored area, so the current study will add to this field.

The significance of the infusion of neutral ideals in books is incalculable. Racial and ethnic prejudices are a prevalent part of our society and people do not even realize at times that they are being prejudiced towards a certain group. Such social problems are like termites that slowly weaken the foundations of any society; therefore, awareness must be raised among the masses regarding these issues. This study will be significant in creating awareness among the students as well as the teachers and make them able to critically analyse the deeper meanings of every word within a text. Moreover, if the corpus that has been built for this study can also be expanded and further research can be conducted on it from alternate perspectives.

2. Literature Review

The aim of this study is to critically analyse the *us versus them* binaries portrayed in postcolonial short stories through the use of adjectives. In order to do this, first of all, some of the most notable and relevant studies have been elaborated in this section.

Discourse is an exploitative tool and it is used by politicians, writers, and motivational speakers alike to cumulate the support of public. Such techniques are used by politicians in political discourse in order to elicit emotions and achieve specific objectives (Bhatia, 2006). In order to explore the hidden meanings lying underneath these discourses, several theories have been at practice in the field of linguistics. One such approach is Critical Discourse Analysis or CDA. It is an interdisciplinary approach to the study of discourse that views language as a form of social practice (Van Dijk, 1998). Being a proponent of CDA, Van Dijk defines it as, “a type of discourse analytical research that primarily studies the way social power abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk in the social and political context” (1998, p. 352). This approach is concerned with the analysis of discourse within socio-political contexts and it helps in unveiling the underlying discursive practices, such as in-group and out-group polarization, ethnic prejudice, power abuse, racial discrimination and exploitation of basic rights.

Language, society and politics are unavoidably interlinked, so the main objective of critical discourse analysis is to identify language use as social practice. Wodak & Meyer believe CDA is not interested in inspecting a linguistic unit alone rather it is more interested in social phenomena which are not as simple as a single unit but more complex and therefore a “multidisciplinary and multi-methodical approach” is required to study them (2009, p. 2).

Critical Discourse Analysis is an interdisciplinary way to deal with language use as a social practice. The language users do not function in segregation rather they need a cultural, social and psychological framework. Its main concern is looking at the ways in which relations, identity, knowledge and power are developed through text and talks within societies. CDA is concerned with analyzing discourse through a critical method that is magnified by linguistic and social analysis. The key issues that Critical Discourse Analysis concentrates on are the power relations, dominance, inequality as well as the way social power abuse is established, opposed and recreated by the written and spoken texts in social and political setting (Tannen, Hamilton, & Schiffrin, 2015).

CDA is a worldview which is identified in interpreting the belief systems and power through organized examination of language, according to Van Dijk (2001). He believes that the purpose of CDA is to methodically interrogate the connection between the organization of the discourse and formation of ideologies. Additionally, it concentrates on how socio-

political discourse controls, broadcasts, induces and exhibits the power connection in the general public (Van Dijk, 2001).

To assist the theoretical framework of CDA within this study, the methodology of corpus has been used to identify the adjectives which are used to create in-group and out-group polarization within postcolonial discourse. This biasness is created due to the *us versus them* dichotomy within discourse (Van Dijk, 1998). The prejudice between the colonizers and the colonized has been prevalent throughout the history. In order to enhance their positive self-image, the colonizers started producing a negative other image of the colonized (Said, 1978). Such prejudices led to practices such as racism and the emergence of the occident and the orient. The West started a negative discourse about the orient in every field through producing literature in which they legitimized their act of invasion by showing a picture of the orient that was not real (Ambesange, 2016). West has created a hierarchy of the privileged and underprivileged in form of these binary oppositions. These binaries are universal, such as light and dark, as well as cultural, such as self and other (Caldas-Coulthard & Coulthard, 1996). Such dichotomies are also present in postcolonial short stories. They have been scrutinized by studying the adjectives used for the representation of us and them with the help of corpus tools.

Corpus methodology is a research method that is dependent upon qualitative and quantitative analysis both. It “utilizes a large and principled collection of natural texts, known as a ‘corpus’ as the basis for analysis” (Biber, Conrad, & Reppen, 1998, p. 4). It is a relatively new approach in the field of research and with its emergence a new door opened for the researchers. Corpus linguistics can help in identifying the significance of including and excluding certain linguistic and textual features. It uses empirical evidence as a reliable source in order to classify and investigate the linguistic structures being used (Brezina, McEnery, & Wattam, 2015). One of those lexical items can be adjectives, as well.

According to Burke and Swales (2003), adjectives can be categorized into seven different classes in accordance with their denotations like aesthetic appeal, assessment, deviance, relevance, size, acuity and strength. They investigated evaluative adjectives in seven classes within academic writing and speech. They identified from their research that strength and relevance adjectives were more regular in academic writing than academic speech.

Research that has examined the adjectives used to represent the *us versus them* binary in electronic and print media has generally found evidence of biasness. A study conducted on the newspaper coverage of American mass shootings indicated that the adjectives used to describe non-white attackers were negative and were almost always related to their ethnicity. However, the white mass shooters were never labeled according to their ethnicity. This type of language usage contributes towards racial and ethnic discrimination among masses (Tanvir, Khoula, & Zahra, 2018). Another corpus-based study was conducted on the representation of trans people in the British press from 2013 to 2015 which showed similar results (Zottola, 2018).

A study (Pasha, 2011) examined the Islamic ideologies presented in an Egyptian newspaper using Van Dijk’s *Ideological Square* (1998). It found that the newspaper discourse emphasised on positive ideologies while deemphasising the negative ones. It portrayed negative ideologies to deemphasise on the good deeds of the Muslim society. Another study was carried out on news reports produced by The Los Angeles Times and

Tehran Times (Ahmadian & Farahani, 2014). This study investigated the ideological differences in the discourse of both newspapers about Iran's Nuclear Program. The results showed that both the newspapers used biased and prejudiced language in order to emphasise their positive self-image and create a negative other image for Iran.

In a study, the process of evaluation is used as it is defined by Hunston and Thompson (2010), that evaluation is a mean which is used to express the writer's or speaker's attitude and feelings towards their production of language. To make a sentence evaluative, many linguistic features are used; however, adjective is one of the most important features making a sentence evaluative.

Marzá (2011) conducted a study on the use of evaluative adjectives in promotional hotel websites. She analysed the syntactic behavior and collocate patterns within the discourse found on hotel websites. Although she expected to find out a relatively large amount of data on evaluative elements, the results were quite surprising. The findings showed that the hotel websites discourse is not as direct and hyperbolic as one would expect rather it uses subtle language. This means that even though the website discourse includes positive self-representation, yet it doesn't solely rely on evaluative adjectives and uses other lexical items, too. Other studies using corpus-based techniques have tried to investigate different forms of *othering* through discourse. A study investigated the concept of *othering* in the representations of hosts in commercial tourism discourse. The researcher, using the Corpus-Assisted Discourse Studies (CADS), compared external (other) and internal (self) representations and found that the Western tourists industry generalizes the role of hosts and as a result showcases them as someone who is more like a servant to the tourists (Jaworska, 2016). Such representation of one group of people leads to forming stereotypical views about them.

The current study has probed eleven short stories from different eras to see how a discourse of *us versus them* was produced through the use of adjectives. This study has dealt with a practical problem of social prejudice. This becomes a cause for the reinforcement of many preconceived notions about the colonizers as well as the colonized. So, current research has tried to bring forth the hidden elements of discrimination and *othering* within the adjectives that are used in the stories to help the readers better understand these dichotomies. For this purpose, the study examines following research questions:

1. How is prejudice reflected against the colonized in the selected short stories through the use of adjectives?
2. How is the representation of prejudice in the short stories disseminating and reinforcing the binary of us and them?

3. Research Methodology

The current study is quantitative in nature as the frequency of the words was calculated by using corpus tools and then the findings were qualitatively analysed. So, this study is both quantitative as well as qualitative in nature. The theoretical framework selected to conduct this study is the *Ideological Square* given by Van Dijk (1998, p. 35). In postcolonial discourse, when the positive aspects of *us* are emphasized and the negative aspects of *us* are de-emphasized, the positive aspects of *them* are de-emphasized and the negative aspects of *them* are emphasized. This concept is defined as *Ideological Square* by Van Dijk (1998, p. 35). This strategy works on both levels; form and meaning (Van Dijk,

2006). In this study, we have analysed how this concept is deployed in the selected postcolonial short stories by scrutinizing through the adjectives used to represent the colonized and the colonizers. A corpus of 53861 words was developed from eleven postcolonial short stories which are as follows:

Table 1: Short stories included in the corpus

S.No.	Short Story	Author
1	<i>The Man Who Would Be King (1888)</i>	Rudyard Kipling
2	<i>An Outpost of Progress (1897)</i>	Joseph Conrad
3	<i>Sacrificial Egg (1959)</i>	Chinua Achebe
4	<i>The Madman (1972)</i>	Chinua Achebe
5	<i>Recitatif (1983)</i>	Toni Morrison
6	<i>Once Upon a Time (1989)</i>	Nadine Gordimer
7	<i>The Revenge of Her Race (2002)</i>	Mary Beaumont
8	<i>The Yellow Wallpaper (2008)</i>	Charlotte Perkins Gilman
9	<i>A Rose for Emily (2008)</i>	William Faulkner
10	<i>Dead Men's Path (2008)</i>	Chinua Achebe
11	<i>Our Dead, Your Dead (2011)</i>	Kamila Shamsie

Socio-political manipulation and mind-control is ideological most of the times, and almost always includes power abuse by the dominant, hegemonic rulers (Van Dijk, 2006). To analyse how and in what ways this manipulation is achieved, Van Dijk (1995) has expanded on how various discursive strategies are used to ensure that the audiences or the readers bend to the speaker's or writer's will. Moreover, discourse employs various strategies to emphasize and/or de-emphasize meanings (Van Dijk, 2006). To this end, the strategy of positive self-representation, and negative other representation is most widely used in discourse, especially political discourse (Van Dijk, 2001). This strategy works such that in-group and out-group polarization is created, thereby creating an Us/Them binary, that is readily "coded in text and talk" (Van Dijk, 2006, p. 126). Van Dijk (2001) has also listed different discursive strategies that are utilized by speakers/writers to exhibit positive self-presentation and negative other-presentation. Within CDA, these discourse levels help to highlight the hidden ideologies that are coded within a specific written or spoken discourse through the strategy of emphasizing and de-emphasizing in-group and out-group properties.

3.1. The Corpus

The instruments used to examine the colonizer/colonized binaries through adjectives in the texts are corpus-based. The data selected for this study was eleven postcolonial short stories and it was retrieved from different websites in .pdf format. The short stories were selected through convenience sampling technique, which is defined as "a technique of sampling where the readily available primary data source will be used for the research without additional requirements" (Dudovskiy, 2016). To analyse this data, the corpus software, WordSmith Tools 5.0 (Scott, 2008) was used. WordSmith Tools 5.0 has three main options: Concord, Keywords and Wordlist, however, for this study only the Concord and Wordlist options were used. A wordlist was generated with the help of Wordlist option in the software. This wordlist was pasted in the CLAWS Tagger (Rayson & Garside, 1998) and a tagged wordlist was obtained which was then scrutinized for identifying adjectives that were projecting the colonizer/colonized binary. These adjectives, taken from the KWIC list

in the corpus, were then selected for qualitative analysis keeping in mind the above explained theoretical framework given by Van Dijk (1998).

4. Data Analysis

The current section has aided the research to move a step ahead by analyzing the data with the help of the selected framework. Using the theoretical framework given by Van Dijk known as the *Ideological Square*, (1998) the data has been analysed. The analysis has further been carried out by examining the social identities constructed through the binaries of us and them. The corpus that has been prepared for this study has a number of different adjectives within it. After short listing those to the ones that are showing *us versus them* binary the researchers further selected every fifth adjective out of the seventy-seven as a detailed analysis of each adjective was not within the scope of this study.

Table 2: Adjectives selected for analysis

S.No.	Adjectives	S.No.	Adjectives
1	Able	11	bad
2	Dank	12	energetic
3	Good	13	imaginative
4	infuriating	14	modern
5	peculiar	15	puzzling
6	repellant	16	separate
7	singular	17	smouldering
8	summery	18	tiny
9	unclean	19	wavering
10	worse		

The current study has analysed the adjectives in two ways. First of all, it analysed which adjectives are used with negative connotations and then those that are used with positive connotations. After analyzing both types of adjectives, their overall frequency was calculated to estimate which type of adjectives are used more in the corpus. The *us versus them* theme is a discursive phenomenon that constructs an ontological, philosophical and ideological dichotomy amongst various identities (Said, 1993). The corpus shows this binary at several places.

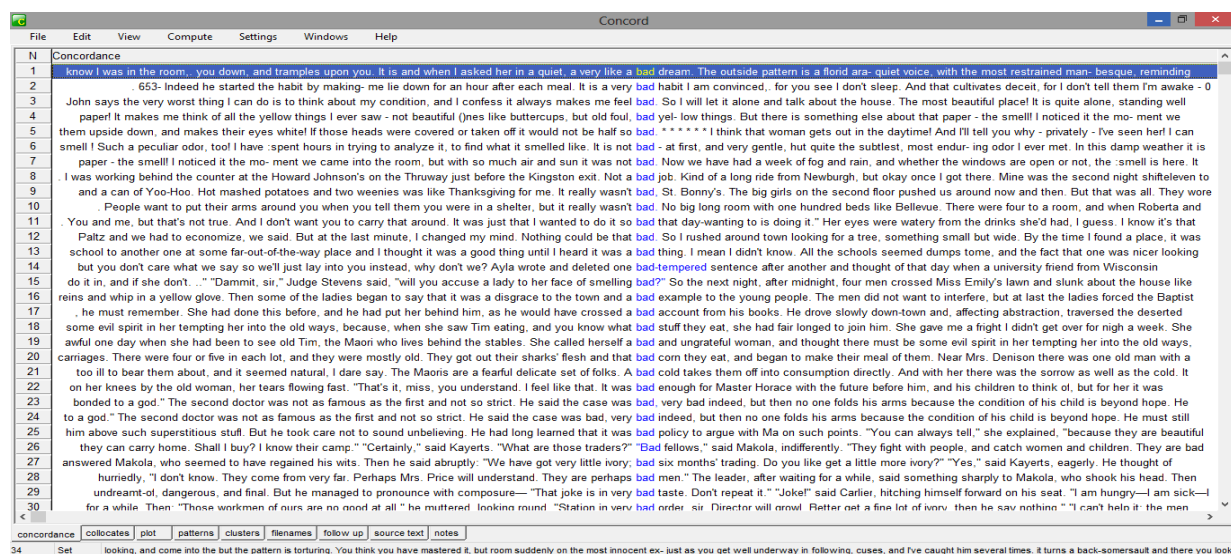


Figure 1: Concordance of *bad*

The adjective *bad* has been used in the corpus to refer to the colonized as being a bad example for the colonizers. They are seen as a bad influence for the newer generations of the colonizers. They are seen as a threat and a disgrace to the colonizers as Miss Emily in *A Rose for Emily* (Faulkner, 2008) was seen to be a bad example for the people of the town, especially the younger ones.

As Said mentions in *Orientalism* (2006), the colonizers (or the West to be precise) sees the colonized to be *lazy, mysterious, imaginative, unnatural* and *irrational*. This corpus gives us a clear picture of all such adjectives being used for the colonized. The colonizer thinks that the colonized are nothing but imaginative *if* they ever point out the problems that are around them due to the colonizers. In *The Yellow Wallpaper* (Gilman, 2008), the woman is representing the colonized and whenever she tries to talk to her husband about her thoughts and feelings, he politely shuts her out by that she is imaginative with a *power and habit of story making*. This shows that the colonizer has no regard for the colonized and it snubs them of their right to speak their heart out by labeling them with such titles.

The colonizers not only have a problem with the thoughts and feelings of the colonized rather they have a major dislike for everything that is related to them whether it is their way of living or the color of their houses or even with the smell that they think is a trademark of their houses. The adjectives used in the corpus to describe this are peculiar, close and dank which have negative connotations associated with them, showing the margin that is created by the colonizer for the colonized. The colonizer has created a specific boundary for the colonized which they cannot come out of.

Looking at the negative adjective *repellent*, it can be seen that even the colors associated with the supposedly inferior class are presented as something entirely repulsive and intolerable. Just to refer to the color of the wallpaper in the room (representing the colonized land) four different adjectives are used, like *revolting, smouldering, unclean* and not to forget *repellent*. This shows the intensity of hatred and disgust that is in the minds of the colonizers for the colonized. Two of the four words, revolting and repellent have very strong negative connotations and are only used to express utter disgust.

Another adjective *worse* is also used in the corpus to describe the condition of material related to the colonized people whether it is their furniture or their wallpaper. Everything seems worse to the eye of the colonizer. The colonizer thinks of himself as the superior know-it-all while the colonized, to them, are nothing more than objects.

Obi of the *Dead Men's Path* (Achebe, 2008), prefers *modern* methods in place for these *superannuated* and *unprogressive* practices. This is also a representation of the *us versus them* binary showing he believes that the natives, who are traditional people, are backward and obsolete as their methods and practices are now outdated. He, representing the colonizers, believes that only he can bring *change* and a *modern* environment within the backward land of Ndume, representing the colonized. In doing so, they forget that what they believe is right may not be right for others but in creating their positive self-image the colonizers will see their group succeed instead of having a win-win situation for both the groups (Said, 2006).

Another positive adjective that is used to represent the colonizers is *energetic* along with *young*, however; on the other hand, negative adjectives like *puzzling* and *nervous* are used for the colonized.

A very interesting adjective is *able* when it occurs in *The Yellow Wallpaper* (Gilman, 2008). The story is written in the first-person narrative with the colonized being the speaker. Throughout the story the adjective *able* has been used in two contexts. In the first place, the narrator says, “*But he said I wasn't able to go, nor able to stand it*” (Gilman, 2008, p. 651), where *he* stands for John, the colonizer and *I* for the narrator. However, as the story progresses the narrator says at point “*I don't want to. I don't feel able*” (Gilman, 2008, p. 651). This shows that how the colonizer was successful in making the colonized doubt their abilities. The colonizer has created these binaries in such a manner that even the ones who have been marginalized and victimized do not even understand that they have been, rather they start feeling as if the discourse created by the colonizer is real.

After looking at the selected adjectives individually, when we scrutinized the entire list of adjectives, it was found that from the 19 adjectives only 4 were used in positive connotations and that too for the colonizers. While the remaining 15 were used in negative connotations and all of them were used for the colonized.

Table 3: Positive and negative adjectives projecting the us vs. them binary

S.No.	Positive Adjectives	S.No.	Negative Adjectives
1	Able	1	bad
2	energetic	2	dank
3	Good	3	imaginative
4	modern	4	infuriating
		5	peculiar
		6	puzzling
		7	repellant
		8	separate
		9	singular
		10	smouldering
		11	summery
		12	tiny
		13	unclean
		14	wavering
		15	worse

5. Conclusion

This study examined the role of short stories in propagating the stereotypes related to a group of individuals. It analyses the *us versus them* dichotomies present in the selected short stories that are the main feature of postcolonial discourse. A corpus-based approach was used to analyse the postcolonial short stories. For that purpose, a corpus of 53861 words was developed from eleven postcolonial short stories. Within the data, the adjectives were identified and analysed for showing any colonizer/colonized binaries. The data was analysed in the light of CDA approach presented by Van Dijk (1993) and the concept of *us* and *them* also given by Van Dijk (1998) in his *Ideological Square* framework.

There are a number of adjectives used to create the dichotomies of *us* and *them* within the data. However, analyzing all the adjectives was not within the scope of this study. Therefore, every fifth adjective was taken for analysis so, only 19 adjectives were analysed in detail. The result of the analysis shows that the adjectives used in the corpus exhibit a struggle between the binaries i.e. new and old, traditional and modern, the colonizer and the

colonized. The colonizers create these binaries of *us* and *them* in order to marginalize the colonized group and as a result they lose a chance to live in harmony with each other. The findings revealed that in the 19 adjectives only 4 were positive while the remaining 15 had negative connotations. Moreover, the positive adjectives were used for the colonizer while all the negative ones were used for the colonized.

The use of such adjectives that help in creating a positive self-image and a negative other image aids in the promotion of stereotypes. Such strategies are used by writers to subtly reinforce certain ideas in the minds of people regarding every race, class, nation, ethnicity, and gender. These discourses help creating the dichotomies and divisions of black and white, right and wrong, *us* and *them* which ultimately leads to biasness in people's behavior.

In this way, the researchers have tried to bring forth the cultural clash and social bias between the colonizers and the colonized. Moreover, the research has substantiated its objectives by answering the question about how prejudice is shown in the postcolonial short stories. Also, the results of the research showed that negative adjectives were used to represent the colonized while positive ones were used to represent the colonizer which proved that the research was in line with the theoretical framework given by Van Dijk (1998, p. 35). However, future researchers can expand the corpus and further study the dichotomies of *us versus them* in the corpora. These dichotomies can also be studied in any other genre apart from fiction. So, future research can be conducted on newspaper representations of *us* and *them*.

References

- Achebe, C. (2008). Dead men's path. In R. S. Gwynn (Ed.), *Literature: A pocket anthology* (4th ed.). London, England: Pearson.
- Ahmadian, M. & Farahani, E. (2014). A critical discourse analysis of The Los Angeles Times and Tehran Times on the representation of Iran's Nuclear Program. *Theory and Practice in Language Studies*, 4(10), 2114–2122. doi: 10.4304/tpls.4.10.2114-2122
- Ambesange, V. P. (2016). Postcolonialism: Edward Said & Gayatri Spivak. *Research Journal of Recent Sciences*, 5(8), 47–50.
- Bhatia, A. (2006). Critical discourse analysis of political press conferences. *Discourse & Society*, 17(2), 173–203. doi: 10.1177/0957926506058057
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, England: Cambridge University Press.
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173. doi: 10.1075/ijcl.20.2.01bre
- Buchanan, I. (2010, December 4). Binary opposition. In *Oxford reference - A dictionary of critical theory*. Oxford, England: Publisher: Oxford University Press.
- Budiyanto, A., & Latifah, E. (2018). Aruna's ambiguity and the paradoxical flavors of her tongue: Postcoloniality of Laksmi Pamuntjak's culinary novel. *Jurnal Pendidikan Bahasa dan Sastra*, 18(2), 221–234. doi:10.17509/bs_jbps.v18i2
- Caldas-Coulthard, C. R., & Coulthard, M. (1996). Discourse, power, and access. In *Texts and practices: Readings in critical discourse analysis* (pp. 84–104). London, United Kingdom: Routledge.
- Chae, Y. (2015). Postcolonial ecofeminism in Arundhati Roy's the God of small things. *Journal of Postcolonial Writing*, 51(5), 519–530. doi:10.1080/17449855.2015.1070010
- Crystal, D., & Davy, D. (2016). *Investigating English style*. London, United Kingdom: Routledge.
- Dudovskiy, J. (2016). Convenience sampling – Research methodology. Retrieved April 3, 2019, from <https://research-methodology.net/sampling-in-primary-data-collection/convenience-sampling/>
- Faulkner, W. (2008). A rose for Emily. In R. S. Gwynn (Ed.), *Literature: A pocket anthology* (4th ed.). London, England: Pearson.
- Gilman, C. P. (2008). The yellow wallpaper. In R. S. Gwynn (Ed.), *Literature: A pocket anthology* (4th ed.). London, England: Pearson.
- Golden, A. (2015, April 19). A brief history of postcolonial literature, part I. Retrieved July 18, 2019, from <https://blog.bookstellyouwhy.com/a-brief-history-of-postcolonial-literature-part-i>
- Gordimer, N. (n.d.). Once upon a time. Retrieved April 9, 2019 from https://archive.org/stream/GordimerOnceUponATime/Gordimer_Once_Upon_a_Time_djvu.txt
- Hunston, S., & Thompson, G. (2010). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford, England: Oxford University Press.
- Jaworska, S. (2016). A comparative corpus-assisted discourse study of the representations of hosts in promotional tourism discourse. *Corpora*, 11 (1). 83–11. doi:10.3366/cor.2016.0086
- Laguarta Bueno, C. (2018). The plight of not belonging: Jean Rhys's "let them call it jazz" and "the day they burned the books". *ES Review. Spanish Journal of English Studies*, (39), 157–172. doi:10.24197/ersjes.39.2018.157-172
- Mahmood, A., Mahmmod, R., & Nawaz, S. (2014). A corpus driven study of adjectives in Sidhwa's fiction. *Global Journal of Human-Social Science (G)*, 14(2). 59–64. Retrieved from <https://socialscienceresearch.org/index.php/GJHSS/article/view/1150>
- Marzá, N. E. (2011). A comprehensive corpus-based study of the use of evaluative adjectives in promotional hotel websites. *ODISEA. Revista de estudiosingleses*, (12), 97–123. doi: 10.25115/odisea.v0i12.222
- Morrison, T. (n.d). Recitatif. *Genius*. Retrieved April 9, 2019 from <https://genius.com/Toni-morrison-recitatif-annotated>

- Pasha, T. (2011). *Islamists in the headlines: Critical discourse analysis of the representation of the Muslim Brotherhood in Egyptian newspapers* (Unpublished Doctoral dissertation). University of Utah, USA.
- Rayson, P., & Garside, R. (1998). The claws web tagger. *ICAME Journal*, 22, 121–123.
- Said, E. W. (1978). *Orientalism*. New York, NY: Pantheon Books.
- Said, E. W. (1993). *Culture and imperialism*. London, England: Chatto & Windus.
- Said, E. W. (2006). *Orientalism*. In B. Ashcroft, G. Griffith & H. Tiffin (Ed.), *The postcolonial studies reader* (pp. 24–27). New York, NY: Routledge.
- Salami, A., & Tabari, B. H. (2018). Things fall apart and Chinua Achebe's postcolonial discourse. *International Journal on Studies in English Language and Literature (IJSELL)*, 6(3), 19–28. doi:10.20431/2347-3134.0603004
- Scott, M. (2008). WordSmith Tools version 5, Liverpool: Lexical Analysis Software.
- Shamsie, K. (2011, September 6). Our dead, your dead. *The Guardian*. Retrieved from <https://www.theguardian.com/books/2011/sep/06/9-11-stories-kamila-shamsie>
- Sinclair, J. M., Jones, S., & Daley, R. (2004). *English collocation studies: The OSTI report*. R. Krishnamurthy (Ed.). London, United Kingdom: Bloomsbury.
- Swales, J., & Burke, A. (2003). “It’s really fascinating work”: Differences in evaluative adjectives across academic registers. *Language and Computers*, 1–18. doi: 10.1163/9789004334410_002
- Tannen, D., Hamilton, H. E., & Schiffrin, D. (2015). Critical discourse analysis. In *The handbook of discourse analysis* (2nd ed., pp. 466–485). doi: 10.1002/9781118584194
- Tanvir, O., Khoula, & Zahra, T. (2018). Newspaper racial agenda: A critical discourse analysis. *Corporum: Journal of Corpus Linguistics*, 1(2), 1–16.
- Van Dijk, T. A. (1990). Social cognition and discourse. In H. Giles & W. P. Robinson *Handbook of language and social psychology* (pp.163–183). London, England: John Wiley & Sons.
- Van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, 4(2), 249–283. doi: 10.1177/0957926593004002006
- Van Dijk, T. A. (1995). Ideological discourse analysis. In E. Ventola & A. Solin (Eds.), *Interdisciplinary approaches to discourse analysis* (pp. 35–161). Helsinki, Finland: University of Helsinki.
- Van Dijk, T. A. (1998). Opinions and ideologies in the press. In A. Bell & P. Garrett (Eds.), *Approaches to media discourse* (pp. 21–63). Oxford, England: Wiley-Blackwell.
- Van Dijk, T. A. (2001). Multidisciplinary CDA: A plea for diversity. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 95–120). London, England: SAGE Publications Ltd. doi: 10.4135/9780857028020.n5
- Van Dijk, T. A. (2006). Ideology and discourse analysis. *Journal of Political Ideologies*, 11(2), 115–140. doi: 10.1080/13569310600687908
- Wodak, R., & Meyer, M. (2009). Critical discourse analysis: History, agenda, theory and methodology. In *Methods of critical discourse analysis* (2nd ed., pp. 1–33). doi: 10.4135/978085702802
- Zottola, A. (2018). Transgender identity labels in the British press (2013–2015): A corpus-based discourse analysis. *Corpus Linguistics in Language and Sexuality Studies: Developments and Prospects*, 7(2), 237–262. doi: 10.1075/jls.17017.zot

Future Workshops and Seminars

Topic	Date	Trainer
Corpus Approaches to Teaching Academic Writing	September 16, 2019	Dr. Akhtar Abbas Assistant Professor, Air University, Islamabad
Discourse Analysis: Exploring Keywords and Analyzing Keyness	October 21, 2019	Dr. Ameer Sultan, Lecturer, IIUI
Analyzing Collocations through AntConc	November 17, 2019	Dr. Tehseen Zahra Assistant Professor, Air University, Islamabad
Text Analysis through Wordsmith Tools	December 15, 2019	Dr. Tehseen Zahra Assistant Professor, Air University, Islamabad
Using Online Corpora for English Language Teaching	January 13, 2020	Dr. Tehseen Zahra Assistant Professor, Air University, Islamabad
Discourse Analysis: Exploring Semantic Prosodies and Semantic References	February 17, 2020	Dr. Tehseen Zahra Assistant Professor, Air University, Islamabad
Analysing Metadiscursive Practices in Text	March 9, 2020	Dr. Akhtar Abbas Assistant Professor, Air University, Islamabad
Corpus Approach Multimodal Genre Analysis	April 20, 2020	Dr. Akhtar Abbas Assistant Professor, Air University, Islamabad
Using Natural Language Toolkit NLTK for Corpus Analysis	May 18, 2020	Dr. Khalid Mahmood, Lecturer, IIUI

For more information contact:

Dr. Tehseen Zahra
crcworksem@gmail.com