

Woman and Media: A Corpus-based Critical Discourse Analysis of Written Press

Ms. Robina Masud
Department of English

1. Introduction

Freedom of expression is a basic human right. Without it, the ability to express oneself in the fullest capacity is impossible. There is multitude of institutions and organizations that work globally to protect, propagate and reinforce this agenda through various treaties, charters and MOUs. Putting this differently; freedom of expression provides an assurance to the citizens of any country or region to work and act freely and effectively under the democratic system. But unfortunately, freedom of expression does most extend to men than to women in Pakistani society. This may also be observed in print media discourses where men employ certain type of language (using linguistic techniques) to express themselves more freely than women. Sun (2014) explains the importance of freedom of expression as "... a democratic society, freedom of expression not only empowers the decision makers to fully appreciate the strength of the diversity of opinions but also encourages the acceptance of such diverse views".

The uses of certain expressions in print media discourses present distinctions in the use of language by both men and women. Despite great advances in recognizing women's rights in every sphere of life, there seems to be marginalization when it comes to express views in print media. This may lead to gender-based discrimination in publishing the writings of women or they may not be allowed to publish on certain types of topics by the concerned agencies. It is difficult to make an assessment on the global patterns of writings by women but the first quantitative study conducted in 1995 known as the global media monitoring project shows that only 17% of the world's news media persons represent women (MediaWatch, 1995). As there are known differences in speech on the basis of gender (Lakoff, 1972), there surely exists the same in print media discourses. The use of mitigation devices (*I think, sort of*), qualifiers (*much beautiful, very sad*), and use of more tag questions (*Isn't it?*) are the characteristic representation of women's language (ibid). So a speculation is made that because women use gender specific language which is more women in its nature, the publishing agencies may not publish their work.

Pakistan is among those countries where freedom of expression for women in the print media is limited to the genre of fashion industry at the most. There are less number of women who write on topics such as politics, religion, social issues (including women's issues) and technology. Without a considerable number of women writers participating in the important issues of a country, it is difficult to listen to their views from the female perspective. Media being the largest source of disseminating information in the present world, it is important that both the genders contribute to the governance of any democracy in order to build mutual trust among them. But such climate exists in result of impartial and independent media (Sen, 1999, Norris, 2004, Castells & Himanen, 2014).

Therefore, the present study endeavors to look at the linguistic differences in corpus of online press media discourses on the basis of gender and their relation to women and their freedom of expression (*through mixed method analysis of linguistic elements*).

2. Statement of the Problem

From the researcher's perspective, Pakistani online newspapers show less number of women writers' representation in the media industry. This leads to the problem that women do not contribute to the major issues that are important for a balanced functioning of any society. However, any major newspaper in the country now-a-days employs women for their own monetary benefits including advertisements or branding. The small numbers of women who do create content are again limited to topics that are feminine.

To unfold the issue, the researcher intends to carry out this research in an attempt to understand the phenomenon of freedom of expression in the print media industry on the basis of linguistic differences. The researcher believes that there are certain linguistic differences that render women not to talk about some issues and hence express themselves because of their language. It is also to be believed that the choice of words (lexis), sentence pattern (syntax), the selection of topics (themes), collocations, the frequencies of words and the use of keywords all contribute to the writing expression. Another important aspect that may be explored is to show differences in opinions and the tendencies towards the selection of topics on the basis of gender.

3. Purpose of the Study

The study tries to analyze and understand the major differences in written media discourses on the basis of gender. These differences will then form the foundation for the qualitative analysis of the data collected through quantitative means. Through numeric data, qualitative generalizations will be made forming the conclusions on the subject matter of women, their freedom of expression in the media. Hence, the study provides useful insights through analysis as to why women represent less in media and if they do, what are the implications of their content.

4. Significance of the Study

- i. The study may highlight the role of women in constructing and portraying their worldview by considering that language is a mode of communication that shapes the way we think.
- ii. By using mixed method analysis, one constructs the narrative that language may be the hurdle for the less representation of women in the press media.
- iii. Most importantly, the research may prove to be highly beneficial for the media industry, providing statistical data to them so that they may provide equal opportunities to work to both the genders.

5. Research Questions

Following are some of the research questions or goals that triggered the foundation of this research and serve as a guide to drive this study in the right direction.

- i. How freedom is constructed and projected in online press media discourses?
- ii. What are the linguistic differences between men and women in presenting their worldviews through online press media discourses?
- iii. How do the themes (in terms of ordering focus and emphasis) differ for both the genders?
- iv. What collocational patterns do appear for the writers of both genders and what do they represent?

6. Literature Review

6.1 Freedom of Expression and Media

Freedom of expression is a matter of serious concern for the balanced functioning of a society. It is not only protected through law, but a question of morality and ethics. It is so fundamental to the rights of individuals that limiting it may cause suffrage and damage to the society. But defining freedom of expression is a difficult task as some utterances may inflict harm to the emotions of other people, groups or societies and thereafter requires intervention (Rønning, 2009).

In today's complex world, the need to address publish interests is becoming more and more common in the media industry. The emergence of various social media platforms including facebook, instagram, blogs and so on has increased the demand for more "public" (Voltmer, 2015) oriented content. The ever increasing competition among these media companies has sparked 'more spicy' content creation rather than the one that is important for the social and mental well being of the people. In the process to satisfy the needs of the public, women are ignored as content creators and they only participate in the issues that are demanded by the media industries. On the contrary, there must be equality and rights should be protected of both the genders in whatever topics they want to write about for the media industry.

Some countries like Scandinavian have been trying to adopt the model of 'rights-based' approach to protect the rights of women, minorities and ethnic groups in terms of media participation. One of the definitions of rights-based approach is given by the Council of Europe (2011:32) as: "A rights-based approach is a conceptual framework for a process of development that is based on international human rights standards and directed at promoting and protecting human rights, analyzing inequalities, and redressing discriminatory practices and the unjust distribution of power". Amartya Sen (2004) argued that human rights hold special significance to individuals and societies and these are linked to "survivability in unobstructed discussion" (p. 320) and therefore requires freedom of expression and to be heard in dialogues in civil ways. The media ecosystem that facilities people in spreading information in the modern world through digitalized ways also present many problems related to privacy, authorship, copyrights, rights of women and minorities to publish their content (Ziccardi, 2013).

6.2 Critical Discourse Analysis

CDA (Critical Discourse Analysis) is developed from Critical Linguistics at the University of East Anglia to understand how media manipulates the language (Fowler, 1979). Van Dijk (2009, p.62) prefers to call it Critical Discourse Studies because the word 'studies' encapsulates broader meanings hence making it an interdisciplinary field using various methods and theories for analyses. The very nature of CDA deals with language as 'discourse' embedded in social context, critically examining it to find power relations that exist in

the text through multiple methods and theories. Emil Benveniste (1971) propounded the term discourse as a replacement to language to suggest that language is an abstract system while discourse is the social use of language for particular purposes. But to examine discourse cognitively to access how journalists create it and how the readers interpret it is a difficult task. Instead, discourse analysts try to analyze ‘trace of discourse’ (Partington, Duguid & Taylor, 2013, p.2-3).

A discourse may also be defined in terms of linguistics structures such as words, metaphors and pragmatic strategies that are associated with particular genres of study, for example, media discourse (news), medical discourse (Bakhtin, 1981), political discourse, religious discourse, law discourse, trade discourse and so on. This emphasizes the fact that particular discourses are akin to particular social contexts that make these discourses unique containing certain specific linguistic structures. Such a definition of discourse may further broaden its meaning which the social constructionists believe create reality for the readers through linguistic and social practices (Burr, 1995). A similar version of the definition of discourse is found in the works of Foucault (1972) and Hall’s (1992:291).

6.3 Representation

Language is unique in the sense that it constructs a certain version of reality for its users instead of projecting the exact reality which to no one is definitive. Explaining this phenomenon, Bakhtin (1981) illustrates the process through metaphor by saying, “a word is construed as a ray of light cast upon its referent” but each word can only highlight one part, side or aspect of the whole object. Similarly, linguistic structures illuminate certain aspects of a word (referent) and ignore the rest. Therefore, the words like ‘broadcasting’ and ‘on air’ are used in different discourses but they refer to the same referent thus drawing attention towards to varying aspects of the each word. For the first word, it refers to the business or profession of radio and TV while the latter refers to the process of recording something. Hall’s (1992) provides comprehensive definition of discourse by saying that a discourse is composed of multiple representations that collectively presents the object in a unified way. Conclusively, discourse in the social constructionist view provides broad assumption about the world around through unified set of implicit representations.

6.4 Corpus Linguistics (CL)

The field of corpus linguistics deals with the analysis of machine-readable large sized files with the help of various tools and software (McEnery & Wilson, 1996). It is based on the experimental or empiricist tradition where observable patterns of language are analyzed for authentic quantitative results. After the electronic corpus emergence in 1960s, it is now believed that most of the intuitions based on traditionalist approach are incorrect about the language.

CL draws upon partial quantitative analysis by measuring the probabilities and frequencies of the language features (Partington et al., 2013). Almost all corpus tools are structured to find out the frequency of the language features that occur in the text by measuring the statistical significance of them. But CL “is much more than just bean-counting” (Biber & Conrad, 2001) because it is primarily based on quantitative evidences which are further elaborated through qualitative methods hence making a kind of mixed research. For example, the patterns that are redundant in a piece of text are looked at in more detail with reference to their social contexts of use. Even when a pattern is high in frequency, its significance does not only depend upon the quantitative results but the qualitative analysis of the abstract level at which these are used must be considered. After all, qualitative analysis is necessary to give meanings to the statistical results. Therefore, a mixed method is necessary to make meaningful generalizations about the objective results drawn.

6.5 Lexical Grammar

CL is primarily concerned with dealing theory of lexical grammar (Sinclair, 1991, Halliday, 1966). This theory of lexical grammar changes the popular perceptions that an orthographic word is the product of one linguistic choice. Instead, it is proposed by Sinclair (1991, p.110) that language patterning or its processing is based on “idiom principle” where language is made up of multiple word patterns known as “extended lexical units” varying in length and internal structures. Thus, discourses (texts) are not fixed like walls but constellations where words attract other words and form lexical or extended lexical units.

The patterns of lexical units vary in texts. Words that frequently co-occur in the presence of co-textual context are known as collocations. Such an association is most often directional. For example, the word ‘right’ collocates with direction, person, color, decision and so on but the opposite is incorrect, i.e., it works in one direction. This means that the word ‘right’ collocates as: right direction, right person, right color, and right decision and so on but the vice versa is not correct. Moreover, relationships of words on the basis of grammar are known as colligations whereby a word is associated with specific word class like noun, pronoun, verb (Hoey, 1997).

For example, Sinclair observed that the word ‘budge’ frequently shows up after a helping/modal auxiliary verb.

On the basis of semantic criteria, words also co-occur frequently with certain semantic fields (units of words that share meanings together. This process of co-occurring with certain semantic fields is named as “semantic preference” by Stubbs (2001, p.65). For example, the phrase ‘class of’ co-occurs with water and cola referring to cold drinks. Semantic preference may further develop ‘semantic prosody’, a process in which the meaning of a word is shaped by its semantic preference. Then there are phenomenon referred as ‘evaluative prosodies (co-occurring with specific evaluative polarity) and discourse prosodies (co-occurring with certain semantic fields).

7. Research Methodology

The research is based on an amalgamated approach combining both quantitative and qualitative methods. The first step involves collecting a corpus of nearly 100 million words from 20 or more online international newspapers. Within these newspapers and publishing, content and data related to women and their rights of freedom of expression by different countries over the years are particularly focused on. This is being done to ensure that each extract has the same underlying ideology and adds to the comprehensive nature of the data and in turn the study itself. The corpus is also annotated into metadata, textual markup, and linguistic annotation, for a clearer investigation and analysis.

The analysis here is done using corpus-based discourse rather than corpus-driven discourse analysis. In the corpus-driven approach, the researchers are guided by the corpus and enter the research blindly, prepared to interpret any information that may come up during the research and then formulate further steps (Sinclair, 2004, p. 190), while a corpus-based discourse analysis is where the corpus is selected and used in a way to fit the researcher’s existing set of goals and intentions and approach the corpus with these goals in mind. Though they have different meanings, it does not suggest that one research may follow only a specific form of analysis. But for the present study, the focus will be majorly on the corpus-based discourse analysis such that we can gather relevant discourse and come to a clear understanding of the concepts defined earlier.

The collected data is then processed to acquire statistical figures that will help in drawing decisive conclusions. The collected and annotated corpus is run through computer-aided software like Wmatrix, CQP web, Sketch Engine, or Lancs Box to name a few, for various linguistic analyses. The results derived through these analyses are further interpreted to find specific outcomes based on the following characteristics:

- i. Frequencies (content and function words)
- ii. Collocations
- iii. Semantic preferences
- iv. Key tags or keywords

7.1 Type of corpus

For this study, the most suitable way to extract quality content would be through the written corpus of international papers that have reported incidents or talk about freedom of expression of women. This would allow for a more methodical manner of inspecting and conducting a reliable language analysis of natural or ‘real world’ text with minimal experimental interference.

While working with the bounty of corpora, we would also look into the characteristics of the corpus to ensure optimal suitability for analysis in reference to the given topic. A few characteristics that are kept in mind during collecting the written corpus are the amount of data that can be processed through this method, the reliability of the content to cover the specific topic, the depth of the content, and the implicative power that can be derived through the analysis of the corpus. These contribute to the four major characteristics of corpus-based discourse and can be termed as extent, reach, depth, and strength (Hunston, 2011a).

7.2 Sampling Technique

For this particular research, the researcher intends to gather data, or corpora, from different online international newspapers. The newspapers that specifically highlight articles related to freedom of expression of women are used to attain more holistic analysis and results.

7.3 Sampling Type

Here, for our given goals and research questions, the best sampling type would be the monolingual balanced or sample annotated corpora (Biber, 1993 & Leech, 2007).

References

- [1]. Bakhtin, M. (1981). *The dialogic imagination: four essays*. (C. Emerson, C. & H. Holquist Trans.) Austin: University of Texas Press
- [2]. Benveniste, É. ([1966] 1971). *Problems in General Linguistics* (M. E. Meek Trans.). Coral Gables: University of Miami Press
- [3]. Biber, D. & Conrad, S. (2001). Register variation: a corpus approach. In D. Schiffrin, D. Tannen & H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 175-198). London: Blackwell.
- [4]. Biber, D. 1993. 'Representativeness in corpus design', *Literary and Linguistic Computing* 8 (4): 24357.
- [5]. Burr, V. (1995). *An introduction to social constructionism*. London: Routledge.
- [6]. Castells, Manuel and Pekka Himanen (Eds.) (2014). *Reconceptualizing Development in the Global Information Age*. Oxford: Oxford University Press.
- [7]. Council of Europe (2011). Commissioner for Human Rights, Media Pluralism and Human Rights (2011). available at <https://wcd.coe.int/ViewDoc.jsp?id=1881589>
- [8]. Foucault, M. (1972). *The Archaeology of Knowledge*. London: Tavistock Publications.
- [9]. Fowler, R. (1979). *Language and control*. London: Routledge & Kegan Paul.
- [10]. Hall, S. (1992). Discourse and Power. In S. Hall & B. Gieben (Eds.), *Formations of modernity* (pp. 291-295). Oxford, Open University Press.
- [11]. Halliday, M. (1966). Lexis as a Linguistic Level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In Memory of J. R. Firth*. London: Longman.
- [12]. Hoey, M. (1997). From concordance to text structure: new uses for computer corpora. In B. Lewandowska-Tomaszczyk and P. J. Melia (Eds.), *International conference on practical applications in language corpora: PALC'99* (pp. 2–23). Lodz, Poland: Lodz U. P.
- [13]. Hunston, S. 2011a. *Corpus Approaches to Evaluation: phraseology and evaluative language*. New York: Routledge.
- [14]. Lakoff, Robin. 1972. Language in context. *Language*, 48:907--924. 1975. *Language and Woman's Place*. New York: Harper and Row.
- [15]. Leech, G. 2007. *New resources, or just better old ones? The Holy Grail of Representativeness*. Doi: https://doi.org/10.1163/9789401203791_009
- [16]. McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- [17]. Media Watch (1995) *Women's Participation in the News: Global Media Monitoring Project*. Toronto: MediaWatch.
- [18]. Norris, Pippa (2004). *Giving Voice to the Voiceless: Good Governance, Human Development & Mass Communications*. Cambridge: Harvard University/John F. Kennedy School of Government.
- [19]. Norris, Pippa (2012). To Them that Hath' News Media and Knowledge Gaps. *Zeitschrift für Vergleichende Politikwissenschaft* 16 (1): s. 71-98.
- [20]. Partington, A., Duguid, A. & Taylor, C. (2013). *Patterns and meanings in discourse theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins Publishing Company.
- [21]. Rønning, H. (2009). 'The Contemporary Challenge to the Concept of Universal Human Rights and Freedom of Expression', in Kierulf, A. and Rønning, H. (eds.) *Freedom of Speech Abridged? Göteborg: Nordicom*.
- [22]. Sen, Amartya (1999). *Development as Freedom*. Oxford: Oxford University Press.
- [23]. Sen, A. (2004). 'Elements of a Theory of Human Rights.' *Philosophy and Public Affairs* Fall 2004: 32: 4: 315-356.
- [24]. Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- [25]. Sinclair, J. (2004). *Trust the Text*. London: Routledge.
- [26]. Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- [27]. Sun, L 2014, The role of diversity on freedom of speech in democratic societies, *International Journal of Sustainable Human Development*, vol. 2, Issue 2, p.44-51, accessed 15 October 2015 <http://hdl.handle.net/1959.14/303014>.
- [28]. Voltmer, I. (2015). 'Rethinking 'public service' in a globalized digital ecology.' *Open Democracy*. <https://www.opendemocracy.net/ourbeeb/ingrid-volkmer/rethinking-'public-service'-in-globalized-digital-ecology> (accessed November 10th 2015).
- [29]. Ziccardi, G. (2013). *Resistance, Liberation Technology, and Human Rights in the Digital Age*. Berlin: Springer.