

Doing Corpus Linguistics: Toward a Conceptual Framework for Indicator of Gender in English Language and Education

* Abdul Ghaffar Bhatti, Assistant Professor

** Muhammad Imran, Assistant Professor (Corresponding Author)

*** Muhammad Younas, PhD Scholar

Abstract

Technology plays a pivotal role in the ESL teaching and education sector. In language teaching, gender and language research mostly favors the idea of potential differences in language use between men and women. This paper explores different indicators of gender in the writing of males and females in a large subset of the British National Corpus (BNC) covering the domain of fiction with the application of the Corpus tool. Robin Lakoff's four key linguistic terms that mark female language have been used as benchmarks against which the study has been conducted. Previous researchers like Argamon, Koppel, and Shimon claim that females use more pronouns and a smaller number of nouns as compared to men. The hits and frequencies of Lakoff's terms and researchers' claims have been checked on BNC to get at the empirical findings. Taking general corpus BNC, corpus research method has been used to answer the research questions. The study found a substantial difference in the documents authored by male and female written text. It was also found that females use many more pronouns and males use many more nouns. Assumptions made regarding Lakoff's terms have been partially substantiated since the results vary a little concerning the use of empty adjectives like 'cute' and 'divine'. The work is a valuable addition to the existing corpus of knowledge about gender differences in language and it provides space for researchers to work in even broader perspectives.

Keywords: Education; technology; pronouns, color terms, empty adjectives, tag-questions, BNC.

Introduction

The relationship between gender and language has been a subject of fascination for many researchers primarily because it attempts to ascertain gender differences from a linguistic point of view. It has also helped feminists a great deal in dealing with the themes such as female subordination and marginality, oppressive male structures and patriarchy etc. using the parameters of language used by males and females. Language could be seen as a powerful indicator of diverse female marginalization as well as the establishment of oppressive patriarchal patterns in society. Not only does it reflect a particular mindset but also mirrors the social structures responsible for producing individuals thinking certain ways. Gender-differentiated language also allows one to approach the writer psychoanalytically, unearthing, thereby certain unconscious motives and desires operating in language. Analyses and interpretations of language from a gender-difference point of view usually lead us to think in the framework of binaries like male/female, dominance/subordination, and center/periphery and alike. Besides, the differences between male and female language are reflective of two distinct subcultures in which women have intrinsic qualities like grace, beauty, politeness, intimacy, delicacy etc. manifested through language.

For Lakoff, there is a great concordance between femininity and unassertive speech she defines as 'women's speech.' According to her, in a male-dominated society women are pressured to show the feminine qualities of weakness and subordination toward men. Thus, "it is entirely predictable and, given the pressure towards social conformity, rational that women should demonstrate these qualities in their speech as well as in other aspects of their behavior" (McAlinden & O'Leary, 1989, p.76). The present study explores the following four features in addition to the use of pronouns and nouns in fiction writings in British National Corpus (BNC). These features form a framework for testing the claim that women writers use the following more than the male.

* University of Education, Lahore, Multan Campus. Email: abdulghaffarbhatti72@gmail.com

** Govt College Havelian, Abbotabad. Email: Imran.pk756@gmail.com

*** School of Education, Soochow University, China. Email: younaskherani@stu.suda.edu.cn

1. Lexical hedges or fillers, e.g. *you know, sort of, kind of, you see.*
2. Tag questions, e.g. *isn't it? Won't he?*
3. 'Empty' adjectives, e.g. *divine, charming, cute, sweet.*
4. Precise color terms, e.g. *purple, crimson, red, green etc.*

Research Questions

1. What are the differences between male and female written documents in the use of pronouns and nouns?
2. How does BNC-based study validate gender differences in language as per Lakoff's key terms used for this study?

Literature Review

In Present-day British English, women have been shown to use fewer nouns and more personal pronouns than men. Rayson *et al.* (1997) show this for conversation; even in formal written texts, however, women seem to use more personal pronouns, while men use more nouns and certain types of noun specifiers (Argamon *et al.*, 2003). Haberman, Afzaal, Ghaffar, and Alfadda, (2020) argue that students of education may better gain insights from native etchers particularly related to lexical bundles. According to Argamon *et al.* (2003, p.321); Afzaal, (2020), these findings lend support to the notion that men's style is more 'informational' and women more 'involved'.

While dealing with the specifiers and use of pronouns, Halliday and Hassan (1994) presented that these are projected in nominal groups. Besides, through the active use of pronoun, a message is highlighted that identity is revealed to the reader whereas the specifiers reveals the information to the reader. The major difference in the writing of male authors and female reveals how these writers project the things, institutions, and objects. Biber (1998) elaborated the difference in his work related to the Seventeenth and Eighteenth-century and categorized it as the "involvement-informational dimension". He further stated that the exhibition of linguistics devices showed the identified involvement whereas men's writing projected as "informational". The contrasting views of relationships are apparent: negotiating with a desire for solidarity in women, maintaining status and hierarchical order in men. The frustration that occurs between women and men in conversation can be better understood "...by reference to systematic differences in how women and men tend to signal meaning in conversation" (ibid, p.7). Lakoff (1975) argued the role of language forms in the usage of both male and female authors and confronted the language produced by female authors. The meta-analysis on epistemic modality examines Lakoff's hypothesis about women's language through linguistic choices. Furthermore, several features of language devices were recognized for female authors which were present epistemic modality about a statement in an explicit manner (Holmes, 1983). Female authors employ these linguistic forms for three functions: to express genuine uncertainty about a proposition; to show politeness by way of alleviating the effect of the speech; or to apologize illegitimately for making a statement at all even though they are assertive about the facts (Holmes, 1975, p. 54).

Holmes (1990) exasperated to differentiate women's writing in terms of Lakoff's arguments about men and women's style of writing comprising of linguistic choices in their writings. An important step is to remedy flaws in terms of research on Lakoff's theories concerning men's and women's writings. First, it was imperative to recognize the linguistic choices together with their functions; second, it was crucial to formulate a scheme for the methods of collecting and analyzing data.

Besides, researchers such as Leaper and Ayres (2007) categorized the differences in gender in the communication through methodological choices such as observation length and gender of the first author as well as contextual factors such as group size and gender distribution of the group. However, the approach of meta-analysis did not follow the theoretical assumptions provided by Lakoff (1973, 1975, and 1977) which formed complicated results in gender differences studies later. To comprehensively examine previous findings with different methods and contexts, he investigated methodological and contextual moderators concerning differences in tentative speech between men and women.

Because the probability and size of gender differences may vary across different methodological devices, Leaper (2011) experimented with five methodological factors as potential moderators in the meta-analysis, of which the first was an operational definition. In view of the fact, Lakoff's (1973, 1975) arguments of gender differences in tentative language and taking into account

in particular that the possibility and extent of discrepancies/magnitude of variation between men and women may be affected across certain kinds of tentative language.

The second methodological choice Leaper (2011) investigated was the length of observation. Fagot (1985) claimed that the magnitude of an average difference of various behavioral patterns may become greater as the observation periods become longer.

Third, the recording method was examined as a possible methodological moderator that may affect gender differences in the speech forms. Moreover, Leaper and Ayres (2007) exposed that the accuracy of the result increased when the method of videotape was used in observation.

The gender of the first author is the fourth methodological variable that may moderate the outcome of a research. To validate whether this moderator may affect the force and direction of the findings (Anderson & Leaper, 1998; Eagly & Carli, 1981), Leaper (2011) explored the effect size of the author's gender by way of sampling.

Besides, Leaper (2011) considered the variation of tentative speech over the years on gender differences between male and female writing in language. Twenge (2001) noticed that assertiveness tended to smaller differences between the tentative speech of men and women.

Contextual Moderators

In addition to the aforementioned methodological variables, six possible contextual moderator variables were considered in terms of differences between male and female authors in tentative speech (Leaper, 2011).

The first moderator Leaper (2011) tested was the distribution of men and women within the group or the dyad. Past meta-analysis has demonstrated that if the differences of tentative speech between women and men indicate the social status and power relations, then the effect size increases when they are interacting with mixed-gender partners than that in same-gender ones (Carli, 1990; Hannover, 2000; Leaper & Ayres, 2007). As an alternative, if social rules and predispositions affect the gender difference, the magnitude of the average difference is greater in same-gender interactions (Carli, 1990; Leaper & Ayres, 2007; McMillan et al., 1977; Palomares, 2009).

Leaper (2011) investigated the interaction of same-gender partners against those of mixed gender as well as the connection between familiar partners and unknown ones to examine the prior findings and exposed/uncovered that the latter is associated with the expectation of consent for gender-typed social behavior (Deaux & Major, 1987; Leaper & Friedman, 2007).

To verify if the expectation of others' approval is related to gender differences, Leaper (2011) compared studies involving communications among strangers to those among familiar persons and anticipate a greater magnitude of differences between men and women in interactions between strangers.

The third potential moderator of gender differences in tentative language was the student status. To test whether the findings of prior research apply to other age groups other than college students, Leaper (2011) adopted the approach of sampling advocated by Leaper and Ayres (2007) to investigate gender differences between undergraduate and non-college participants in tentative speech and proposed a smaller gender-typed differences among college-age participants. However, the analyses on other potential factors that he could conduct such as social status, economic condition, and ethnicity were constrained by the number of available researches.

Fourth, group size is an important factor to consider when researching tentative speech because McMillan et al. (1977) implicated that variations in tentative speech tend to reveal gender differences in interpersonal intimacy. Besides, many studies have demonstrated the dyadic interactions establish interpersonal intimacy in tentative language, whereas group interactions foster competitiveness (Benenson, Nicholson, Waite, Roy, & Simpson, 2001; Leaper & Ayres, 2007; Solano & Dunnam, 1985).

Limitations of the present study include the range of moderators for differences between males and females because they were limited to variables of prior findings. Other possible moderators showed no significance across various researches and therefore they were not included in the present meta-analysis. On the other hand, if Leaper (2011) categorized the differences among various levels of variables, it would apply across a broad range of participants. Besides, the present study mainly adopted the approach of prepared tasks. Researchers such as Leaper and Ayres (2007) suggested that other approaches such as interviews and more gender-typed activities may be more revealing in

variations in tentative language. To sum up the limitations, Leaper (2011) did not explore the interaction effects of gender-typed differences in terms of multitudinous moderator factors.

Nevertheless, one of the main contributions of the current study is to build on prior knowledge to support Lakoff's (1975, 1977) claims concerning female's speech (Aries, 1996; Crawford, 1995; LaFrance, 2001) and to favor Lakoff's proposals about female's preference for tentative language as compared to male. This meta-analysis on moderator variables followed the assumptions that sensitivity and intimacy rather than an absence of assertiveness and competitiveness are reflected through tentative language. Furthermore, Leaper (2011) also addressed the significance of self-presentational factors/issues to scrutinize the gender differences in the speech.

Natural gender is equivalent to the sex of the individual and usually fuses with or is subordinate to other noun groups in countries such as France, Swahili, or Russia. However, people sometimes use it as specific pronouns and suffixes that are masculine, feminine, or neuter. Currently, the social tendency to gender equality has exposed differences between linguistic forms to open discussion and variation. The prior cross-cultural studies indicated that there were great gender differences of linguistic choices on the characteristics of sound, word forms, as well as lexis concerning pronominal use (Haas, 1944). Haas's study on linguistic variation showed that only aged women used archaic words in Koasati, whereas in Muskogee they only appeared in tales that include female speech. Gender differences of the utterer and the receiver may affect their linguistic choices. Previous qualitative studies usually indicate that this gender dichotomy is observed by all individuals, is frequently quoted, and is taught by caregivers. However, some latest quantitative investigations reveal that the truth is the other way round and is sometimes against the instincts of linguistic scholars and also the ordinary people. The institutionalized differentiation of the sexes is not familiarized by the community, and they are not taught to children directly.

Method

The Corpus approach is a rather recent technological application in language teaching and gender-related research. It provides empirical and verifiable data for a better understanding of both socio-cultural and sociolinguistic investigations in gender studies. To avoid any baseless speculation concerning the data analysis, the corpus-based approach is employed for getting precise and accurate results. The researchers have checked a large subset (fiction) of the British National Corpus (BNC) covering male and female writings. The study employed the software AntConc version 3.1.1 for the analysis of data. AntConc identifies concordances, frequencies, and keyword analysis of the language item, and compares the results with BNC corpus.

Results and Discussion

1. Color Terms

One assumption about the difference between male and female language is that women use more color terms than men do. This is because they are more interested in home decorations, fashions, dressing trends etc. To investigate the use of color terms by male and female authors, we searched the fiction section of the BNC. The search shows in table 1 the number of hits and frequency figures for a selection of color terms. Overall, it can be seen that the writings by women exhibit more use of color terms than male writings except for 'black' color as it seems the favorite color of men (as shown in table 1).

Table 1: male & female authors' use of color terms in the BNC (Fiction)

	Male authors'		Female authors	
	No. of Hits	Frequency	No. of Hits	Frequency
Black	3163	433.6	3104	369.86
White	3035	416.05	3250	387.26
Red	858	254.7	2022	240.94
Green	1438	197.13	1790	213.29
Pink	417	57.16	750	89.37
Purple	177	24.26	254	30.27
Crimson	72	9.87	143	17.04
Mauve	24	3.29	47	5.6
Indigo	12	1.65	33	3.93

Whereas the color terms like 'red', 'pink', 'purple', 'crimson' and 'mauve' have been used extensively by the female writers.

2. Hedges

According to Robin Lakoff, women use more hedges in their language. She links this claim with the notion that women by hedgings show greater insecurity and tentativeness. Hedges, however, can also be a sign of politeness and softness. (Holmes, 1986). After checking the frequency of the hedges like *sort of*, *kind of*, *you see* and *you know* were checked in the male and female writings, we got the following results.

Table 2: Frequency of hedges in Male writings

Hedges	No. of Hits	Frequency
sort of	2057	281.98
kind of	1600	219.33
you know	3778	517.9
you see	1156	158.47

Table 3: Frequency of hedges in Female writings

Hedges	No. of hits	Frequency
sort of	1913	227.95
kind of	1869	222.7
you know	4288	510.94
you see	1284	153.00

This comparison validates Lakoff's claim that females use more hedges in their language as compared to men. This leads to confirm the notion that women are more tentative and often uncertain in their arguments.

3. Tag questions:

The present study looked for the following tags in the BNC fiction section and came up with the results as follows.

Table 4: male & female authors' use of tag questions in the BNC (Fiction)

	Male authors'		Female authors	
	No. of Hits	Frequency	No. of Hits	Frequency
Can't you?	23	3.15	32	3.81
Don't they?	35	4.8	44	5.24
Isn't it?	396	54.29	509	60.65
Won't he?	3	0.41	4	0.48
Will you?	124	17.0	186	22.16
Is it?	163	69.5	168	71.73

The results shown above (table 4) fully confirm Lakoff's stance that women use more tag questions than men.

4. Empty Adjectives

Empty adjectives like charming, divine, cute, and sweet were checked in BNC to note the following differences in male and female writings.

Table 5: male & female authors' use of empty adjectives in the BNC (Fiction)

	Male authors'		Female authors	
	No. of Hits	Frequency	No. of Hits	Frequency
Charming	109	14.94	322	38.37
Cute	45	6.17	27	3.22
Divine	75	10.28	59	7.03
Sweet	496	67.99	890	106.05

The results shown above do not fully validate Lakoff's claim concerning the use of empty adjectives. This can be seen in the use of adjectives like 'cute' and 'divine' as they are used more by men than women. However, the results (in table 5) partially support Lakoff's assumption.

5. Pronouns and Nouns

Argamon et al in one of their study argue that women use more pronouns whereas men use more nouns. The results based on empirical findings of BNC fiction inquiries in the following table verify their claim.

Table 6: male & female authors' use of nouns and pronouns in the BNC (Fiction)

	Male authors'		Female authors	
	No. of Hits	Frequency	No. of Hits	Frequency
Nouns	1567699	214906.96	1634306	194738.75
Pronouns	967680	132653.76	1274907	151913.9

(The screenshots of all the aforesaid individual search queries have been attached as annexure.)

Conclusion

Possible variations between male and female writing styles have been explored in the perspectives of the key terms provided by Lakoff. Using a corpus-based approach, the study found significant variations in the documents written by male and female writers' use of nouns and pronouns. Moreover, it was observed through the analysis that both the writers have used certain linguistic devices to create a text of their own. The key linguistic terms introduced by Robin Lakoff concerning gender differences in language have been examined and validated through the study on the BNC (Fiction section). Findings showed a significant difference between male and female write up in the personal pronoun usage. Assumptions made regarding Lakoff's terms have been partially substantiated since the results vary a little concerning the use of empty adjectives like 'cute' and 'divine'. The general study of Argamon et al on the use of nouns and pronouns has also been confirmed through the analysis of the BNC. In the end, the study also focuses on the work to be conducted because of the large difference that occurred in the result which opens a new dimension for future research in the field of corpus linguistics.

References

- Afzaal, M. (2020). Book review: Kennet Lynggaard, Discourse Analysis, and European Union Politics (Palgrave Studies in European Union Politics). *Discourse Studies*, 22(5), 632–634. <https://doi.org/10.1177/1461445620921656>
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-linguistic Comparison* (Cambridge University Press, Cambridge)
- Biber, D., S. Conrad & R. Reppen, (1998). *Corpus Linguistics Investigating Language Structure and Use* (Cambridge University Press, Cambridge).
- Cameron, D., F. McAlinden & K. O'Leary (1988). Lakoff in context: the social and linguistic function of tag questions, in J. Coates and D. Cameron (eds.) *Women in their speech communities* (Longman, New York), pp. 74-93
- Coates, J. (ed.) (1998) *Language and Gender: A Reader* (Blackwell, Oxford)
- Halliday, M. A. K. (1994). *Introduction to Functional Grammar* (2nd ed.) (Arnold, London).
- Holmes, D. (1998). The evolution of stylometry in humanities scholarship, *Literary and Linguistic Computing*, 13(3), pp. 111-117.
- Holmes, J. (1984). 'Women's language': A functional approach, *General Linguistics* 24(3).
- Holmes, J. (1988). Paying compliments: A sex-preferential positive politeness strategy. *Journal of Pragmatics*, 12(3), pp. 445—465.
- Holmes, J. (1989). Sex differences and apologies: One aspect of communicative competence. *Applied Linguistics*, 10(2), pp. 194—213.
- Holmes, J. (1990). Hedges and boosters in women's and men's speech, *Language & Communication* 10(3).
- Holmes, J. (1993). Women's talk: The question of sociolinguistic universals, *Australian Journal of Communications* 20, 3.
- Haberman, P., Afzaal, M., Ghaffar, A., & Alfadda, H. (2020). Various Roles in the Development of EFL Learners' English Skills. *International Journal of Instruction*, 13(4),
- Koppel, M., Argamon, S. & A. R. Shimoni (2001). Automatically determining the gender of a text's author. Bar-Ilan University Technical Report BIU-TR-01-32.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change, *Language Variation and Change*.
- Lakoff, R. T. (1975). *Language and Women's Place* (Harper Colophon Books, New York).
- Palander-Collin, M. (1999). Male and female styles in 17th-century correspondence, *Language Variation and Change* 11, pp. 123-141.
- Rayson, P., G. Leech, & M. Hodges, (1997). Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the *British National Corpus*, *International Journal of Corpus Linguistics* 2, pp. 133—152.
- Tannen, D. (1990). Gender differences in topical coherence: Creating involvement in best friends' talk. *Discourse Processes*, 13, 73—90.
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich, *Language in Society*.