

Male-Female Discourse Difference in Terms of Lexical Density

¹Manizheh Alami, ²Maryam Sabbah and ³Mohammad Iranmanesh

¹Department of English and Literature, Islamic Azad University, Tabriz Branch, Tabriz, Iran

²Instructional Technology and Multimedia, University Sains Malaysia

³Department of Management, Main Campus, University Sains Malaysia, Penang, Malaysia

Abstract: The development of gender roles often begins as early as infancy. Being at the centre, gender manifests itself in any subtle and trivial aspect of our social life. From the time we are very small, it is ever present in any aspect of our life, in conversation, humor, conflict and so on. The overwhelming studies on the differences between men/women speech style represents the significance of the issue. The present study is an attempt to investigate whether the speaker's gender (being a male/female) contributes to the lexical density of their discourse. In other words, whether the lexical density of discourse is sensitive to the gender of the speaker? It is a data-driven, empirical study based on the transcribed recorded talk-in interactions between men and women. Text Content Analysis Tool (TCAT) was used to measure the lexical density of male/female speaker's discourse and to count the total number of words used by male/female speakers. The results of Chi-square test show that there is not a statistically significance difference between the lexical density of men and women discourse ($p > 0.157$). However, there is a negative relationship between the lexical density of discourse and discourse length. In other words, the more word counts (689 words) the lower lexical density (0.33.67%) and vice versa the less word used by the speaker (31) the higher lexical density of the discourse (90.32%).

Keywords: Discourse analysis, gender, lexical density, TCAT, word count

INTRODUCTION

In the past couple of decades, the studies on the difference between male/female linguistic behaviors increased dramatically. The scope of these studies covers variety of subjects, to name a few; the amount of speech delivered by each gender (Brizendine, 2006; Mehl *et al.*, 2007), the frequency of interruptions (West and Zimmerman, 1983, 1987), the politeness strategies used by each gender (Holmes, 1995) the use of minimal responses in male/female discourse (Maltz and Broker, 1982) the frequency of tag questions (Lakoff, 1975) the type of discourse markers (Ostman, 1983; Erman, 1987) and so on.

Likewise, the density of lexical in discourse has been of great interest for scholars. However, the majority of the studies conducted on Lexical Density (henceforth LD) are of comparative nature which pertain to the modality of discourse (written vs. spoken (Halliday, 1986)), different genres, e.g., business telephone conversation vs. radio state funeral commentary (Stubbs, 1986), different rhetorical structures, e.g., life discussion among students (Ure, 1971) and the conversation among students (Subhi and Johns-Lewis, 1989).

Objectives of the study: The present study is an attempt to investigate whether the speaker's gender (being a male/female) contributes to the LD of discourse or not. In other words, whether discourse LD is sensitive to the gender of the speaker (male vs. female)? Furthermore, our concern was to find out if there is a correlation between the length of discourse and its LD.

LITERATURE REVIEW

Sex/Gender dichotomy: The reason why some scholars confuse and conflate the terms 'sex' and 'gender' is due to the lack of a precise definition of sex/gender. Generally sex refers to those biological and physiological characteristics that define men and women, whereas gender accounts for those socially constructed roles, behaviors, activities and attributes that a given society considers appropriate for men or women.

- **Background studies on gender and language:** Lakoff (1975) article Language and Women's Place has been considered as a trigger to gender studies, while ushering the study of language and

gender in related disciplines for over three decades. Her pioneering work argues that women have a different way of speaking from men- a way of speaking that both reflects and produces a subordinate position for women in society. In her article she suggests that women's speech typically displays a range of features such as:

- Hedges
- Super polite forms
- Tag questions
- Question intonation in declarative statements,
- Speaking in italics
- Empty adjectives
- Hyper-corrected grammar and pronunciation
- Lack of sense of humor
- Direct quotation
- Special lexicon

In the past decades, the studies on gender and communication have fallen into two paradigms:

- i. Gender as cross-cultural difference
- ii. Gender as social power/dominance

- **Difference paradigm:** The advocates of the first approach believe that women and men speak differently because of fundamental differences in their relation to language, perhaps due to different socialization and early experience. Applying the Gumperz (1982) cross-cultural perspective the proponents of 'difference model' explain the differences in male/female language use in terms of cross-cultural differences. For Deborah Tannen, a well-known proponent of this approach, men's and women's styles are so different that she considers "cross-gender communication as cross-cultural." In her book 'You just don't understand' (1990) she posits that the main reason for the difference in men's/women's linguistic behavior is that men and women try to accomplish different things with talk. Men approach conversation as a contest, so they prefer to lead it in a direction in which they can take central role by for example telling joke, displaying information or skill, which she calls "report-talk" (public speaking). Whereas most women's conversation is a way of establishing community and creating connection, which she calls "rapport talk" (private speaking) (Tannen, 1990). She believes that men approach the world as a place to achieve and maintain status while women approach it as a network of connections to seek support and consensus. In a somewhat similar vein, American anthropologists Maltz and Broker (1982) in support of their stand with regard to two-culture model, (difference approach) argue that the main reason for a massive miscommunication in

male-female interactions is that they learn and use genderlects i.e., two separate sets of rules for engaging and interpreting conversation.

- **Dominance paradigm:** The 'dominance' model, a feminist oriented perspective, stresses that differences between men's and women's speech style arise because of the male's dominance over women which persists in order to keep women subordinated to men. Associated with this framework are studies conducted by Julia Penelope (1988), Spender (1981), Cameron (2007) and Fishman (1980) to name a few.

What is lexical density? Words can be classified as either content words or grammatical function words. Lexical items (L) are the major content words which carry information. They fall into four grammatical categories: (Yule, 2010)

- Nouns
- Adjectives
- Adverbs
- Main verbs

Grammatical items/ function words (G) serve to express relations between content words and are including: Auxiliary verbs, Modals, Pronouns, Prepositions, Determiners and Conjunctions.

The term LD often coincides with the notion of "information Packaging" (Johansson 2008). It could be measured either over the whole text or over clause. It seems that LD per text is more informative as it is independent of clause length. It is proved that the higher LD, the more informative the text and thus the more difficult to read it. In other words, sentences which are long and lexically dense are more difficult to read due to the fact that the information density of text depends largely to the amount of content words used in the text/discourse. Thus the sentence length and LD can affect readability and style as well.

According to Yates (1996) LD is "a measure of information density within a text.

Halliday defines LD as "a number of lexical items as the proportion of running words" (1985: 64). In his book (1994) Halliday argues that the written language is not only more complex in terms of lexical density than spoken language but they construe reality in different ways i.e., spoken language resembles dynamic aspect of reality while written language represents the effective account of the finished product.

Lexical Density can be defined as a percentage by the following formula:

$$LD = (N_{lex} / N) \times 100$$

LD = Analyzed text's lexical density

N_{lex} = Number of lexical words in the analyzed text

N = Total number of words in the analyzed text

Background studies on lexical density: Stegen (2007) investigation was an attempt to show whether the differences found between oral and written texts, in terms of LD, in other languages also hold for Bantu languages. He found that for Tanzanian Rengi language (Banto), the oral version of two narratives had a higher LD (56 and 54.7%) than the written versions (50.3 and 46.6%). He attributes these differences to the nature of Bantu language which is an agglutinating language. Having corroborated Halliday's claim of higher LD in written than in oral texts, he states that LD is probably more indicative of the colloquial vs. literary style than of oral vs. written medium. Through estimating LD in interviews and conversations with the same subjects, Subhi and Johnson-Lewis (1989) conclude that LD is higher in interview but the difference is not statistically significant. They refer to 8 factors that should be controlled in experimental studies of LD including:

- Basis for calculating LD
- Expected interruption and length of speaking turn
- Function of component units of text
- Self-consciousness/self-monitoring
- Personal attribute
- Group attribute
- Planning time
- Topic

In her study Johansson (2008) compares two measures of lexical density and lexical diversity among different age groups. She concludes that both lexical density and lexical diversity can be used for "modality and developmental differences" (p: 76), however, they cannot be used interchangeably. She further suggests that lexical diversity is "a better measure to be used for detecting differences between age groups than lexical density" (Ibid: 77).

Lexical density measurement: There are two kinds of methods to arrive at the LD ratio in the spoken/written discourse. The first approach is manual, whereby the status of all words in the text is specified by the analyst, after which percentages are worked out (Ure, 1971).

The second method is automatic and depends mainly on computer programs like the one devised by Stubbs (1986). The manual method, although is time consuming, has a greater degree of accuracy since each problem is dealt with by the human linguist in its real context. Automatic analysis based on tailor-made software, though sufficient and reliable to a great extent, is not without problem. Some of these problems are:

- Verbs such as can will occur as main verb or as a noun in certain contexts
- 2-In the case of phrasal verbs, the status of preposition or particle element is sometimes difficult to determine for example: (Halliday, 1967)

- She made up her face
- She made up her story
- They made up and kissed
- She made up the hill at speed
- Auxiliary verbs such as *be*, *have do* can be used as either grammatical or content words according to the grammatical contexts in which they are used
- One more general problem is in word classification. What one research counts as lexical, another may classify as grammatical. For example, Stubbs (1986) lists *be* as lexical/grammatical, while Ure (1971) counts it as grammatical

MATERIALS AND METHODS

The transcribed casual conversations presented in (Eggs and Slade, 1997) book *Analyzing Casual Conversation* constitute the data for the present study. Not living in English speaking countries, recording the native English speakers' conversations in naturally occurring settings was impractical for the authors. Thus, the authentic casual conversations presented in Eggs and Slade's book was used as the corpus. The conversations are real-life interactions of English (Anglo-Australian) speakers conversing in the informal and spontaneous situations. They are excerpts of casual conversations from a variety of contexts (e.g., tearoom at a hospital, at a coffee break at work, in a dinner party, in a parked car, in a lunch break among work mates and). They were recorded between 1983 and 1995.

To measure the lexical density of male/female speakers, the Text Content Analysis Tool was used. TCAT provides statistical information for the given text including; word count, unique words, number of sentences, average words per sentence, lexical density and Gunning Fog Readability Index. The statistical tool of SPSS (version 16) was used for analyzing the data.

Research design: This study was carried out on 50 excerpts from the everyday, casual conversations among 25 male and 25 female native English speakers. After each speaker's discourse being reported, the emerged texts were analyzed by TCAT. From among the information provided by TCAT the obtained LD for each speaker's discourse and the Total Word Count were used to answer the research questions. Having access to the LD of the speakers' discourse, the mean of LD for both groups of speakers was calculated. Table 1 illustrates the comparison between the mean of male/female speakers' LD in the corpus.

DISCUSSION

As illustrated in Table 1, the mean LD for the male speakers was estimated 59.69% whereas for female speakers it was 64.86%. Although the male/female LD differ sparingly, to make sure that this difference is not

Table 1: Lexical density of male/female speakers

Gender	Lexical density
Male	59.69%
Female	64.86%

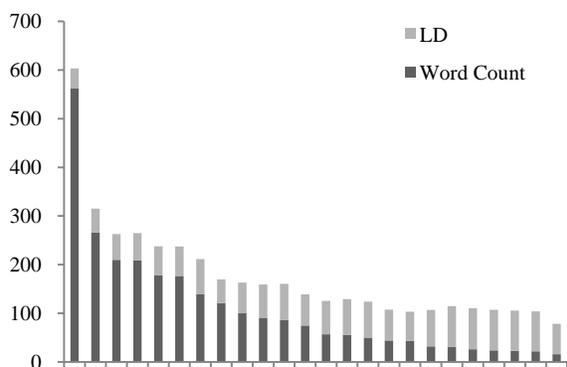


Fig. 1: The relationship between the number of words and LD in women discourse

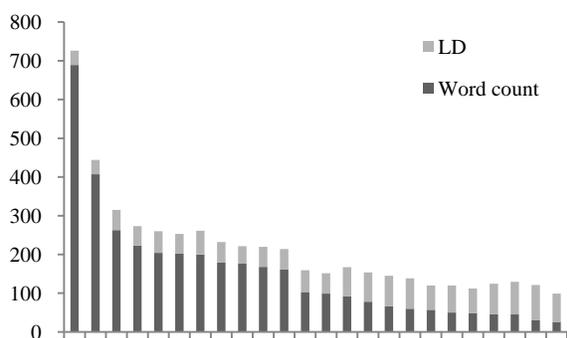


Fig. 2: The relationship between the number of words and LD in men discourse

meaningful the Chi-square test was used to make a crosstab comparison, the results of which show that there is statistically insignificant difference between the LD of their discourses ($p > 0.157$) (Appendix). In other words, the Lexical Density of a discourse is not sensitive to the gender of the speaker as an effective factor in speech style.

Another concern of the present study was to find out the relationship between LD and the length of

discourse. Figures 1 and 2 show the relationship between the numbers of words used in discourse and its LD in female and male speakers speech respectively.

Having a close look to the Figs 1 and 2 indicates an interesting fact regarding the relationship between the total number of words uttered by the speakers and the LD ratio of their discourse. As illustrates in these figures, there is a negative relationship between the total number of words employed by the male/female speakers and their discourse LD. In other words, in both gender groups the more words used by the speaker (689 words by men and 562 words by women), the lower LD of discourse (33.67% for men and 41.10% for women) and vice versa, the less number of words used by them (31 words by men and 16 words by women) the higher LD (90.32% for men and 62.50% for women). This negative relationship could pertain to the fact that the information density of the discourse depends mainly on the amount of content words used by the speaker. In other words, the length of discourse does not contribute to its information packaging. In general, an individual could express his/her meaning via a short discourse and be sufficiently informative. Furthermore, an extensive use of pronouns and discourse markers in combination with consecutive conjunctions are exclusively the features of spoken discourse.

CONCLUSION

Having a review of background literature on LD indicates that the difference between male/female discourses in terms of LD has not been addressed so far. Bearing in mind that measuring LD can have application in computer analysis of language, we attempted to compare LD of men/women spoken discourse, using the casual conversations among native English speakers. The results of the study show that male and female discourses are almost equally dense. In other words, the gender of the speaker has no effect on the lexical density of discourse. Another reading of the data indicates that there is a negative relationship between the length of discourse and its LD for both gender groups.

Appendix

	Chi-square tests				
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson chi-square	2.000 ^a	1	0.157		
Continuity correction ^b	0.000	1	1.000		
Likelihood ratio	2.773	1	0.096		
Fisher's exact test				1.000	0.500
Linear-by-linear association	1.000	1	0.317		
N of valid cases	2				

^a: 4 cells (100.0%) have expected count less than 5; the minimum expected count is 0.50; ^b: Computed only for a 2x2 table

REFERENCES

- Brizendine, L., 2006. *The Female Brain*. Morgan Road Books, US.
- Cameron, D., 2007. *The Myth of Mars and Venus*. OUP, Oxford.
- Eggins, S. and D. Slade, 1997. *Analyzing Casual Conversation*. Equinox, London.
- Erman, B., 1987. Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *J. Pragmatic.*, 33(9): 1337-1359.
- Fishman, P.M., 1980. Conversation Insecurity. In: Giles, H., W.P., Robinson and P.M. Smith (Eds.), *Language: Social Psychological Perspective*. Pergamon Press, Oxford, pp: 127-132.
- Gumperz, J., 1982. *Discourse Strategies*. CUP, Cambridge.
- Halliday, M.A.K., 1967. *Journal of Linguistics*. 3.1, 1967; 3.2, 1967; 4.2, 1968, Reprinted in full in Halliday, M.A.K., 2005, *Studies in English Language*, Volume 7 in the *Collected Works of M.A.K. Halliday*. Edited by J.J. Webster, Continuum, London and New York.
- Halliday, M.A.K., 1986. *Spoken and Written Language*. Deakin University, Victoria.
- Holmes, J., 1995. *Women, Men and Politeness*. Longman, London and New York.
- Johansson, V., 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers*, 53: 61-79.
- Lakoff, R.T. 1975. *Language and Women's Place*. Harper and Row, New York.
- Maltz, D.N. and R.A. Broker, 1982. A Cultural Approach to Male- Female Miscommunication. In: J. Coates (Ed.), *Language and Gender*, pp: 417-434.
- Mehl, M.R., S. Vazire, N. Ramirez-Esparza, R.B. Slatcler and J.W. Pennebaker, 2007. Are women really more talkative than men? *Science*, 317(82).
- Ostman, J.O., 1983. *You Know: A Discourse Functional Approach*. Amsterdam: John Benjamins.
- Penelope, J., 1988. *Prescribe Passivity: The Language of Sexism*. University of Nebraska-Lincoln, Faculty Publication, Department of English.
- Spender, D., 1981. *Man Made Language*. Pandora Books, New York.
- Stegen, O., 2007. Lexical Density in Oral vs. Written Rangi Texts. *SOAS Working Papers in Linguistics*, 1(15): 173-184.
- Stubbs, M., 1986. Lexical Density: A Computational Technique and Some Findings. In: Coulter, M. (Ed.), *Talking About Text*. English Language, Birmingham.
- Subhi, Z. and C. Johns-Lewis, 1989. Lexical density in interview and conversation. *York Papers Linguistics*, 14: 89-100.
- Tannen, D., 1990. You just don't Understand: Women and Men in Conversation. *Balantine Books*, New York, pp: 74-95.
- Ure, J., 1971. Lexical Density and Register Differentiation. In: Parren, G.E. and J.M. Trim (Eds.), *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics*. CUP, Cambridge, pp: 443-452.
- West, C. and D.H. Zimmerman, 1983. Small Insults: A Study of Interruptions in Cross-Sex Conversations between Unacquainted Persons. In: Throne, B., C. Kramarae and N. Henley (Eds.), *Language, Gender and Society*. New Bury House, Cambridge, MA.
- West, C. and D.H. Zimmerman, 1987. Doing gender. *Gender Soc.*, 1: 125-151.
- Yates, S.J., 1996. Oral and Written Linguistic Aspects of Computer Conferencing: A Corpus-Based Study. In: Susan, C.H. (Ed.), *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. John Benjamins Publishing Co., Philadelphia, pp: 37.
- Yule, G., 2010. *The Study of Language*. 4th Edn., Cambridge University Press, Cambridge.