

## Informational and Interactional Features of Written and Spoken Pakistani English

Behzad Anwar  
Saba Maryam

### Abstract

*Spoken language is different from written at various levels. These divergences help to form various genres and registers. This study conducts a Multidimensional Analysis (MDA) of Pakistani written and spoken English. For MDA analysis, a large corpus based on written and spoken data has been compiled. Biber's (1988) proposed multidimensional analysis model provides theoretical grounding for this present study. The study employs MAT software (version 1.3) for the individual and, later, the comparative analysis of written and spoken corpus. The results show that on D1, informational vs. involved, written text is informational and spoken data is interactional.*

**Keywords:** Multidimensional Analysis, Pakistani English, Written and Spoken Corpus

### 1. Introduction

This study explores the features of written and spoken Pakistani English by applying multidimensional model. Pakistani English has been explored by many researchers (Baumgardner, 1996; Talaat 2002; Mehboob, 2003; Anwar & Talaat 2011; Rehman, 2014). These researchers highlight the different linguistic, social and cultural features but no study has yet been done to explore the dimensions of Pakistani English through multidimensional model.

Written text is different from spoken (Biber, 1986). Written and spoken languages are different from each other for different reasons (Halliday, 1989). These two modes of language have been the topic of great interest for many researchers (Chafe & Danielewicz, 1987; Pawley & Syder, 1983). These differences among registers mark the boundary of written and spoken registers in the presence of variant linguistic features and these two registers depict distinctive terms and have divergent characteristics from each other.

By investigating the differences between written and spoken, the language forms a new path towards a new register (Biber, 1988; Biber et al.1999; Halliday, 1989). This study explores the features of written and spoken Pakistani English with the implementation of first dimension of multidimensional model by using written and spoken corpus. Multidimensional is an approach which is used for register analysis. This methodological approach is basically proposed for the comparative study of spoken and written register variation (Biber, 1986, 1988).

The focus of the present study is to investigate the variant linguistic features of Pakistani English. In Pakistani English, corpus MD methodological approach is used to highlight the informational and interactional linguistic features at dimension 1.

### Research Questions

This research aims to answer the following research questions:

1. How are written and spoken genres different from each other with reference to Pakistani English?
2. What kind of differences can be found in these genres regarding Dimension 1 of multidimensional analysis?

## 2. Literature Review

Written and spoken registers have been used by different scholars to establish the differences and similarities. The focus of register assessment shifts from contrastive analysis to dimension analysis. This dimension analysis was first practiced by Biber (1988). He used the written and spoken registers to investigate the dimensions and developed MD analysis. He himself applied it on academic corpus.

Academic discourse has been an area of great interest for register analysis. Many linguists explored academic register in different dimensions by applying multidimensional model (Biber, 1988; Nini 2013) especially University register by the implementation of multidimensional model. Biber (1988) provided a detailed interpretation of multidimensional analysis of academic prose of selected disciplines. It included data from different registers like, social sciences, pure sciences and medical. A comparative study was done to distinguish assorted dimensions of different registers.

In the last decade of 20<sup>th</sup> century, the focus was only on the written register. Atkinson (1992) used medical research articles for MD analysis. He used scientific corpus to develop the register characteristics of the data. The same perspective was elaborated by Biber and Finegan (1994). Their corpus was interpreted in relation to other texts. MD on University discourse with minor changes was also applied by Atkinson (1992). He collected the data from University as previous researches had been done in this particular register. All the researchers analyzed different dimensions of MD model but did not elaborate them completely.

Conrad (1994) analyzed professional academic journal under the umbrella of multidimensional approach. This study provided the development process of writing among the students of college. Conrad (1996) provided a thorough MD assessment of research articles and textbooks. It was the contrastive analysis from two major academic disciplines; ecology and American history.

Hyland's (1999) focused on the multidimensional features of Meta discourse of textbooks and research articles. The research was conducted on large scale for the MDA. Conrad (1996) evaluated a large collection of data for register understanding. It included common composition textbook selections, textbooks of ecology courses, and professional research articles of ecology.

Many scholars used the same model in many departments and disciplines of University (Faigley & Hansen, 1985; Horowitz, 1986). Historical development of language was also studied and its development in the field of particular disciplines (Halliday, 1988). Many other specialized discourses had been used by different writers (Biber & Finegan, 1994; Atkinson, 1992, 1996, 1999; Conrad, 1996, 2001) to explore multidimensional patterns.

Spoken and written variation in English (Biber, 2001) was the initiative step in the research studies of the present age. It is worth-mentioning that these studies do not entail separate MD analyses. These studies applied the dimensions to some new discourse domain, but they did not undertake new MD analyses i.e. involving a new factor analysis.

The present research deals with the comparative study of written and spoken data. This comparison brings the similarities and differences of spoken and written corpus in the form of dimensions.

### 3. Methodology & Data Collection

This study is based on corpus which comprises Pakistani written and spoken data. This corpus based research interprets data quantitatively. Corpus methodology provides the real/actual use of language. As this study focuses on the analysis of Pakistani English, for this very reason, corpus comprising of written and spoken Pakistani data has been compiled. The written corpus consists of Pakistani English novels by different Pakistani male and female writers. Short stories and edited works of fiction are also a part of written corpus. Spoken data is collected from different TV channels specifically “PTV WORLD”. Three different types of programs are part of it as, morning shows, political talk shows and sermons/ speeches by different religious persons.

#### 3.1 Corpus Building

For this research project, corpus (written and spoken) of Pakistani English is compiled. Written data consists of Pakistani fiction and spoken data consist of (Pakistani) English programs.

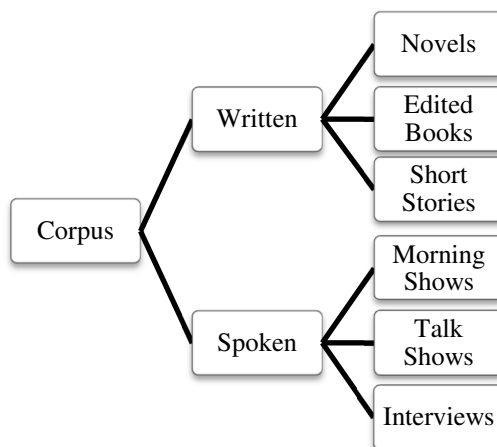


Figure 3.1: Presentation of Corpus Building

For this purpose data has been collected from different internet and other sources and converted into computer readable form. A few novels were available only in hard form; the researchers scanned those files first and saved them in PDF files. The remaining novels and stories were available in soft form. Written files were converted into plain text format from Microsoft word and PDF files.

The collection of spoken data was time taking and very laborious task. The researchers listened the programs first and then transcribed them. The transcribed data was transformed into plain text format. The total number of token (words) of each corpus (Written and Spoken) is presented in the following table

Table 3.1: Number of tokens in Written and Spoken Corpus

Corpus	Min.	Max.	Mean
Written	-22.23	19.45	-8.80
Spoken	14	22.89	17.35
Corpus Type	Category	Token	Percentage
Written Corpus	Novels	1,841,008	87
	Stories	184,179	8
	Edited Books	89,460	5
<b>Total Written Corpus</b>		2114647	
Spoken Corpus	Talk Shows	1350035	46
	Interviews	709535	24
	Morning Shows	909525	30
<b>Total Spoken Corpus</b>		2935864	
<b>Total</b>	5050511		

### 3.2 Research Tool

Multidimensional analysis tagger (MAT) has been used as a research tool. It tags the data and analyzes the data in input files. This programme is designed by Nini (2013) to process data according to Biber's model (1988). It is a methodological approach that applies statistical techniques to investigate variation in different genres and registers (Biber 1985, 1986, 1988, 2004).

## 4. Data Analysis and Results

The basic purpose of the study is the multidimensional analysis of the features of spoken and written Pakistani English and to highlight the differences of written and spoken Pakistani English on the basis of genres/registers. Biber (1988) establishes five dimensions for register variation in English. These are Involved vs. Informational production, Narrative vs. Non-Narrative concerns, Elaborated vs. Situation-dependent reference, Overt expression of Persuasion and Abstract vs. Non-abstract style.

### 4.1 Written Corpus at Dimension 1

Written and spoken are different genres and depict variation in style and content. It is a primary truth that written text is different from spoken (Biber, 1986). The table below shows result scores in their minimum, maximum and mean values.

**Table 4.1 Presentation of Dimension 1**

Corpus	Min.	Max.	Mean
Written	-22.23	19.45	-8.80
Spoken	14	22.89	17.35

The scores in the above table present the data in accordance with minimum and maximum scores. First dimension indicates negative scores and lies in the informational category, as the first dimension is the comparison of involved vs. informational. The first dimension shows negative scores which indicates that text is not interactional. It is informationally dense like academic prose. According to the result of Biber et al. (1999), informational text is prose or fiction and interviews while face to face discussions are the types of interactional text. The factors of informational axis are given below.

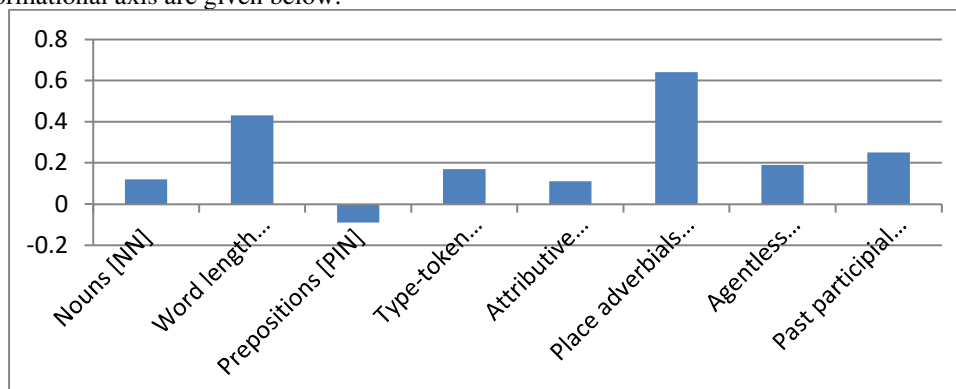


Figure 4.1: Linguistic Features and Scores (WC)

#### 4.2 Spoken Corpus at Dimension 1

Spoken corpus, as Biber (1986) mentions, is different from written corpus regarding dimensions. In the above table (4.1), positive values refer to the high scores of the Dimension 1. MDA presents linguistics items and these are illustrated with dimension scores and their values. Dimension 1 is the comparison of interactional vs. informational text. D1 indicates the positive value (17.35) which marks spoken text interactional. It is not informationally opaque but true interactive and communicative. Minimum, maximum and mean scores are all at positive scale. Spoken text is interactional and has high value at first dimension (Biber & Finegan, 1999). The factors of interactional perspective are given in the graph below which shows the positive value of dimension 1.

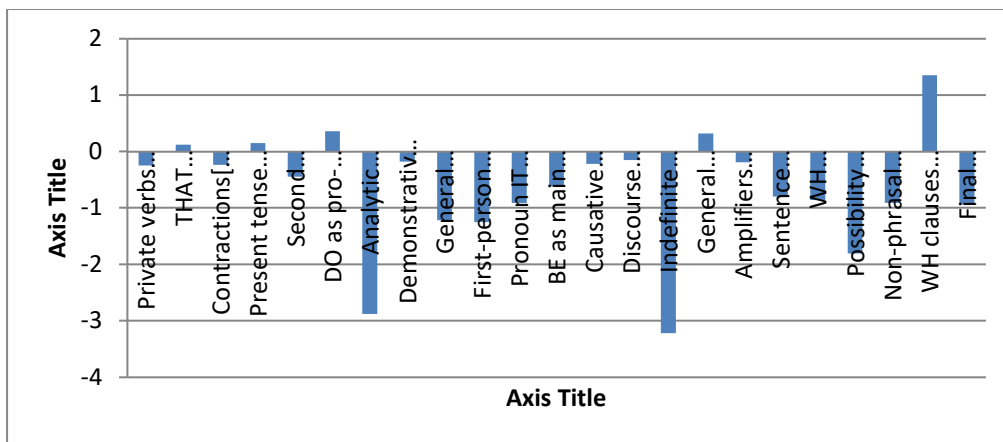


Figure 4.2: Linguistic Features and Scores of D1 (SC)

### 4.3 Comparison of Written and Spoken Corpus

Dimension 1 is theco-relation between interactional and informational axis which marks this particular corpus into one of the categories. The first dimension (involved vs. informational) shows negative scores which indicate that text is not interactional. The presence of noun (except nominalization, gerund and abstract noun) and adjective make it less non-interactive register. It is informationally dense like academic prose.

Table 4.2: Comparison of Written and Corpus at Dimension 1

Written Corpus	Talk Shows			Morning Shows			Interviews		
	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.	Mean
	10	12.99	12.03	6	12.96	11.55	13	22.89	22.76
Spoken Corpus	Novels			Stories			Edited Books		
	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.	Mean
	-23.76	-11.33	-20.14	-20.01	-7.77	-24.37	-33.06	-10.79	-22.75

According to Biber (1988, 2003) if any register is informational the other will automatically be interactive. In spoken register, dimension 1 is associated with interactive domain and the personal involvement of the interlocutors.

Interactive stance is measured through pronouns (first, second, it), stance (that-clauses with likelihood verbs and factual verbs, factual adverbials, hedges, private verbs) and formulaic language (contractions, that-omission, common vocabulary, lexical bundles). Biber (2003) also presents the same features in his study of University spoken register.

According to Biber (1988) and Biber et al. (1999), when these features (pronouns, present tense, that deletion) are presented particularly in any text, it indicates that the speakers aim to express ideas and opinions. It explicitly shows the involvement of the speakers in the communication process. These constructive features on Dimension 1 are associated with major functional domains such as inter-activeness and personal involvement (Biber, 2003).

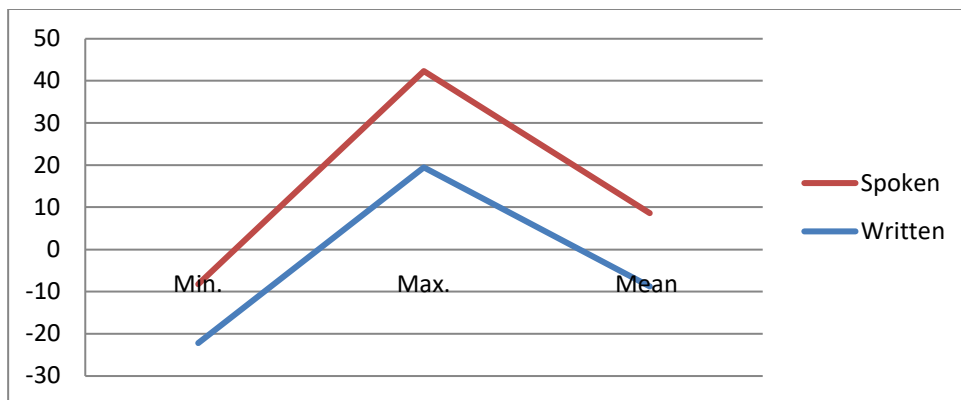


Figure 4.3: **Written and Spoken Corpus at D1**

All these features are presented in the above figures (4.1, 4.2). Present tense and pronouns are used by the speakers in the discussion which show their explicit interaction with each other. Audience also feels interactive as it is the discussion about present time issues. The above graph depicts the comparison of both the genres clearly.

## 5. Conclusion

According to Yule and Brown (1988), the basic purpose of the speech is to establish and maintain the relations of humans (those who are involved in conversation). We use language for dealings and we also interact with each other. The findings establish the fact that registers are different from each other on the basis of dimensions. The purpose of the researchers is also to highlight the dimensions which are the basic reasons of their differences (written and spoken). These differences are basically dimensions not oppositions. Dimensions are the complementary distribution of linguistic items which place any text into its genre type and specify its category.

D1 is the complementary distribution of informational and interactional text dimension. Written text is informational and spoken data is interactional. Biber's (1988) results also establish the fact that broadcast and face to face conversations are interactional and academic prose/fiction is informational.

The difference in both the corpus reflects the true relationship of both the registers in a quantitative manner. The difference, basically, treats the register in separate manner and highlights the differences in both kinds of registers.

## References

- Anwar, B. & Talaat, M. (2011). English in non-native context: Distinctive features of Pakistani journalistic English. *EFL Annual Research Journal*, vol 13.
- Atkinson, D. (1992). The Evaluation of Medical Research Writing from 1737 to 1985: The Case of Edinburgh Medical Journal. *Applied Linguistics*. Vol. 13, No. 4.
- Atkinson, D. (1996). The Philosophical Transactions of the Royal Society of London, 1675–1975: A Sociohistorical discourse analysis. *Language in Society*, 25, 333–371.
- Atkinson, D. (1999). *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675–1975*. Maywah, NJ: Lawrence Erlbaum.
- Baumgardner, R. J. (1996). Word Formation in Pakistani English. *English World Wide* 19:2, 205–246, John Benjamins Publishing Co.
- Biber, D. (1985). Investigating Macroscopic Textual Variation through Multi-Feature/ Multi-Dimensional Analysis. *Linguistics* 337- 60.
- Biber, D. (1986). Spoken and Written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, 384-414
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D., (2001). Variation among university spoken and written registers: a new multi-dimensional approach. In: Leistyna, P., Meyer, Ch.F. (Eds.), *Corpus Analysis. Language Structure and Language Use*. Rodopi, Amsterdam, New York.
- Biber, D., (2003). Variation among university spoken and written registers: a new multi-dimensional approach. In: Leistyna, P., Meyer, Ch.F. (Eds.), *Corpus Analysis. Language Structure and Language Use*. Rodopi, Amsterdam, New York.
- Biber, D. (2004). *Longman grammar of spoken and written English* (4. impr. Ed.). Harlow: Longman.
- Biber, D. & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, 9, 93–124.
- Biber, D. & Finegan, E. (1994). Intra-textual Variation within Medical Research Articles. In N. Oostdijk & P. de Haan (Ed), *Corpus Based Research into Language*. Amsterdam: Rodopi.
- Biber, D., Johansson, S., G. Leech, S. Conrad, & E. Finegan (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Chafe, W., & Danielewicz, J. (1987). *Properties of spoken and written language*. Berkeley, CA: Center for the Study of Writing.
- Conrad, S. (1994). *Variation in Academic Writing: Textbooks and Research Articles across Disciplines*. Paper Presented at the Annual Conference of the American Association of Applied Linguistics, Baltimore.
- Conrad, S. (1996). *Academic discourse in IIVD disciplines: Professional kiting and student development in biology und history*. Unpublished doctoral dissertation. Northern Arizona University.
- Conrad, S. (2001). Variation among disciplinary texts: A comparison of texts about American nuclear arms policy. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 84–93). London: Longman.
- Faigley, L., & Hansen, K. (1985). Learning to write in the social sciences. *College Composition and Communication*, 36. 140- 149.
- Halliday. M. (1988). On the language of physical science. In M. Ghadessy (Ed.). *Registers of written English* (pp. I63 17X). London: Pmter.



- Halliday, M. (1989). *Spoken and written language* (Vol. 2, illustrated, p. 109). Oxford University Press.
- Horowitz, L. (1986). *A Writer's guide to research*. Cincinnati, Ohio: Writer's Digest Books.
- Hyland, K. (1999). Talking to students: Met discourse in introductory course books. *English for Specific Purposes*, 18, 3–26.
- Mehboob, A. (2003). The English Language in Pakistan: A brief overview of its History and Linguistics. *Pakistan Journal of Language*, vol 4, no. 1.
- Nini, A. (2013). *Multidimensional Analysis Tagger 1.0-* Manual. Retrieved from: <http://sites.google.com/site/multidimensionaltagger>.
- Pawley, A., & Syder, F. (1983). *Two puzzles for linguistic theory: Native-like selection and native-like fluency*. In J. Richards and R. Schmidt (eds.), *Language and Communication*, London: Longman 191-226.
- Rehman, T. (2014). *Pakistani English: The Linguistic Description of Non Native Variety of English*, Islamabad, Institute of Pakistan Studies, Quaid-i- Azam University.
- Talaat, M. (2002). *Forms and Functions of English in Pakistan*. Unpublished Ph.D Thesis, Department of English Bahayudin Zakaria University, Multan.
- Yule, G., & Brown, G. (1988). *Discourse Analysis*. Cambridge: Cambridge University Press.