Tracing Linguistic Evidence of Cultural Diversity: A Corpus-based Study

Sadia Ali[1], Ebtesam Abdulhaleem [2]

Abstract
Various factors, including culture, indigenous languages, and social norms, shape English being spoken and written in a foreign context. To trace the influence of these various factors, several registers/ genres can be investigated. However, the newspaper register is the most suitable as it is close to everyday events to study how the language has undergone a change. This study seeks to investigate the influence of these factors on the English used in a foreign context and explores the patterns of linguistic variation in the corpus of Cultural news from South Asian countries. A corpus from South Asian countries was compiled. Biber's (1988) multidimensional model was used. ANOVA was applied to find the statistically significant differences. The results indicate that Cultural press reportage of all the countries varies in producing explicit, non-argumentative and abstract discourse. The results indicate that as the cultural press reportage of every country is different, English used in a foreign context differs in producing newspaper discourse.
Keywords: Culture, multidimensional analysis, press reportage, register analysis, South Asian Englishes.

Introduction

English has recently attracted a great deal of interest of researchers in South Asia. British English remained the benchmark for standardising English spoken and written in South Asia for a while. However, as time passed, the region's native languages, cultures, and sociopolitical influences began appropriating and indigenising the English language (Kachru, 1994). Linguistic studies have continued to centre on the language of print media. Recently, there has been a noticeable revival in interest. A triangle formed by language, culture, and media shows how each influences the other two while also being influenced by them (Rasul et al., 2016). The readership of the newspaper is quite broad. Newspaper language reflects society's language as a whole

[1] Department of English, College of Science and Humanities, Prince Sattam Bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia, sad.ali@psau.edu.sa
[2] King Saud University, Riyadh, Kingdom of Saudi Arabia, eabdulhaleem@ksu.edu.sa

(Westin, 2002), making it a valuable instrument for tracking language changes in society and providing clues about shifting social norms. So, for this study, newspaper language is selected to explore linguistic variations.

Moreover, there was a need to study the cultural differences regarding language. The study uses the most comprehensive approach, i.e., a multidimensional approach for studying linguistic variations. Multidimensional analysis refers to Biber's seminal study, which he conducted in 1988, comparing spoken and written registers of English. His study established five textual dimensions: Involved vs Informational Discourse, Narrative vs Non-narrative Concerns, Explicit vs Situation Dependent Discourse, Overt Expression of Argumentation/ Persuasion, and Impersonal (Abstract) vs Non-impersonal (Non-Abstract Style). This paper investigates the differences between cultural press reportage (CUR) of six South Asian countries, i.e., Bangladesh, Bhutan, India, Maldives, Nepal and Sri Lanka.

Literature Review

English holds a key position in South Asia's language planning and strategies due to its deep roots and importance in cultural contexts like education, intercultural communication, and several others (Kachru, 1997). It has deep roots in the region, with more speakers in South Asia than in the United Kingdom and the United States. The number of people who speak English in India alone, where the population has surpassed one billion, ranges from 200 to 333 million (Crystal, 2004; Basheer et al., 2022; Sanchez-Pay et al., 2022). Compared to native English speakers, there are an increasing number of non-native English speakers, with Crystal (1996) estimating that there are between 320 and 380 million of them in the Inner Circle, 150-300 million in the Outer Circle, and 100-1000 million in Expanding Circle. Similarly, Kachru (1996) asserts that non-native English speakers outnumber native speakers by a ratio of four (p. 241). Consequently, it is impossible to deny English's dominant position in South Asia when considering both the importance of the language and the number of people who speak it there.

Every word of English has its own norms and features. According to Brown (1995), three components are fundamental to Global Englishes. The first aspect is connected to the enormous variety of models that are available for English; the second is the idea that localised innovations in English are currently the common property of all its users; and the third is the fact that localised innovations in English are based on practical needs. Functional and genetic nativeness are

additional factors to consider in addition to these three. Kachru (2005) analyses the distinction between "functional nativeness" and "genetic nativeness" in the context of Global Englishes. He thinks that languages with a shared past, like Hindi and Bengali, which are both members of the Indo-Aryan family of Indian languages, share a genetic nativeness. Functional nativeness, however, is unrelated to linguistic genetic mapping. It more or less depends on the range and depth of a language in a community.

In contrast to the range, which is related to the functional domain, depth is related to social penetration.

Regarding functional nativeness, there are a few things to take into account:

• the standing that a variety has in society in the transplanted context

• the domains of function regarding language usage

• the ways that local identities are articulated through creative processes

• the language experts who support the process of acculturation in nativisation

• the contributions made to the new narrative through cultural merging

• the way that attitude determines labels for things

Functional nativeness was the deciding factor in South Asian culture. The variants of English are shaped by the diverse variety and depth of English in various South Asian cultures. These nations share many cultural and linguistic characteristics.

A significant body of research has been conducted on different registers using Multidimensional analysis. The works by Ali and Thompson (2022), Ali (2020), Shakir and Deuber (2018), Monaco (2017), Asghar et al. (2018), Biber and Egbert (2016), Hussain et al. (2016), Biber et al., (2015), Egbert (2015), Egbert and Plonsky (2015), Sardinha (2014), Biber & Conrad's (2014), Moskowich and Crespo (2012), Nesi (2008), Biber et al., (2006) are among some of the studies which analysed the language using MD approach. Research is scarce on the cultural category of South Asian Englishes. Only one study, i.e., Ali and Thompson (2022), analysed the cultural category of Afghan and Pakistani cultural reportage compared to British English. Nevertheless, there was a need to find the cultural and linguistic differences between other South Asian countries. So, this study fills the gap and analyses the cultural press reportage of six South Asian countries.

Data and Methodology

A specialised corpus of South Asian Cultural reportage was developed. For the compilation of the corpus, three newspapers from each South Asian country were selected. The cultural category is selected. The Daily Star, The Independent, Daily Observer from Bangladesh, Bhutan Times, The Bhutanese, Kuensel from Bhutan, The Times of India, The Telegraph, The Hindu from India, Maldives Independent, Mihaaru, Maldives Times from Maldives, The Himalayan Times, Kathmandu Tribune, The Kathmandu Post from Nepal and The Daily News, Adederana, Daily Mirror from Sri Lanka have been selected for the compilation of the corpus. Fifty texts from each newspaper were collected. Following is the detail of the countries and the selected newspapers:

Table 1. List of the countries and selected newspapers

Countries	Newspapers
Bangladesh	The Daily Star
	The Independent
	Daily Observer
Bhutan	Bhutan Times
	The Bhutanese
	Kuensel
India	The Times of India
	The Telegraph
	The Hindu
Maldives	Maldives Independent
	Mihaaru
	Maldives Times
Nepal	The Himalayan Times
	Kathmandu Tribune
	The Kathmandu Post
Sri Lanka	The Daily News
	Adaderana
	Daily Mirror

After cleaning the data, the files were marked up with specialized codes. Then the corpus of Cultural press reportage was developed. The corpus was tagged to identify the linguistic features using MAT tagger. The frequencies of the linguistic features were normalized. The next step was the computing of dimension scores. Then ANOVA was applied

to find the statistically significant differences between the cultural press reportage of each country.

Analysis and Discussion

Table 1 presents statistically significant differences among the countries regarding CUR across five textual dimensions introduced by Biber. On D1, Bangladesh, Bhutan, India and Sri Lanka show statistically significant differences. Moreover, there are no statistically significant differences between Maldives and Nepal as the table shows similar letters, meaning their significance value is more significant than 0.05. On D2, Bangladesh, Bhutan, India and Nepal show statistically significant differences. They have different mean sharing letters which show that the significance level is less than 0.05 indicating that there are statistically significant differences between them.

Table 2 Variation across S0uth Asian press reportage with reference to cultural category on five textual dimensions 88 MD analysis

Cultural	Dimension				
	D1	D2	D3	D4	D5
Bangladesh	-19.04 ± 0.86p	-0.91 ± 0.19j-m	4.05 ± 0.28abc	-2.44 ± 0.21mn	2.39 ± 0.25c-f
Bhutan	-17.68 ± 0.72jk	-0.97 ± 0.22ghi	6.37 ± 0.29abc	-1.70 ± 0.23i	1.97 ± 0.22ef
India	-9.58 ± 1.19o	-0.49 ± 0.26i-m	2.71± 0.33b-f	-1.58 ± 0.28k-n	2.06 ± 0.28d-h
Maldives	-18.24 ± 0.70n	0.07 ± 0.21hij	4.64 ± 0.37a-d	-2.43 ± 0.16kl	2.56 ± 0.23d-g
Nepal	-17.05 ± 0.92n	-1.03 ± 0.19jkl	4.18 ± 0.35bcd	-3.14 ± 0.23m	1.76 ± 0.27e-h
Sri Lanka	-18.51 ±0.93mn	-1.24 ± 0.31i-l	5.86 ± 0.28ab	-2.28 ± 0.27kl	1.80 ± 0.28efg

On D3, the difference between India and Nepal is statistically significant across CUR. Analysis of variance (ANOVA) results shows that variations between Bangladesh and India are statistically significant. Results of D4 show that the variations between Bangladeshi and Indian CUR are statistically significant. It is crucial to notice that on D5, there are statistically significant differences between all the countries, as the significance level is less than 0.05.

Further, figure 1 illustrates the results of CUR of South Asian countries regarding Biber's 1988 textual dimensions. The first dimension consists of a continuum ranging from positive "informational" discourse to negative "involved" discourse. CUR is often considered to include "culture and art news widely, where some of the news are of current culture and art events, some others consist of interviews with artists

or information about them" (Ozturk, 2014, p.137). The results of the present study conform to this claim. On dimension 1, the CUR of South Asian countries is informational instead of involved in discourse production.

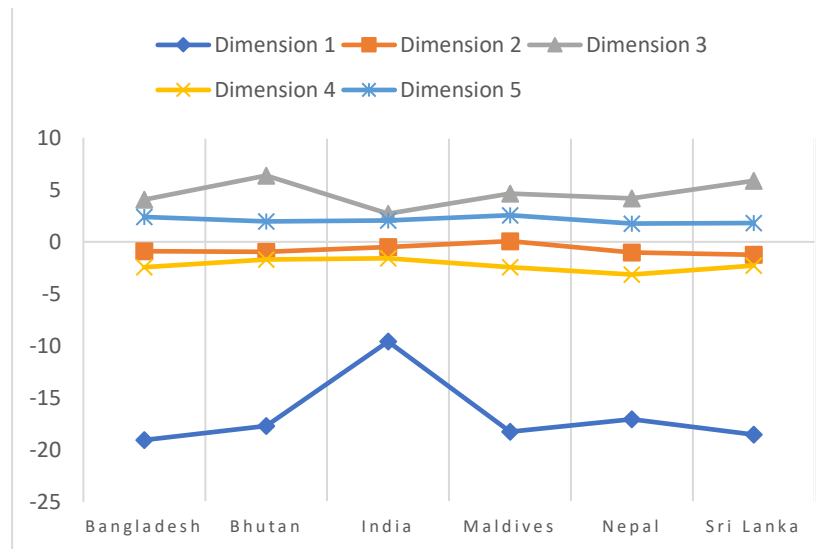**Fig.1. CUR of South Asian countries on five textual dimensions**

Figure 1 represents the CUR of South Asian countries. On negative polarity, Bangladeshi CUR is the most prominent in producing informational discourse, with mean scores of -19.04, respectively. In comparison, Sri Lanka and Maldives show slightly less informational discourse, with a mean score of -18.51 and -18.24, respectively. There is a slight difference between the mean scores of Bhutan, and Nepal (i.e. -17.68, & -17.05, respectively), showing informational discourse. CUR of India (- 9.58) is the least informational among these countries.

'Cultural press reportage (e.g., reportage of theatre or music events) is the most involved type of reportage' (Biber, 1995, p. 191). In his study, the sub-genre ranges from highly informative (financial press reportage, -20) to highly involved (personal telephone conversation, 40) (Biber, 1988, 1995). However, the results of the present study show that the cultural press reportage of all the South Asian countries shows informational discourse. This informational discourse follows the readership demands. Facchinetti et al. (2015) finds the competition among news agencies responsible for an increased focus on people and, consequently, on informational discourse.

A feature like nouns indicates the information focus of the text. Chafe (1985) defines preposition as 'a device for integrating information into idea unit' (p. 237), and attributive adjectives detailed information

1252

(Biber, 1988). These linguistic features performed a shared function of producing informational discourse.

On D2, positive scores indicate that discourse is primarily narrative, while negative scores show that its fundamental aim is non-narrative. Elliott (2012) found CUR texts as 'recollection of a first-hand observer' (p. 73). Table 1 provides evidence in favour of this claim by showing that on D2 (narrative vs non-narrative concern), most South Asian countries produce non-narrative discourse dealing 'with more immediate matters' (Biber, 1991, p. 109). Biber's (1988) study had no negative linguistic features. However, in this study, present tense verbs, the pronoun it, that deletion and place adverbial together produce non-narrative discourse. With a mean score of -1.24, CUR of Sri Lanka shows the highest negative scores indicating that their primary concern is non-narrative. In contrast, Maldives, with a mean value of (0.07), shows that its primary concern is narrative. Nepal (-1.03) is slightly less non-narrative than Sri Lanka. The mean scores of Bhutan (- 0.97) and India (- 0.49) also show fewer non-narrative concerns. Various cross-cultural factors in South Asia influence the language of South Asian press reportage.

The mean score of Sri Lankan CUR is maximum on negative polarity. The following example is taken from the Sri Lankan newspaper, Adraderana. The words in Italics are examples of linguistic features that produce non-narrative discourse.

Pre sident Maithripala Sirisena sa ys that Sri La nka has been a ble t0 0verc0 me a dark e ra and reillu mine unsurm0untable asp rati0ns f0r dem0cracy, fr eed0m, and human dignity. "H0wever, the r ealizati0n 0f the true meaning 0f freedom and independence li es in 0ur p0ten tial t0 free 0urseIves fr0 m the shackles of a II kinds and f0r ms of p0verty, ill--health, an d c0nfines 0f d ivisive identities of r ace, religi0 n and caste, an d enrich 0urseIves, a s a s0ciety th at values equity a nd rights 0f all." (CSSCUAT4)

Figure 1 indicates that on D3 (Explicit vs Situation Dependent discourse), India (2.71) produces the least explicit discourse among other South Asian countries. There is a slight difference between the mean scores of Bhutan and Sri Lanka (i.e., 6.37 & 5.86, respectively). Maldives (4.64), Nepal (4.18) and Bangladesh (4.05) show a slight difference in their mean scores displaying explicit discourse trend. Wh clauses are among the other features that produce explicit discourse. Shakir and Hill and Hengchen (2019) note that 'wh- pronoun-relative clauses are used for more exact and explicit reference' (p. 237).

In Biber's (1988) analysis, factor 4 has positive features. He suggests that these positive linguistic features together mark persuasion. Biber named this dimension as an 'Overt expression of persuasion/

argumentation'. In journalism, mostly argumentative writing is used (Steensen, 2019). However, the results show that third-person pronouns, private verbs and hedges, in this study, produce non-argumentative discourse. It is apparent from figure 1 that on D4, the CUR of South Asian newspapers has negative scores to varying degrees. CUR of Nepal (- 3.14) is the most non-argumentative in its discourse style. Interestingly, the mean scores of Bangladesh (- 2.44), Maldives (- 2.43) and Sri Lanka (- 2.28) are pretty similar. These countries show slightly less non-argumentative discourse than Nepal. Bhutan and India, with mean scores of -1.7 and -1.58, respectively, show the least non-argumentative discourse among other South Asian countries.

The words in Italics in the following excerpt from the Nepali newspaper, The Himalayan Times, highlight the presence of non-argumentative discourse.

Yo ungsters and child ren were seen wearing r ed Santa Cla us caps that re ad 'Merry Christmas' as t hey exchan ged greetings. Acc0rding t0 cen sus rep0rt, ab0u t three lakh pe0pl e in the c0untry ar e Chris tians. The g0vernment h ad ann0unced a national holiday t0 celebrate the festival. Mea nwhile, President Bidhya Devi Bhandari extend ed her Christmas greetings t0 all Nepali Christians at h0me and abr0ad, wishing them peace, pr0sperity and g00d health. (CSNCUHTT51)

Dimension 5 includes only positive features: adverbial-conjuncts, agentless passive verb, passive verb + by, passive post-nominal modifier and subordinate conjunction. These features together perform the function of producing an abstract style of discourse. CUR of all the countries shows positive scores to varying degrees. India (2.06) also produces abstract discourse. Maldives (2.56) and Bangladesh (2.39) show slightly less abstract styles. CUR of Bhutan (1.97), Sri Lanka (1.8) and Nepal (1.76) also produce abstract discourse.

So, the results indicate that all the countries are different in producing discourse regarding cultural press reportage. This linguistic recrafting of cultural press reportage might be attributed to the influence of cultural diversity, local languages, educational policies and historical background.

Future Directions
The non-availability of corpora of local languages and register studies on local or regional languages restricted the study to locate the influence of local South Asian languages on South Asian English. The corpus of this study may be helpful for the researchers working on

South Asian English. The results of the present study are significant in terms of providing a possibility to determine how cultural press reportage behaves linguistically in the South Asian context. The study also allows comparisons with future MD analyses in native and non-native contexts.

Acknowledgement:

Bibliography

Ali, S. (2020). Multidimensional Corpus-based Analysis of Newspaper Reportage: A Comparative study of Pakistani, Oter South Asian and British Newspapers [Unpublished doctoral dissertation]. Air University, Islamabad, Pakistan.

Ali, S., & Thompson, P. (2022). Reporting local cultures in the globalised world: How indigenised can English be in the free world? Corpora, 17(2), 197-224. https://doi.org/10.3366/cor.2022.0254

Asghar, S., Mahmood, M., & Asghar, Z. (2018). A Multidimensional Analysis of Pakistani Legal English. International Journal of English Linguistics, 8(5), 215. doi:10.5539/ijel.v8n5p215

Basheer, Z. M., Althahabi, A. M., Ali, M. H., Wafqan, H. M., Abd Al Mahdi, R., Al-Muttar, M. Y. O., & Bkhebukh, A. S. (2022). Determinants of Financial Performance: A Case from Oil and Gas Companies listed in the Iraq Stock Exchange. Cuadernos de Economía, 45(128), 34-44. https://cude.es/submit-a-manuscript/index.php/CUDE/article/view/258

Biber, D. (1988). Variation across speech and writing. Cambridge: Cambridge University Press.

Biber, D. (2006). Stance in spoken and written university registers. Journal of English for Academic Purposes, 5(2), 97-116.

Biber, D., & Conrad, S. (2014). Variation in English: Multi-dimensional studies. Routledge.

Biber, D., & Egbert, J. (2016). Using Multi-Dimensional Analysis to Study Register Variation on the Searchable Web. Corpus Linguistics Research, 2, 1-23. doi:10.18659/clr.2016.1.0.01

Biber, D., Egbert, J., & Davies, M. (2015). Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. Corpora, 10(1), 11-45. doi:10.3366/cor.2015.0065

Biber, D.(1995). Dimensions of register variation: A cross-linguistic comparison. Cambridge: Cambridge University Press.

Brown, N. (1995). Governing Prosperity: Social change and social analysis in Australia in the 1950s. Cambridge: Cambridge University Press.

Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. Literacy, language, and learning: The nature and consequences of reading and writing, 105, 105-123.

Crystal, D. (1996). The past, present and future of English rhythm. Speak Out, 18, 8-13.

Crystal, D. (2004). The Cambridge encyclopedia of the English language. Ernst Klett Sprachen.

Egbert, J. (2015). Publication type and discipline variation in published academic writing: Investigating statistical interaction in corpus data. International Journal of Corpus Linguistics, 20(1), 1-29. doi:10.1075/ijcl.20.1.01egb

Egbert, J., & Plonsky, L. (2015). Success in the abstract: exploring linguistic and stylistic predictors of conference abstract ratings. Corpora, 10(3), 291-313. doi:10.3366/cor.2015.0079

Elliott, A. (Ed.). (2012). Routledge handbook of identity studies. Taylor & Francis.

Facchinetti, R., Brownlees, N., Bös, B., & Fries, U. (2015). News as changing texts: Corpora, methodologies and analysis (2nd ed.). Cambridge Scholars Publishing.

Hill, M. J., & Hengchen, S. (2019). Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. Digital Scholarship in the Humanities. Retrieved from https://doi.org/10.1093/llc/fqz024

Hussain, Z., Mahmood, M. A., & Azher, M. (2016). Register variation in Pakistani English: A Multidimensional approach. Sci.Int, 28(4), 391-402. Retrieved from http://www.sci-int.com/pdf/636385732794496088.pdf

Kachru, B. B. (1997). The Alchemy of English: The Spread, functions and models of non-native Englishes. Oxford: Blackwell.

Kachru, B. B. (2005). Asian Englishes: beyond the canon (Vol. 1). Hong Kong University Press.

Kachru, B.B. (1994). English in South Asia. In R. Burchfield (ed.). The Cambridge History of the English Language. Vol. 5: English in Britain and Overseas: Origins and Developments 497–553. Cambridge: Cambridge University Press.

Monaco, L. M. (2017). A multidimensional analysis of late Modern English scientific texts from the Coruña Corpus (Doctoral dissertation, Universidade da coruna, Uk). Retrieved from https://ruc.udc.es/dspace/bitstream/handle/2183/19322/Monaco_Lei daMaria_TD_2017.pdf?sequence=2&isAllowed=y

Moskowich, I., & Crespo, B. (2012). Astronomy 'playne and simple': The Writing of Science between 1700 and 1900. Amsterdam: John Benjamins.

Nesi, H. (2008). A multidimensional analysis of student writing across levels and disciplines. Retrieved from https://warwick.ac.uk/fac/soc/al/research/collections/bawe/papers/n esi_baal08_abstract_formatted.pdf

Ozturk, R. G. (Ed.). (2014). Handbook of Research on the Impact of Culture and Society on the Entertainment Industry. IGI Global.

Rasul, A., Rasul, B. C., & McDowell, S. D. (2016). The Taliban factor: Conflict in Afghanistan and elite South Asian newspapers. The Journal of International Communication, 22(2), 273-292. doi:10.1080/13216597.2015.1106960

Sanchez-Pay, A., Ortega-Soto, J., Courel-Ibanez, J., & Sanchez-Alcaraz, B. (2022). Tennis Service in Profesional Women Players: Influence of Ranking Position and Players? Hand-dominance. International Journal of Medicine and Sciences of Physical Activity and Sport, 22(86), 255-267. https://doi.org/10.15366/rimcafd2022.86.004

Sardinha, T. B. (2014). 25 years later. Multi-Dimensional Analysis, 25 Years on: A Tribute to Douglas Biber, 60, 81.

Shakir, A., & Deuber, D. (2018). A Multidimensional analysis of Pakistani and U.S. English blogs and columns. English World-Wide, 40(1), 1-23. Retrieved fromhttps://www.researchgate.net/publication/329339900_A_Multidi mensional_Analysis_of_Pakistani_and_US_English_blogs_and_columns

Shakir, M., & Deuber, D. (2019). A Multidimensional study of interactive registers in Pakistani and US English. World Englishes. doi:10.1111/weng.12352

Westin, I. (2002). Language Change in English Newspaper Editorials. Amsterdam, Netherlands: Rodopi.