

European Journal of Linguistics

(EJL)

Word Formation in Bangladesh English



CARI

Word formation in Bangladeshi English

 Mir Sabbir Hasan

University of Bremen

<https://orcid.org/0009-0004-2930-7521>

Accepted: 9th Sep 2024 Received in Revised Form: 9th Oct 2024 Published: 9th Nov 2024

Abstract

Background: This paper aims to investigate the lexical properties of Bangladeshi English (BdE). Bangladesh does not have an official second language. Nevertheless, English has been present in Bangladeshi society for approximately four hundred years. The polity of Bangladesh is also silent about its position in the three-circle model of World English's and does not have any proper plan for the English language. Very few studies have been done on the English of Bangladesh from World English's perspective.

Methodology: Data is collected from five different sources; a survey, the Corpus of Global Web-Based English (GloWbE), the News on the Web (NOW), the Oxford English Dictionary, and one corpus compiled with 92 speeches of the former Prime Minister of Bangladesh. This corpus contains 101,490 words. The freeware corpus analysis toolkit Antconc is used to process the data of this corpus. The mixed method is followed in this work. Data is collected quantitatively and analyzed qualitatively.

Findings: The result of the study reveals that Bangladeshi English has rich lexical properties. These lexis are formed in different morphological processes. A good fraction of these lexis are noun, but adjective, verb, and adverb are also present.

Unique Contribution to Theory, Practice and Policy: These findings can be used as a starting point for further research.

Keywords: *World Englishes, Lexis, Corpus, Bangladesh, Bangladeshi English, GloWbE, NOW.*



Introduction

Lexical borrowing is a natural process in the spread of a language. In the case of Bangladeshi English, it began hundreds of years ago. The first documentation of Bangla words ('Supari') in English can be found in the Oxford English Dictionary (OED) in 1598. Bangladeshi English has a distinct body of vocabulary with loan words from Bangla. Bangladeshi English can also be characterized by productivity and creativity. Productivity and creativity are essential in describing lexis (Anesa 2019, 36).

Crystal viewed productivity and creativity as synonymous. He defined both terms as, "a general term refers to the creative capacity of language users to produce and understand an indefinitely large number of sentences. Most of which they will not have heard or used before. Seen as a property of language, it refers to the 'open endedness' or productivity of patterns, whereby a finite set of sounds, structures, etc., can be used to produce a potentially infinite number of sentences. It contrasts particularly with the unproductive communication systems of animals, and in this context is seen by some linguists as one of the design features of human language" (Crystal 2008, 122, 389). Štekauer refers to creativity as "any deviation from the productive rules" (2005, 224). In a similar vein, Plag argues that "the productivity of a word-formation process can be defined as its general potential to be used to create new words and as the degree to which this potential is exploited by the speakers" (Plag 2006, 127, as cited in Anesa 2019, 37)

Particular focus is devoted to neologism formations, as "among the several approaches available when investigating the lexicon of a language, the emphasis on neologisms allows us to gain a more detailed understanding of lexical developments. It also enables us to discuss the level of productivity and creativity present in different varieties, as well as the reciprocal lexical influences which emerge" (Anesa 2019, 4). According to Schmid, the process of developing neologism has three stages. Not all the stages need to happen one by one. These steps can be repeated by themselves (2008, 1-38 as cited in Anesa 2019, 42):

Lexicalization- by lexicalization, the structural development is referred to in this thesis. First, a form is created that expresses a new meaning sometimes in a specific context. Context can be of two types: user and the context of use. The word gradually assumes semantic meaning and can be described from a diachronic perspective

Institutionalization: Institutionalization is about socio-pragmatic development. At this stage, the word is being accepted by society. The word became context independent and started to be used by the community.

Hypostatization: This is the final stage and is about cognitive perception. At this level, the new word consolidates in the user's mind and becomes a part of the mental lexicon.[A4]

The states of the art

Unlike many other former British colonies, Bangladesh has not received critical attention from the WEs community. Interestingly, the mention of Bangladeshi English can be found in many early papers on World Englishes. In the historic conference at the East-West Center, Honolulu, Hawaii in 1978, Kachru and Quirk mentioned Bangladesh English as a variety of World Englishes (Smith ed. 1981, xiii). Kachru also listed Bangladesh in the "Outer circle" in his three-circle model (Kachru 1992, 356). Colonial politics influenced the Kachruvian approach to World Englishes. Bangladesh has a colonial past, so later scholars on this trend included Bangladesh in ESL varieties. In his circle of World English, Mc Arthur (1998, 97) included Bangladeshi English as one of the South Asian Standardizing English. Gargesh (2006) argued that Bangladesh English is one of the varieties of South Asian English. He also discussed a little about the role of English in radio broadcasts in Bangladesh. Melchers et al. (2001) listed Bangladesh English in the outer circle.

Crystal (2003) referred to Bangladeshi English as one of the several varieties of English that emerged throughout the subcontinent. Durkin has worked on loanwords in English. In this study, he dedicated a short chapter on the languages of South Asia. Bangla has referred to this chapter but not discussed it explicitly (2014). The researchers did not discuss the lexical aspects of Bangladeshi English

Data and Methodology

1. Source of Data

The study presented here draws on data from five different sources:

1. Corpus of Global Web-Based English (GloWbE)
2. News on the Web corpus (NOW)
3. Bangladesh Corpus.
4. The Oxford English Dictionary (OED)
5. Survey

Corpus of Global Web-Based English (GloWbE)

The Corpus of Global Web-Based English (GloWbE) is one of the largest available corpora with 1.9 billion words of text from 22 different countries; Bangladesh is one of them. GloWbE is an online corpus with a unique interface that allows several types of searches. This corpus is unique because it allows comparisons between different varieties of English (Davies 2013). A large corpus like GloWbE is beneficial for studying lexical innovation given the expected low frequency of lexis. In this thesis, GloWbE was extremely useful for verifying purposes and checking frequency and data comparison.

Table 1: Outline of GloWbE Bangladesh.

General (may also include blogs)			(only) Blogs			Total		
Web sites	Web pages	Words	Web sites	Web pages	Words	Web sites	Web pages	Words
4415	30,813		2,332	14,246		5,712	45,059	
		28,700,158			10,922,869			39,658,255

News on the Web corpus (NOW)

The News on the Web (NOW) corpus is one of the largest available corpora with a massive 14.1 billion words. NOW is a newspaper-based corpus. The data is taken from web-based newspapers and magazines from 2010 to the present time. The corpus grows by about 180-200 million words each month (from about 300,000 new articles), or about two billion words each year (Davies 2013). NOW is different from GloWbE in that sense; GloWbE contains both formal and informal data but NOW contains purely formal data. The developer compared NOW with Google trends. He said,

Google Trends shows you what people are searching for, the NOW Corpus is the only structured corpus that shows you what is actually happening in the language -- virtually right up to the present time. For example, see the frequency of words since 2010, as well as new words and phrases from the last few years" (Davies 2013).

NOW Bangladesh has over 85 million words.

Table 2: Outline of NOW Bangladesh

Number of sources	Number of articles	Years
26	93994	2010- on going

The GloWbE and NOW are appropriate for verification purposes as both corpora provide the frequency of any word of all 22 sub corpora on a single page. GloWbE and NOW are also helpful to compare data among varieties. GloWbE and NOW have an interface where users can compare data of one variety with the other 21 varieties or across the groups of varieties. Finally, the apparent reason for choosing GloWbE and NOW, till the date of this research, GloWbE and NOW are the only corpora that contain Bangladesh components.

Bangladesh Corpus (BD Corpus)

GloWbE and NOW are based on written materials from websites and blogs. Investigating spoken discourse is also vital for the analysis of new lexical items. The International Corpus of English (ICE) has transcribed spoken texts. But the Bangladesh variety is not available in ICE.

To bridge the gap, a corpus is compiled consisting of 91 speeches of the previous Prime Minister of Bangladesh, Sheikh Hasina Wazed. This corpus contains 101,490 words. Sheikh Hasina Wazed is the 10th Prime Minister of Bangladesh, having held office since January 2009. She had previously served as the Prime Minister from 1996 to 2001. She is also the President of one of the largest political parties of Bangladesh, the Bangladesh Awami League since 1981 (Augustyn 2019). Sheikh Hasina Wazed is the daughter of the Father of the Nation Sheikh Mujibur Rahman. According to Forbes, Sheikh Hasina Wazed is the world's 43rd most powerful woman ("The World's 100 Most Powerful Women List" 2021). Usually, she gives her speech in Bangla. However, sometimes, especially in different seminars, diplomatic visits, and meetings of different international organizations, she delivered speeches in English. Usually, Prime Minister Sheikh Hasina gives a speech addressing the nation every year. It is about different achievements and activities of the government. The government-sponsored National News Agency of Bangladesh (BSS) provides an English version of her address to the nation.

Often speeches of the PM are pre-written. There is no information available if these speeches are pre-written or not. So it cannot be claimed as pure data of spoken language. This paper will mark these data as written to be spoken data.

To ensure validity and reliability, all of the data were retrieved from the official websites of the Prime Minister's office, United Nations, OIC, Permanent Mission of Bangladesh to the United Nations, and the National News Agency of Bangladesh (BSS). The table illustrates the word count of different sources:

Table 3: Word count of BD Corpus.

Website	Number of words
The Prime Minister's office	70,433
UN/ Permanent Mission of Bangladesh to the United Nations	17,537
OIC	1,187
BSS	12,333
Total	101,490 _[A6]

The timeframe of these speeches is from 2009 to 2022. AntConc software is used to process data in the BD Corpus.

Survey

NOW and the Bangladesh corpus contain formal data. GloWbE houses a mixture of formal and informal data. So, a survey is conducted to get more raw information. Google form is used to do this survey. One hundred six native speakers of Bangla participated in this survey. The participants are high school students, university students, or fresh graduates who are active in social media and blogging. They were asked the following five questions:

1. Which language do you use in your daily activities?
2. Do you use Bangla words in English writing or speech in a formal context?
3. Do you use Bangla words in English in an informal context? Such as in social networks, messaging, or blogs.
4. Do you intentionally use any misspelled English words?
5. Comments/Give some examples of such Bangla words, and misspelled English words.

The main objective of this survey was to collect words. This question is left to the very end to avoid the observer paradox. Two hundred eighty-seven words are collected from this survey.

6. The Oxford English Dictionary (OED)

The Oxford English Dictionary (OED) is widely accepted as the authority of the English language. It is the principal historical dictionary that traces the historical development of the English language, providing a comprehensive resource. OED houses 600,000 words— past and present— from across the English-speaking world. The OED was first published in 1884 by the Oxford University Press. The first electronic version was made available in 1988 and the online version in 2000. The existing entries are revised every three months and add new words (Oxford English Dictionary).

The dictionary's online version has an advanced search option that allows users to browse words based on origin. Fifty words (as of December 2021) are found with Bangla origin. The OED has an interface where users can see which words are frequent in which variety of English. But, the words of Bangla origin do not have this information. So, it cannot be claimed that all the 50 words are affiliated with BdE. In this paper, only those words are considered as part of BdE if it matches one of the following conditions-

1. If the words are more frequent in GloWbE Bangladesh or NOW Bangladesh than in other varieties,
2. If the word appeared in Bangladesh Corpus at least once,

3. If the words appeared in the survey.

So, in short, the type of data of this paper:

Table 4: Data type of the thesis.

Data type	Informal and formal written	Formal written	Formal written (to be spoken)
Source	GloWbE, Survey	NOW, OED	Bangladesh corpus

Besides the sources mentioned above, in some specific cases, another material (taken, for example, from newspaper articles) has also been used to investigate the use of certain neologisms and syntactic structures.

1.2 Research Methodology

The thesis aims to investigate the lexical properties of Bangladeshi English through qualitative and quantitative analyses. The qualitative part is collecting and identifying relevant data. The quantitative part includes counting lexis and calculating their frequency. The following are the steps on how to conduct the research.

Step 1. Identification of correct data

Several methods have been developed to investigate lexis over time. Previously, it was predominantly manual. With the advancement of computer sciences, at present, the popular approach is automatic or semiautomatic. The semiautomatic method is followed. Quite diversified data is collected from five different sources. Participants of the survey provided 287 words. Many words are English, Urdu, and Arabic. Eighty-four words were collected from the Bangladesh corpus, and the Oxford English Dictionary (OED) listed 50 words that originated from Bangla. The GloWbE and NOW have a unique interface that allow to search words the most frequent words of a variety. Several words were collected from The GloWbE and NOW. First, the possible candidate of the lexicon of BdE is identified. The words are not considered if,

1. Appearing in one source, the frequency of the words in GloWbE Bangladesh or NOW Bangladesh is not among the top 3.
2. If the word is frequent in other varieties,

Step 2. Classification of data.

The WE scholarly community has developed several methods to show the lexical category of English and its varieties. Schneider (2011, 209) discussed ten categories:

A. Broader category

1. Name of indigenous Plant and Animal.
2. Objects and customs alien to British settlers.

B. Based on the first two

1. Physical objects
2. Food
3. Clothing
4. Social standing and customs
5. Term for religion and superstition

C. Etymologically

1. Innovation
2. New compound
3. Derivation

In his study focusing on the translation of neologisms (1988,150), Newmark proposed twelve types of lexis:

A. Existing lexical items with new senses:

1. Words
2. Collocations.

B. New forms:

1. New coinages
2. Derived words
3. Abbreviations
4. Collocations
5. Eponyms
6. Phrasal words
7. Transferred words (new and old referents)
8. Acronyms (new and old referents)
9. Pseudo-neologisms
10. Internationalism.

The lexical classification analyzed here is based on Schneider (2011) and Newmark (1988). Based on their classification, the main categorization taken into account in the analysis of this paper are the following:

A. Broader category

1. Name of Place, indigenous community, plant, and animal
2. Objects and customs alien to the inner circle
3. Objects, customs, and concepts exist in the inner circle but still use the indigenous term.
4. English words used in a different sense from British & American English,
5. English words related to administration which are uncommon or obsolete in British and American English,
6. English words related to the recent economic outburst.

B. Etymologically

1. Affixation
2. Compounding
3. Hybrid
4. Abbreviation
5. Acronym
6. Blending
7. Word borrowing
8. Calque

C. Based on sources

1. Appeared in four/five sources
2. Appeared in two/three sources
3. Appeared in one source
4. *Step 3. Calculating Quantitative data*

After classifying the data, the frequency of the lexicon is calculated in the Corpus of Global Web-Based English (GloWbE), Bangladesh corpus, and the News on the Web corpus (NOW). The corpus based approach has some potential limitations.

Firstly, there are no encoding standards and an agreed spelling system of BdE. There are a few spellings for a single word in practice. Such as Riksha/Rickshaw, Iztema/Istema/Ijtema, Qawmi/Kaomi. The ‘customized word list’ function of the GloWbE and NOW has been used to resolve this issue. This function was proven very useful for this purpose. It shows the combined frequency of several words.

Secondly, one major problem regarding the frequency is the source. One word can appear ten times in the corpus, but all the words or most of the words can be from a single article, website, blog, or writer. For example, the word ‘Grameen’ (something of or from the village) is one of the most frequent words in GloWbE Bangladesh and NOW Bangladesh. The frequency of Grameen in GloWbE is 144, and in NOW, it is 11.3. However, when contexts are checked, most of the word in GloWbE is from a few websites of an investment corporation called Grameen Foundation, and in NOW, most of the newspaper articles are also about a bank of this foundation called Grameen Bank. So, this study does not consider any word that appears multiple times in single sources.

Thirdly, the problem is regarding the plural markers. As these words are not recognized, the corpus machine cannot mark the plural form of these words. For example, GloWbE and NOW both show ‘parishad’ (Office) and ‘parishads’ as different words, but actually, parishads is the plural form of parishad. So, words are searched with all possible forms of spelling grammatical markers.

Fourthly, the English words used in a different sense and English words related to administration are addressed. These words are used in both local and original meanings. For example, the word DC is the abbreviation of the word “Deputy Commissioner” and the name the capital of the United States. Both uses exist in the corpus. In such a scenario, a random sample of 100 words is taken from the corpus and calculated the frequency of the uses. Both GloWbE and NOW have an interface that allows generating a random sample of 100 words. These words are marked with the star *.

And finally, several words are also proper nouns. For example, the word Alim (scholar of Islam) is also a common noun for Bangladeshi boys. Random sampling and star marking are used in this case as well.

Based on frequency, the lexis is divided into three categories:

1. High frequent words
2. Frequent words
3. Less frequent words

For statistical reflection, the frequency of BD corpus has been shown per 100,000. And, all the numbers are rounded to the nearest integer unless otherwise specified.

Step 4. Illustrating the shared lexis of Bangladeshi English & Indian English and Bangladeshi English & Pakistani English.

As mentioned in the previous chapter, GloWbE has data from 20 different varieties, including Bangladesh, India, and Pakistan. GloWbE allows users to compare data between varieties. This feature of GloWbE is used to address this. Moreover, in the time of word quantification, the presence of the words in Indian and Pakistani English is also checked.

Analysis and Result

I found three hundred thirty-nine words that can be described as unique to Bangladeshi English. Words from all the categories are found in this study. In terms of acceptance, the lexical items used in Bangladeshi English are primarily of two types. The first class consists of those lexical items which have been assimilated across varieties of language, specifically in British English. The latest (as of January 2024) online edition of the Oxford English Dictionary (OED) listed 50 words that originated from Bangla.

The second type of lexis comprises those items which have not necessarily made their way to British English. These lexis are used frequently in Bangladeshi English. A large majority of this second class of words are not included in any dictionaries of English and till today, there are no glossaries or dictionaries available focusing on Bangladesh English.

A pattern can be drawn from the frequencies-

1. Most of the high frequent words appeared in all five or four sources. These words can fall into Schmid's hypostatization stage. These words are consolidated in the mind of the users and become a part of the mental lexicon. Such as Deputy Commissioner, Iftar, in charge, Readymade.
2. The frequent words mainly appeared in three or two sources. These words would fall into Schmid's institutionalization and hypostatization stage. Some words are already accepted, and others are accepted by society. The words are becoming context independent and started to be used by the community. Such as Auto-rickshaw, Black money, Beel.
3. The difference between frequent and less frequent words is quite difficult to explain in terms of nature. Most of these words appeared in two or one source. However, these words can fall in the hypostatization stage on the practical level. For example, the word 'Remittance fighter' is only found in the NOW corpus with a frequency of 0.07. But, this word is trendy. Other words like ain (law) and Brinjal (Eggplant) are not expected to appear frequently. Such as Adhunik, Adibasi, Alu.

Conclusion

The paper appears to affirm that Bangladeshi English has a considerable number of lexis. These words are the result of sophisticated morphological processes. These lexises are of different word classes but are not as versatile as British or American English. A large fraction of these lexis are nouns. Adjectives, adverbs, and verbs are also present and emerging.

The primary limitation of this study was the lack of resources, both corpus, and literary resources. Several words are found while looking for resources used by Bangladeshi writers in Bangladeshi English literature, newspaper reports, and journal articles which could not be tested for lack of corpus resources. Those words are not included either in NOW or GloWbE; for example, *pera* (pain), *Bengalisization*, *showing high court* (make someone a fool), *petni* (female ghost), *Pori* (Very beautiful), *BNCC* (critical people). The lack of spoken data is also worth mentioning. The BD corpus compiled to bridge the gap consists of slightly just over 100,000 words. Very little study has been done in this field so far. This research is the most descriptive empirical study in the field of BdE.

This paper suggests that the BdE has rich lexical properties. These findings can be used as a starting point for further research. One possibility is to work on the dictionary/glossary of Bangladeshi English (BdE). Three hundred thirty-nine recorded words of this study will be pretty helpful for a lexicographer.

References

- Anesa, Patrizia. *Lexical Innovation in World Englishes: Cross-Fertilization and Evolving Paradigms*. Routledge, Taylor Et Francis Group, 2019.
- Crystal, David. *A Dictionary of Linguistics and Phonetics*. Blackwell, 2008.
- Crystal, David. *English as a Global Language*. Cambridge University Press, 2003.
- Davies, Mark. *Corpus of News on the Web (NOW)*. 2016-, Available online at <https://www.english-corpora.org/now/>.
- Davies, Mark. *Corpus of Global Web-Based English*, 2013, Available online at <https://www.english-corpora.org/glowbe/>. Accessed 19 February 2021.
- Durkin, Philip. *Borrowed Words: A History of Loanwords in English*. Oxford University Press, 2014.
- Gargesh, Ravinder. "South Asian Englishes." *The Handbook of World Englishes*, Wiley Blackwell, 2006, pp. 90–113.
- Kachru, Braj B. "Models for Non-Native Englishes." *The Other Tongue: English across Cultures*, University of Illinois Press, Urbana, 1992, pp. 48–74.
- McArthur, Tom. *The English Languages*. Cambridge University Press, 1998.
- Melchers, Gunnel, et al. *World Englishes*. Routledge, 2019.
- Smith, L., editor. *English for Cross-Cultural Communication*. Macmillan, 1981.