

Noun Phrase Complexity in Pakistani English: Focus On Syntactic Function and NP Pattern

Muhammad Kamran Abbas Ismail

PhD. Scholar, NUML, Islamabad

kamran.ch@uo.edu.pk

Dr. Azhar Habib

Lectuer, NUML, Islamabad

ahabib@numl.edu.pk

Abstract

Taking a more corpus-based method to examine NPs in newspaper editorials, the current research examined the complexity of the NP in Pakistani English, referred to five parameters proposed by Brato (2020): NP patterns, length of premodification, length of postmodification, complexity of premodification, and complexity of postmodification). The mixed-method research strategy that combines quantitative frequency counts that are corpus-based with qualitative comparative analysis was adopted. The study considered four possible predictors (syntactic function, length, structural complexity of the premodifier, and structural complexity of the postmodifier) using conditional inference trees. Moreover, within the variationist theoretical framework, the integrated model by Brato (2020) was used as an analysis tool in this study the frequency counts of the variants in the corpus were conducted. The analysis of the data showed that syntactic function is a significant predictor of NP complexity in the Pakistani English and the patterns of NPs also have a correlation with syntactic function. The results also show that Pakistani English has structural simplification and nativisation.

Keywords: Pakistani English, Corpus-based, Noun Phrase Complexity, Nativisation

Structural simplification, Editorial

1. Introduction

The new forms of the English language, which are being adopted in multilingual environments, including Pakistan, India, Singapore, and Nigeria are called New Englishes. Platt et al. (1984) came up with the term New Englishes to refer to those varieties of the English language that were the official languages of the former colonies of Great Britain (Lange & Leuckert, 2020). Kachru (1985) classified the varieties of English into three circles which included the inner circle, the outer circle and the expanding circle. The inner circle includes the formed and norm-giving forms of English which are in place in monolingual contexts, including the British and the American English. By contrast, the types of the English language that have developed in multilingual environments, including India, Pakistan, and Singapore, belong to the outer circle varieties of English (Lange & Leuckert, 2020). The outer circle varieties of English are referred to as the Postcolonial English (PCE) by Schneider (2007). But today, scholars specifically call all the outer circle Englishes, in use in multilingual countries, using the term, New Englishes. The American and New Zealand English though they are postcolonial variants of English do not fall in the category of New Englishes. It is only English that is championed in these monolingual nations (Lange & Leuckert, 2020).

The variety of the outer circles, especially the New Englishes is contrasted with the outer circle variety, which has a form and functional difference with the inner circle varieties, including the British and American English (Schneider, 2007). New Englishes that had been viewed through the lenses of deviance to the standards (Native Englishes) are currently viewed through the lenses of differences to these norms. These variations and differences at different levels including the lexical and syntactic levels have been distinctive potential identifiers of new varieties of English. These new varieties, i.e., Pakistani English is a type of New English that has many characteristics, including transfer, interference, and simplification, in contrast to standard varieties of English, like British English and American English (Akinlotan, 2017). These features have become an increasingly popular study of the structure of New Englishes. These are the major themes that a number of committed works (Brato, 2020; Akinlotan and Housen, 2017; Akinlotan, 2016; Schneider, 2007; Gorlach, 1999) have investigated and discovered that New Englishes is undergoing structural nativization.

New Englishes Structural complexity in New Englishes has been thought of as a significant variable in conceptualising and recognising new forms of varieties that emerge. Nevertheless, the forecast of such variability in complexity has received scant attention in the academic literature so far with most of the studies concentrating on the individuality of grammatical characteristics (e.g., adverb, articles, yes/no questions etc.) in Pakistani English newspaper corpora (Buriro, Zaki & Asharaf, 2023), corpus-based study of grammatical phrasal novelty of Pakistani novels (Izhar, Shahzadi & Bashir et al., 2024).

Therefore, the structural simplicity/complexity has emerged as an important variable according to which the New Englishes and inner-circle varieties may be traced (Akinlotan & Housen, 2017). New Englishes are supposed to show the structural simplification on the various levels i.e. morphology, lexis and syntax (Gorlach, 1998). Since nouns are taken to be the bricks or building blocks of a language structure, nouns are a crucial aspect of a language as it has been argued by Algeo (1995, p. 203) that semantic content of sentences is principally carried by the nouns. A noun is a headword in a noun phrase and the noun phrase involves a head element with or without one or more dependents, modifiers, in the form of postmodifiers and premodifiers (Huddleston and Pullum et al. 2002:329).

It is worth noting that the outer circle varieties of the simplification of structures can manifest themselves in the noun phrase structure i.e., New Englishes. As mentioned by Schilk and Schaub (2016, p.58), descriptions of varieties of English often identify the term noun phrase (NP) as the source of structural variation. The assumption has been that NP is more or less complex syntactically, and the conditions that cause NP to be more or less complex in New Englishes are largely unclear. Since different opinions exist regarding the complexity of NP in New Englishes (i.e., length of NP, the number of phrases an NP comprises, or sentential NPs), NP simplicity/complexity in new Englishes deserves more focus in the research.

The research design in the current study is corpus-based where data will be analyzed using a self-compiled corpora of Pakistani Written English (PWE), on the basis of newspaper editorials in English. Through the consideration of the formulae of self-developed corpus, this paper explores the factors predicting noun phrases structure, including syntactic function (grammatical functional), the length of a noun phrase, the complexity of premodification, and the complexity of postmodification as a measure of the quality of a noun phrase to be

sentential. In this manner, the research paper has taken a corpus based variationist approach to examine structure complexity of noun phrase in Pakistani English.

Therefore, the given work is focused on the structural description of the NPs in Pakistani English. Based on the variationist paradigm of Berlage (2014) and the analytical position of Brato (2020), the given work examines the structural differences in NPs in Pakistani English.

2. Research Methodology

2.1 Corpus

The current study focused on the untagged electronic corpus of editorials from Pakistani English newspapers. The corpus from the editorials of the selected newspapers was compiled using online text collector software. The compiled files were parts-of-speech (POS) tagged in Treetagger. Only potential NP heads (token tagged as NN1 and NN2) were extracted randomly through AntConc version 4.2.0 from a large text corpus comprising the editorials of the targeted newspapers, using regular expressions in AntConc (Anthony, 2022). The Regular expressions were built and used in this study. NPs with pronoun heads were excluded from the data as they carry no modifiers (Biber et al., 1999). Usually, they are not premodified. Therefore, they were excluded from the analysis.

2.2 Research Design

This exploratory study provided insights into noun phrase complexity in Pakistani English by examining the elements that make an NP complex in this variety. It proposed a corpus-based research design to explore the structural complexity of NPs based on Brato's (2020) five parameters: pattern of NP, the length of premodification in terms of the number of words, the length of postmodification in terms of the number of words, the complexity of the structure of premodification, and the complexity of the structure of postmodification. The corpus for the present study comprises newspaper editorials, and the size of the corpus was five million words. This corpus-based study employed Brato's (2020) analytical framework to conduct corpus-based frequency counts of the variants. This study also avoids theoretical commitment to any specific syntactic theories, is exploratory in design, and is underpinned by an interpretivist epistemological research paradigm (Lincoln & Guba, 2003). Using conditional inference trees, the study took into account only two potential predictors: NP Patterns and syntactic function.

3. Results

The following are the results of the study for each parameter. The results for ctree and random forest are also discussed.

3.1 NP Pattern of Pakistani English

The current study followed De Haan's (1993) proposed classification for the distribution of NP patterns separated by corpus and syntactic function. De Haan's (1993) model classifies noun phrases into four categories: unmodified, premodified, postmodified, and pre-and postmodified NPs. The study found that unmodified NPs occur most predominantly in Pakistani English, followed by pre- and postmodified NPs. Moreover, figure 1 shows that the overall realisations of unmodified NPs account for 50 per cent in The Dawn, which ranks highest, followed by 46 per cent in The News and 45 per cent in The Express Tribune. The Nation ranks lowest in terms of the realisation of unmodified NPs, accounting for almost 15 per cent of the total. The distribution of NP patterns separated by corpus and syntactic function in Pakistani English newspapers is shown in figures 1 and 2 below.

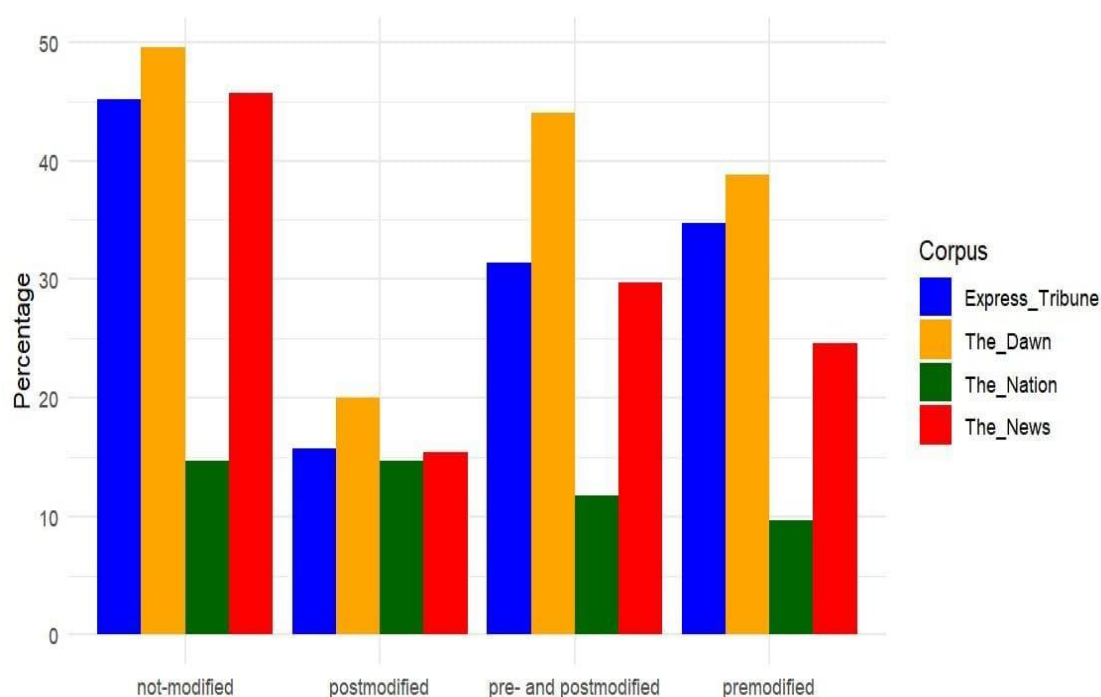


Figure 1 Descriptive Statistics for NP Patterns (De Haan, 1993) by Corpus (Newspapers)

Premodified NPs, in which the head noun is preceded by modifiers such as adjectives, were more frequently found in The Dawn than in the other newspapers. Figure 1 shows that the English daily 'DAWN' ranks the highest with almost 39 per cent premodified NPs, followed by The Express Tribune and The News with 35 per cent and 25 per cent

premodified NPs, respectively. The Nation recorded the lowest percentage of premodified NP patterns, which is 10 percent.

Postmodified NPs are where the head noun is followed by modifiers such as prepositional phrases, relative clauses, non-finite clauses, or finite clauses. The above figure demonstrates that 20 per cent of postmodified NPs were realised in The Dawn, followed by The Express Tribune with 16 per cent, and with a slight difference, only 15 per cent of postmodified NPs were found in The News. The Nation again recorded the lowest percentage of premodified NP patterns. Moreover, postmodified NPs are the lowest NP pattern in this corpus.

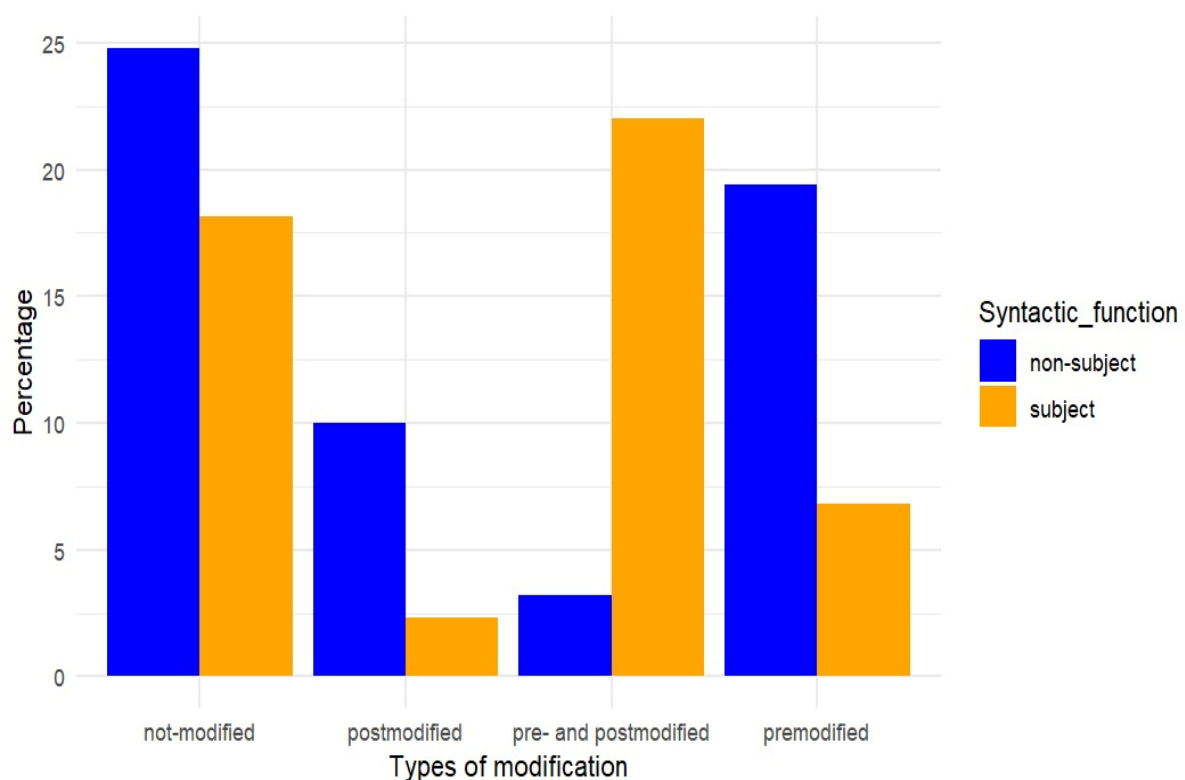


Figure 2 Descriptive Statistics for NP Patterns (De Haan, 1993) by Syntactic Function

Regarding syntactic function, the data exhibits a binary split between subject and non-subject. The present study found that the non-subject positions are most prevalent in all types of NP patterns, i.e., unmodified, premodified, and postmodified. However, they greatly differ in their syntactic functions. Further, the comparison of non-subject and subject positions reveals that unmodified, premodified, and postmodified occur more frequently in non-subject

positions than in subject positions. Figure 2 shows that almost 25 per cent of unmodified NPs occur in the non-subject or object position, whereas almost 18 per cent of unmodified NPs are found in the subject position. The premodified NPs are the second NP pattern more likely realised at the non-subject position than the subject position. Almost 19 per cent of premodified NPs are found at the non-subject or object position, while only 7 per cent of premodified NPs occur at the subject position.

As illustrated in the example below:

“His strong position on governance has sparked debate.” (The Express Tribune)

Postmodified NPs have also been found more frequently in non-subject positions than in subject positions. Figure 2 also shows that almost 10 per cent of postmodified NPs were found in the non-subject or object position, while 2 per cent of postmodified NPs were found in the subject position. In contrast to the other three types of NP patterns, the pre- and postmodified NPs are more likely found at the subject position than the non-subject position. Figure 2 reveals that almost 23 per cent of postmodified NPs occur at subject positions, whereas only 3 per cent of pre-and postmodified NPs were found at non-subject positions. As the example below shows:

“If Fata is ever to be normalised and not simply returned to the pre-insurgency state of affairs, it has to have a sustainable responsive system of government running **the gamut of civil administration to the administration of justice to law and order**”. (The Dawn)

Thus, the results of the present study show that NPs at non-subject positions occur more frequently than those at subject positions. Moreover, syntactic function is an important predictor of NP complexity in Pakistani English.

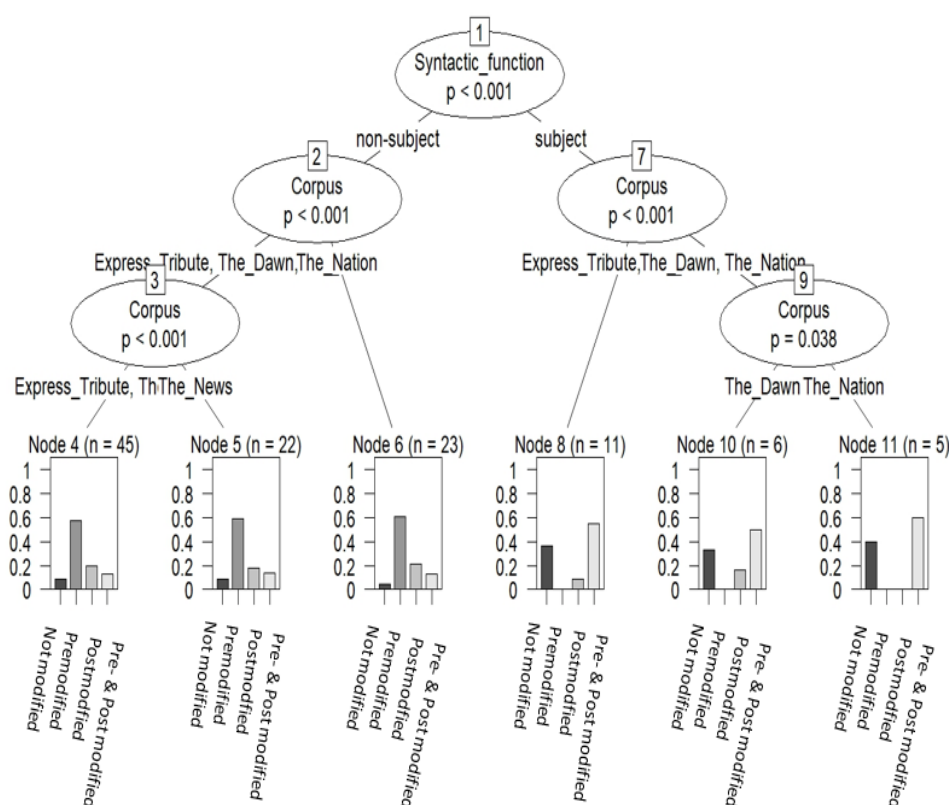


Figure 3 Conditional inference tree modeling the dependent variable NP pattern

In Figure 3 above, the ctree shows that the syntactic function is a significant factor in the distribution of patterns. Syntactic function separates the NP pattern (corpus) into two splits: non-subject (object) on the left hand and subject on the right hand. The random forest results describe a visible correlation between corpus (0.038, 0.001) and syntactic function (0.001).

The descriptive data findings of NP patterns of NP complexity in Pakistani English newspaper corpora show that the unmodified NPs are the most predominant, then there are pre- and postmodified NP patterns. Furthermore, this corpus has the lowest usage of postmodification NPs.

In addition, comparison of non-subject and subject positions demonstrate that non-subject positions emerge quite often in premodified and postmodified NPs whereas unmodified NPs and pre- and postmodified NPs were also observed in the subject positions. Also, syntactic

role and premodification, as opposed to postmodification, are stronger predictors in Pakistan English (the present study). It is indicated in figure 2 that the unmodified NPs are more common in the non-subject position than in the subject position. Furthermore, the postmodification is least frequent and also it is often used in the non-subject position. In the current study, syntactic function and premodification predictors are stronger predictors in comparison to postmodification in Pakistan English.

The research study also revealed that the most common three types of NP pattern that are unmodified, pre modified and post modified are the most common in the non-subject positions. The current research arrives at a conclusion that syntax is an important variable in the distribution of patterns of noun phrases in New Englishes and there is a high level of relationship between the complexity of the noun phrases and the syntactic functions in Pakistani language. The results of the current study are also consistent with that of other studies (Brato, 2020; Akinlotan and Housen, 2017; Schilk and Schaub, 2016).

4. Conclusion

The study of the complexity of the noun phrase in the Pakistani English newspaper discourse is an under-researched field and the current research provided valuable information on the topic. Through a combination of major complexity indicators to explore the connection between salient aspects and syntactic processes, the study attempted to acquire a fine perspective of NP complexity in Pakistani English. The current research arrives at a conclusion that syntax is an important variable in the distribution of patterns of noun phrases in New Englishes and there is a high level of relationship between the complexity of the noun phrases and the syntactic functions in Pakistani language.

This corpus-based work found that Pakistani English uses the not-modified structure very often, and the postmodifications come next, which is characterized by a preference to use postmodifications less often. The high-position of the non-subject (object) positions characterizes the NP complexity of the selected corpora, which is a major difference between the other varieties. The current research found that to some extent, the structural simplification hypothesis of noun phrases in terms of Pakistani English is supported by the structure of noun phrases in Pakistani English, based on quantitative and qualitative analyses to establish the level of complexity change in the structures of noun phrases in Pakistani English (Görlach, 1998).

References

- Akinlotan, M. (2016). Genitive alternation in New Englishes: The case of Nigerian English. *Token: A Journal of English Linguistics*, 5, 59–73.
- Akinlotan, M., & Housen, A. (2017). Noun phrase complexity in Nigerian English: Syntactic function and length outweigh genre in predicting noun phrase complexity. *English Today*, 33(3), 31–38. <https://doi.org/10.1017/S0266078416000626>
- Algeo, J. (1995). *The Cambridge history of the English language* (Vol. 4). Cambridge University Press.
- Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Waseda University, Tokyo, Japan. Available from <https://www.laurenceanthony.net/software>.
- Berlage, E. (2014). *Noun phrase complexity in English*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., Reppen, R., & Leech, G. (1999). Corpus linguistics: Investigating language structure and use. *International Journal of Corpus Linguistics*, 4(1), 185–188.
- Brato, T. (2020). *The noun phrase across varieties of English: A cross-varietal and diachronic perspective on NP complexity*. Heidelberg University Publishing.
- Buriro, W. M., Zaki, S., & Ashraf, A. (2023). Grammatical features of Pakistani English: A corpus analysis of Pakistani English newspapers and magazines. *Linguistic Forum-A Journal of Linguistics*, 5(3), 89–113.
- Clark, V. L. P., & Creswell, J. W. (2008). *The mixed methods reader*. Sage.
- Crosthwaite, P. (2023). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge.
- De Haan, P. (1993). Sentence length in running text. In C. Souter and E. Atwell (Eds.), *Corpus-Based Computational Linguistics* (pp. 147–161). Brill.
- Görlach, M. (1998). Varieties of English world-wide: Where we stand. *Links & Letters*, 5, 13–36.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*.

Cambridge University Press.

Izhar, M., Shahzadi, L., Ali, S., Bashir, T., & Tabasum, R. (2024). An analysis of novel expressions in Pakistani English using corpus-based research. *Remittances Review*, 9(2), 4195–4215.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk and H. Widdowson (Eds.), *English in the World: Teaching and learning the Language and literatures* (pp. 11–30).

Cambridge:Cambridge University Press.

Lange, C., & Leuckert, S. (2020). *Corpus linguistics for World Englishes: A guide for research*. Routledge.

Lincoln, Y. S., & Guba, E. G. (2003). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (pp. 253–291). Sage.

Platt, J. T., Weber, H., & Ho, M. L. (1984). *The New Englishes*. London: Routledge & Kegan Paul.

Schilk, M., & Schaub, S. (2016). Noun phrase complexity across varieties of English: Focus on syntactic function and text type. *English World-Wide*, 37(1), 58–85.

Schneider, E. W. (2007). *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.

Walliman, N. (2022). *Social research methods*. Sage.