

Language Contact and Syntactic Complexity: A Corpus-Based Analysis of Clause Structure in Pakistani English Research Articles

Mairam Arif^{1*} Behzad Anwar² Uswa Shahid³ Quratulain Cheema⁴

¹ PhD Scholar, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan. Email: 25016102-009@uog.edu.pk

² Associate Professor, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan. Email: behzad.anwar@uog.edu.pk

³ PhD Scholar, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan. Email: uswaawan27@gmail.com

⁴ Assistant Professor, Government Graduate College (W) Satellite Town, Gujranwala, Punjab, Pakistan.

Email: quratulaincheema@gmail.com

ABSTRACT: The study investigates the syntactic complexity in Pakistani research articles. The focus of the research is on academic texts produced by Pakistani scholars and published in international journals. A mixed-methods research design is employed to examine the factors contributing to syntactic complexity in Pakistani English academic writing. This research mainly focuses on the complementation clause (that-clause) and other embedded clauses, which contribute to the syntactic complexity in writing. The syntactic measurement is examined using the key indices of Lu (2010), which present a statistical analysis of clausal structure. In this perspective, Weinrich's (2010) syntactic theory is used to explain the qualitative findings. The sample for the study comprises internationally published research articles with Pakistani 1st author publications. The articles are converted into a specialized corpus. There are 96 files or research articles included in the study. Examples of clauses are extracted using AntConc (3.5.9). The findings reveal that dominant complementation clauses are prevalent in Pakistani academic writing. The syntactic structure is often complex, which highlights the embedding of local languages. In academic discourse, language is strongly influenced by a multilingual perspective (Weinrich, 2010). Pakistani English is a highly contact language, and studying it is very significant from an academic writing perspective.

KEYWORDS: Syntactic Complexity, Language Contact, That-clause, Pakistani English, Clausal Complexity

Introduction

The English language is spread globally, and the pluricentric character of English has led to a wide variety of forms. These varieties are treated as a research topic from a World Englishes perspective (Bernardo, 2023). In the context of ELT, World Englishes have steadily gained momentum. English is a global language and holds the prestige that no other language has attained yet. It is the result of the development process that has been ongoing for the past few years (Imran & Wahid, 2020). The English language has undergone a significant transformation since its origin. The historical development of the English language can be traced back to the migration of Germanic tribes, including the Jutes. So, English evolved from the Old English and Middle English periods to the modern era around the 16th century. The major reason for the spread of English is the impact of British Colonial rule in the subcontinent. This

Pages: 29 – 41

Volume: 5

Issue: 7 (July 2026)

Corresponding Author

Mairam Arif

✉ 25016102-009@uog.edu.pk

Cite this Article: Arif, M., Anwar, B., Shahid, U., & Cheema, Q. (2026). Language Contact and Syntactic Complexity: A Corpus-Based Analysis of Clause Structure in Pakistani English Research Articles. *The Regional Tribune*, 5(7), 29-41.

<https://doi.org/10.55737/trt/v-vii.332>

language spread to various regions, including Pakistan. So, the spread of the English language has contributed to the development of various varieties of English. There are various regional and continental variations in the English language. The idea of linguistic variation has necessitated the analysis of linguistic patterns across various registers. Registers include the situationally defined variety of the English language, in which various features are included, and these are inclined to various regions (Imran & Wahid, 2020). The spread of the English language has given rise to various varieties, which are also worth studying. Each one is specified in terms of cultural impacts.

Syntactic Complexity

Syntactic complexity is considered a key component in research in corpus linguistics (Biber et al., 2011). In academic writing, syntactic complexity is widely acknowledged as a key measure of quality and an important predictor of literary quality (Ortega, 2003). In the Pakistani context, the development of syntactic complexity is influenced by extensive L1 interaction (Butt & Anwar, 2019; Anwar et al., 2020). This resulted in changes to writing conventions and other related factors (Anwar et al., 2020), particularly in the Pakistani context of higher education. English is considered the primary medium of instruction because it is an institutionalized, serving variety (Mansoor, 2005). It is also critical for students to navigate such linguistic requirements to achieve academic success and develop as global citizens (Hyland, 2004).

According to Wolfe-Quintero et al. (1998), syntactic complexity refers to the degree of diversity and sophistication of the sentences in language output, which reflects the writer's capacity to employ various grammatical resources to transmit complicated sentences and empirical facts. It is increasingly viewed through the prism of development routes, in which students progress from basic coordination to sophisticated subordination and ultimately to an extremely compact phrasal structure that characterizes advanced academic language (Taguchi et al., 2013). For students, addressing these linguistic challenges is necessary to achieve academic success and subsequently join the global community (Hyland, 2004). The relationship between language and cognition is very important. As learners grow older, their thinking becomes more complex, requiring language forms that convey their messages effectively (Vygotsky, 1986). In academic writing, syntactic complexity is a multifaceted concept, usually categorized into coarse-grained and fine-grained indices (Norris & Ortega, 2009).

In a second language, the skill of writing is a very complex phenomenon. It requires complete grammatical knowledge. ESL learners, especially those involved in academic writing, are often required to acquire more academic knowledge. It requires precision, clarity, and correct syntactic structure. Structural accuracy is the key component of academic writing. Moreover, academic writing is a crucial ability of university students. It serves as a primary means of assessment and information. To create meaning in words, syntax is very important, as it ensures effective and meaningful written communication. Coherence and clear grammatical patterns are facilitated by proper grammar, making the sentences fluent.

English as Lingua Franca (ELF)

Since the prestigious publication of scholarly works, English has been widely used as a performance measure in ELF writing (Hyland, 2016a). Globally, there is a rapid increase in the publication of scholarly writings. It is estimated that there are 6 million academic writings from 17,000 institutions and research organizations worldwide (Hyland, 2016b; Bjork et al., 2009). There is pressure on academics to publish articles, and under this pressure, they submit their work to highly prestigious international journals, most of which are in English (Cargill & O'Connor, 2006). In this context, various writers with diverse backgrounds and first languages are involved in academic publication. In short, the academic literature published in prestigious international journals does not have a single background; rather, it is intended for ELF learners. Thus, investigating the use of ELF journals in scholarly writing may provide

insights into unique issues that make for interesting debate in ELF writing (Wu et al., 2020). The results of the statistics showed that there are 90 percent journals worldwide published in English, even the international scientific indexing (ISI) database is published in the English language. English is widely used in academic publishing, which has clear benefits. Firstly, English is a common language, and it makes international collaboration in publishing easier. Secondly, it enables international writers to promote global exchange that belongs to a distinct linguistic background (Wu et al., 2020). However, research conducted in English has largely focused on the English language as a native language or on comparing it with other English varieties (Swales, 1990; Granger & Tyson, 1996; Martínez, 2005). However, the differences between these suggest that speakers of one native language have issues with the proficiency of the other native language and are instructed to emulate ENL proficiency in academic writing. So, ELF research should focus on the characteristics of ELF academic discourse (Ingvarsdóttir & Arnbjørnsdóttir, 2013). Moreover, the ELF perspective can reduce linguistic injustices that have prompted a growing number of studies (Flowerdew, 2008; Hyland, 2016b). So, it is very important to explain the ELF contribution to the international academic writing context.

The use of the English language in ELF writing has been the subject of various studies. For example, Lorés-Sanz (2016) provided evidence supporting her theory that non-anglophone authors had changed the rhetorical quality of their research paper abstracts. The sample of her study was abstracts from 66 research papers. Applying Swales' (1990) technique, she examined the rhetorical structure of ELF writings between 2012 and 2014. She proposed three hybrid structures for variation, in addition to the traditional writing style, in the abstract section. A detailed description was given of a thorough move to a straightforward structure intertwined with the method, aim, and conclusion. Another related example is the study by Mur-Dueñas (2015), which analysed the evaluative *it*-clause to investigate the lexical and grammatical features of ELF writing. The sample of this study consisted of 48 research articles published in international journals in the field of business and management from 2002 to 2010. The results showed that a greater variety of evaluative adjectives was used in ELF international journals. All these studies are crucial for highlighting the creative use of English in academic writing. In other words, these writings demonstrate the use of the English language at an international level in scholarly works.

Pakistani English

Pakistani English holds a unique position within world Englishes, functioning as an unofficial variety shaped by the socio-cultural, religious, and linguistic landscapes of South Asia (Baumgardner, 1993). There are three distinct historical phases: first, the missionary period, from 1614 to 1765; then, the local demand phase, from 1765 to 1835; and finally, the era of government policy (Kachru, 1986). English is a colonial legacy in Pakistan, and it is considered the language of institutions and the court (Ahmad et al., 2020).

Localized varieties of English have emerged alongside the spread of English, and these are frequently studied under the umbrella of world Englishes. Pakistani English has evolved into a unique language with its own characteristics, influenced by multilingual, educational, and social settings. Scholars have drawn considerable attention to lexical variation. Little is known about the syntactic variation and the nativization process. Moreover, there is also limited corpus-based research on syntactic complexity. The study's research gap concerns the syntactic complexity of tense and aspect in international Pakistani research articles and its effects on the credibility of the academic writing. This study aims to answer the following research questions:

1. What are the different patterns of syntactic complexity used by Pakistani scholars in academic writing?
2. Which clausal aspect is most frequently used in Pakistani research article writing?

3. How does the that-clause contribute to syntactic complexity and facilitate the embedding of various clauses in Pakistani academic discourse?

Literature Review

Language complexity is not an easy concept and is also difficult to handle because it can be approached in various ways (Ortega, 2003; Bulté & Housen, 2012; Pallotti, 2014). Syntactic complexity is crucial in SLA (Shahid et al., 2026) because it is a key measure in SLA research and L2 proficiency (Crossley & McNamara, 2014), and it also determines linguistic ability (Ortega, 2003) to inform specific educational interventions (Ellis, 2009; Ong & Zhang, 2010).

Language variation is not a new phenomenon. The origins of this research can be traced back to Trudgill's (1999) study of various English dialects. He highlighted that, compared with lexical variations, syntactic variations receive less attention in English. He also found that the English language is undergoing continuous change, a consequence of various factors. Various studies highlight the importance of syntax in a language. Syntactic complexity is considered a very important aspect of second language acquisition and is widely used as a reliable measure of language proficiency (Shahid et al., 2026). Moreover, it can be used to measure the index and the developmental status of L2 (Arif et al., 2026). Various studies have investigated syntactic complexity and demonstrated its reliability as a measure of L2 proficiency. For instance, in a study, Martínez (2018) showed a strong correlation between writing quality and syntactic complexity, based on overall essay grades from secondary-level students. In particular, the researcher investigated the use of simple sentences in relation to low-quality writing. The main purpose of the study was to highlight syntactic complexity based on writers' performance in the essays. The author found that using lengthy clausal structures and sentence units is a reliable means of producing complex writing.

Moreover, very few studies have focused on the syntactic complexity of EFL writers. Some studies focused on comparing syntactic complexity indices that measure different constructs. Huang et al. (2024) demonstrated that EFL learners used more complex and lengthy sentences, verb phrases, and T-units. There is less coordination in speaking than in writing. However, Yoon and Park (2021) demonstrated that syntactic complexity did not differ significantly between Korean EFL learners' writing and their speaking, except for complex nominal phrases per clause.

Methodology

To investigate the impact of language contact and syntactic complexity in Pakistani research articles, the study adopts a mixed-methods design. In the vast dataset, the linguistic design is only authentic and possible in the corpus method. The corpus method offers a systematic and scientific approach to analyzing large datasets and evaluating naturally occurring data. The focus of the study is on measuring clausal complexity. Moreover, to uncover the syntactic complexity of Pakistani scholars, the research uses a comparative design. Quantitative findings are revealed by using the corpus software AntConc (3.5.9). Then, examples are extracted from the corpus data and further measured using the syntactic complexity scale presented by Lu (2010). This research explores nativized syntactic patterns adopted by Pakistani English research articles. These patterns and syntactic complexity reflect adaptive and functional language use within a particular sociocultural context rather than as deviations from standards.

Theoretical Framework

The theoretical framing of the research is based on the syntactic measure of Lu (2010). Lu (2010) presented a syntactic measure to investigate syntactic complexity, which serves as the theoretical framework for this study.

Table 1

Theoretical Framework for Measuring Syntactic Complexity by Lu (2010)

Length of production unit	Mean length of clause (MLC)	Word\clauses
	Mean length of sentences (MLS)	Word\sentences
	Mean length of t-Unit (MLT)	Word\T-units
Sentence complexity	Clauses per sentence (CS)	Clauses\Sentences
Sub ordination	Clause per t-unit (C\T)	Clause T-unit
	Complex T-unit per t-unit (CT\T)	Complex t-unit
	Dependent clause per clause (DC\C)	Dependent clause per clause
	Dependent clause per t-unit (DC\T)	Dependent clause per t-unit
Coordination	Coordinate phrase per clause (CP\C)	Coordinate phrase per clause
	Coordinate phrase per t-unit (CP\T)	Coordinate phrase per t-unit
	T-units per sentence (T\S)	T-units per sentences
Particular structure	Complex nominals per clause (CN\C)	Complex clause\clause
	Complex nominals per t-unit (CN\T)	Complex nominals\t-unit
	Verb phrase per t-unit (VP\T)	VP t-unit

The table presents indices of syntactic measurement to investigate syntactic complexity in Pakistani academic writing. By using the framework of Lu (2010), the clausal structure is explored. Further, the implications and discussion of the results are explained using Weinrich's (2010) language contact theory. This theory highlights the effect of local languages on contact with other languages. So, Pakistani English is discussed in the light of language contact theory.

Sample

The research sample is taken from a Pakistani international research article published in the field of English. There is a total of 96 articles published between 202-2026. The first author is Pakistani, and the articles are taken from international publishers, including Elsevier, ScienceDirect, Sage, Francis, Taylor, and Wiley.

Table 2

Details of International Journal Articles Published by Pakistani Authors

Files	96
Word tokens	825424
Word Types	30375

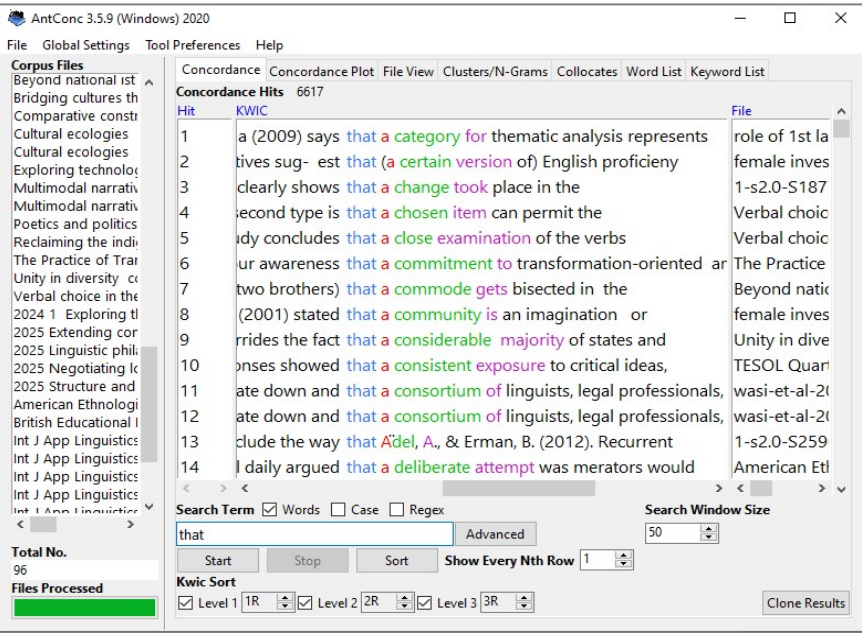
A specialized corpus of international journals written by Pakistani authors is compiled for the purpose of the study. International journals contain articles from various countries, but only the Pakistani 1st-author articles are selected to compile a specialized corpus. There is a specific inclusion criterion. The articles are related to the field of English; it is not specified whether they are from linguistics or literature. But the compulsion is that the authors are Pakistani. The research articles are published from 2022 to 2026. The corpus comprises 96 articles containing 825424 tokens, which are sufficient to provide reliable statistical details. Only open-access articles are included in the criteria. The articles are downloaded as PDF files. Later, these are converted into a plain text file for processing in the AntConc (3.5.9) software. Then the data is prepared for processing in the software, and it also yields reliable statistical results.

Data Analysis

This part of the study provides a detailed data analysis of Pakistani research articles published in international journals. The research aims to determine the syntactic complexity of that-clauses in the corpus of Pakistani articles. The corpus is first analysed using AntConc (3.5.9), as shown in the given figure:

Figure 1

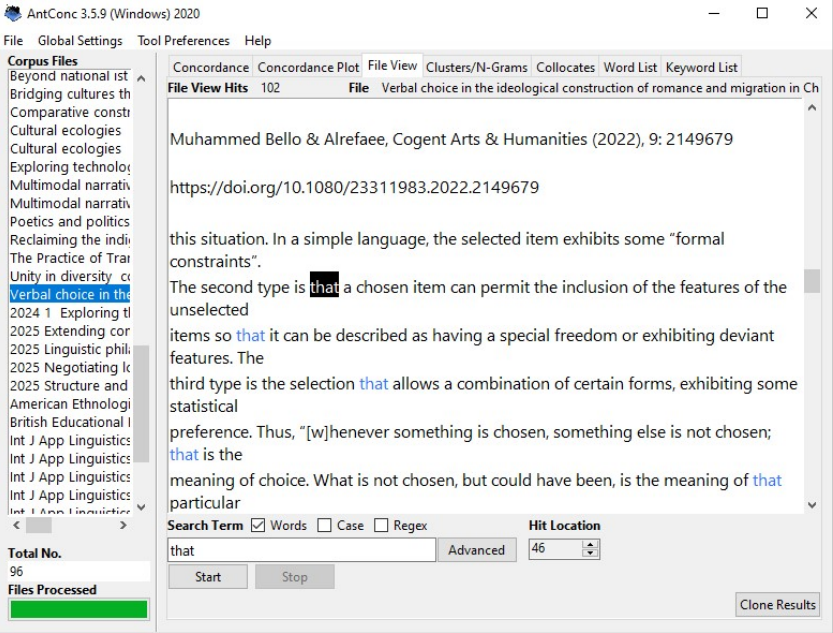
Concordance of That-clause Extracted by AntConc (3.5.9)



First, the corpus is processed in AntConc (3.5.9), which provides quantitative information on the concordance of that-clauses in research papers. It describes the concordance of the given command, which represents the concordance hit of that-clauses in the corpus. Furthermore, when clicking a specific concordance line, the software opens the plot view, which displays KWIC (keyword-in-context). Figure 2 illustrates the way of extracting examples:

Figure 2

File View of That-clauses in the Corpus



This is the software interface from which the data is extracted, and examples are not taken from the relevant corpus of Pakistani research articles. The examples are further quantified using the syntactic complexity measure. There is a one-by-one discussion of the extracts that are taken as examples from the corpus.

Extract 1: *Saldaña (2009) says that a category for thematic analysis represents codes (words or phrases) which symbolically assign summative, salient, essence-capturing, or evocative attributes of language-based data.*

It is analyzed by using the syntactic measure of Lu (2010). It is classified as having only one T-unit in the first example. Then, the syntactic complexity is further explored by using the clause division pattern. The first example consists of the main clause (Saldana says). The main clause was followed by the complement clause (that a category...), then by the relevant clause.

Extract 2: *To analyze and extract themes from the qualitative data. The study found that female learners heavily.*

There are 2 sentences in this extract. It starts with a non-finite purposive clause, and it has only one t-unit. So, there is only one clause considered.

Extract 3: *The data clearly shows that a change took place in the learner's attitudes to responding to class fellows and teachers in greeting, meeting, and helping in teaching/learning. Norton (2010) states that identity is associated with social relationships, and when people change their context, their identities also change. Thus, when participants moved from the high school context to the university context, they changed their attitudes to the relationships in the new context, which, according to Norton's (2010) claim, is identity change.*

The structure of this extract is completely different from the previous one. There are multiple t-units in this example of the corpus. In the example, 3 t-units are found.

Extract 4: *The second type is that a chosen item can permit the inclusion of the features of the unselected items so that it can be described as having a special freedom or exhibiting deviant features. The third type is the selection that allows a combination of certain forms, exhibiting some statistical preference. Thus, "[w]henever something is chosen, something else is not chosen; that is the meaning of choice. What is not chosen, but could have been, is the meaning of that particular instance of choice. When we ask the question 'why did the speaker say that?', what we want to know is why that, why not this, or the other. What difference would it have made if the speaker had chosen something else? We can consider each choice independently, pinpointing it in terms of one system at a time".*

Extract 4 shows the clausal breakdown. There is one main clause in this example. Apart from the main clause, there are other embedded clauses, such as adverbial, complement, and relative clauses, that are used in the sentence structure.

Extract 5: *Specifically, the verbs of romance, which are either finite or non-finite (but usually lexically extensive), reveal the phenomenon at physical and psychological levels, but there are also some verbs that have emotive value, depicting a character's assessment of another character's perception or judgment of a phenomenon. The verbs of migration, which are either copular or extensive, are inclined towards depicting direction, linguistic competence of immigrants, their hope, disposition, and challenges in America. The study concludes that a close examination of the verbs used in constructing romance and migration in the novel is a gateway to a further understanding of the text.*

This example shows extremely high syntactic complexity. There is high coordination embedded in this example. The sentence structure is not easy, as it highlights multiple clause types, including complement, relative, and adverbial clauses. All the embeddedness reflects the complex syntactic structure of Pakistani academic discourse. It highlights the strong academic style.

Extract 6: *She is beautiful and has a magnetic personality that attracts men of all cultures, with whom her sexual liberation enables her to interact with pleasure.*

Compared to other clausal structures, example 6 does not exhibit greater complexity in the sentence. There is one relative clause and two coordinating clauses in it.

Extract 7: *To implement these recommendations, it is suggested of all kinds of experts to bring the crime rate down and that a consortium of linguists, legal professionals, and prevail legitimate justice all over the world representatives from relevant fields be formed.*

This example highlights the clausal breakdown as compared to other examples. The example starts with an infinitive clause, “to implement these...”; later, there is a complement clause, “that a consortium...”, and also a coordination clause that begins with “and prevail legitimate...”. There is a total of 3 clauses and 1 t-unit found in this extract. It highlights the formal academic writing style in Pakistani academic discourse. It also reflects the academic register.

Extract 8: *Warning of the limitations that a dialectology of Pashto may face, Elfenbein (1997: 739) identifies the central role that ‘tribal distinctions’ take in its development (and also that of geography influencing dialect boundaries).*

There is no simple clausal structure which is used in extract 8. There are different types of clauses used, including an embedded relative clause, a complement clause, and a finite clause. For example, “and also that....” It embodies the coordination and complement clause.

Extract 9: *The few concessions that the state gives to religious contemporary Muslim society (Butt & Anwar, 2019; Singha, 2022). But minorities, in the form of electoral reservations, jobs quotas- this official invisibilization of caste does not imply that reli-tas, and some legal reforms, are small enough that they gious minorities solidified into settled categories of governance cannot be comfortably shared between the “largest” and in Pakistan (Saeed, 2017).*

Extract 9 also shows the complex syntactic structure in Pakistani academic English. It contains different types of relative, embedded, and complement clauses. For example, “the few concessions..... But minorities.... and that they give.....”.

Extract 10: *Throughout the mid-20th century, demographic language decided that “it was unnecessary to ask Muslims whether they colonial India had been folded into the binaries of religious were Sunnis or Shias.*

The structure of extract 10 differs from that of the other examples. There is an embedding of different clauses within a single sentence. There are embedded clauses (whether they....), adverbial clauses (throughout....), and complement clauses (decided that....) in one sentence. There is no terseness in the syntactic structure.

After explaining the examples according to their clausal structure, there are tabular and graphical representations of the overall syntactic complexity in the research articles. Below is the table representing the indices of syntactic measurement as structured in Pakistani academic writing.

Table 3
Syntactic Complexity in Internationally Published Pakistani Articles (IPPA)

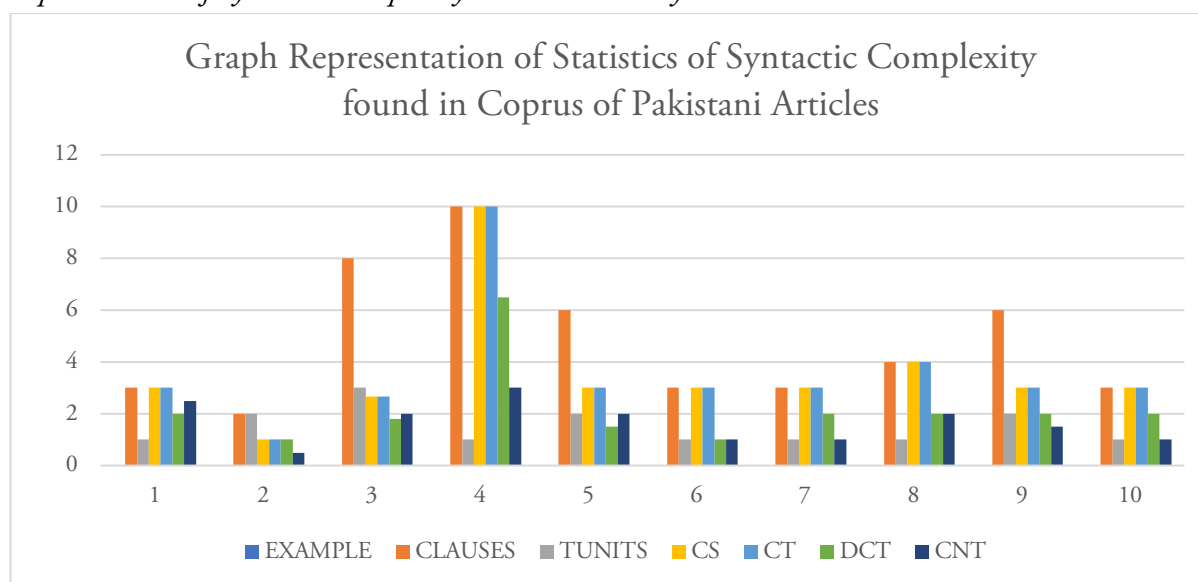
EXAMPLE	CLAUSES	T-UNITS	CS	CT	DCT	CNT
Extract 1	3	1	3	3	2	2.5
Extract 2	2	2	1	1	1	0.5
Extract 3	8	3	2.67	2.67	1.8	2
Extract 4	10	1	10	10	6.5	3

Extract 5	6	2	3	3	1.5	2
Extract 6	3	1	3	3	1	1
Extract 7	3	1	3	3	2	1
Extract 8	4	1	4	4	2	2
Extract 9	6	2	3	3	2	1.5
Extract 10	3	1	3	3	2	1

The above table shows the columns representing the statistical results on syntactic complexity in Internationally Published Pakistani Articles (IPPA). The 1st column shows the number of examples extracted from the corpus using the AntConc software (version 3.5.9). Then there are other columns that separately represent the syntactic complexity following the measurement indices represented by Lu (2010).

Figure 3

Representation of Syntactic Complexity in Internationally Published Pakistani Articles



The above figure also represents the extract-wise division of syntactic complexity, which is found in Internationally Published Pakistani Articles (IPPA).

Discussion

This section of the data analysis provides a qualitative discussion with theoretical foregrounding of Weinrich's language contact theory. The findings lead to various discussion points. The statistical results show that in the corpus extracts, Pakistani authors exhibit a high degree of clause embedding. The results of the corpus also show that the *that-clause* is dominant in the whole corpus. There are more sentences that use *that-clauses*, highlighting the embedding of the clause structure. There is also frequent use of the relevant cause throughout the sentence structure. From the perspective of academic discourse, strong nominalization is found in the corpus. For example, the words "limitations of", "analysis of", and "inclusion of" are used to illustrate nominalization from an academic perspective. So, if these are assessed from a complexity perspective, some sentences are highly complex, and others are not. There is no regular pattern of clause structure in the academic writings of Pakistani articles.

The interpretations of the results are discussed in light of Weinrich's language contact theory. There are a number of complexities in the clauses that are found. It shows a high degree of complexity. The language contact theory shows that multilingual speakers or writers often transfer features from one language to another, creating a high degree of complexity. The reason may be the language, which creates hurdles to producing strong output on the author's part. So multilingual speakers may introduce syntactic complexity in this way, resulting in structural variance.

One of the most noticeable features of syntactic complexity is the high degree of coordination and long structures in Pakistani research articles. The frequent use of long sentences may cause the use of coordinating conjunctions, which are and, or, but. This coordination is found in various examples discussed above. Extracts 3, 4, 5, and 9 are very particular in showing the embedding clausal structure in the corpus. There are substantial clause structures in the examples, which are highly embedded. From a theoretical perspective, multilingual speakers are often used to embed the structure of one language into another. These embedded and highly coordinated cases result from multilingualism found in Pakistan. In Pakistan, various languages are spoken, such as Urdu, Punjabi, and Pashto, so Pakistani writers do not all come from a single region. Although all the articles are authored by Pakistani 1st authors, we found high language contact based on each author's background. So, in academic writing in Pakistan, the additive clausal structure is more prominent than concise sentence structures. For example, extract 4 highlights a number of embedded clauses. This example contains complement sentences, adverbial clauses, and relative clauses. The results show that there is a high DCVT value (6\7) and a CS value of (10). This lengthy elaboration on syntactic complexity indicates multilingual cognitive methods in academic writing. It highlights the process of information gathering, which is gradual rather than concise and presented in shorter form. This led to the discussion that the long indices result from local transfer, in which local languages' organizational tendencies affect academic writing in English.

There is also frequent variation in the syntactic complexity of patterns. There are various examples that highlight it. For instance, in extract 2, the clausal structure is relatively simple, whereas extract 4 contains a more complex sentence. Weinrich's theoretical perspective, content-induced variation, is closely connected to the variation found in the data. It shows that everyone is affected by language perception differently. Not everyone is affected by language in the same way; rather, they are influenced by various factors. These factors may be cultural or emotional background, which can affect language production abilities. It may be a multilingual production ability, which is a greater reason for inefficiency in academic English writing. From the perspective of Pakistani English, there are various social and regional variations. These variations affect the language production ability from an academic perspective. Not all learners can produce the language in the same way with equal efficiency. In Pakistan, there are also different levels of English language proficiency.

Academic discourse is very important because it allows writers to present different points of view, making the construction highly significant. It provides arguments by citing various data sources. From the standpoint of language production, multilingual authors often rely on the well-known reporting style. They offer an unambiguous and steady framework for arranging data. Therefore, the frequent usage of embedding clauses may highlight the convergence of multilingual communication and English academic norms. The frequent use of these clauses indicates that the authors often draw on multilingual resources when writing in academic discourse.

Moreover, a greater number of clause embeddings leads to syntactic complexity. This is the reason for the hurdles in understanding the clausal structure and its meaning. Weinrich highlights that multilingual speakers often highlight the complex syntactic structure, which is the reason for various background factors.

The findings reveal that Pakistani English is a highly language-contact variety embedded in contact with indigenous languages. The syntactic complexity results from English having close contact with other local languages. There are long sentence structures, embedded clauses, and coordination clauses, which together produce complex sentences. The syntactic complexity indices, developed by Lu (2010), also highlight that Pakistani English is in close contact with other local languages.

Conclusion

The study investigated the syntactic complexity found in internationally published Pakistani research Articles. The study is based on the Lu (2010) framework of syntactic complexity of a few academic writings. The results are discussed using the language contact theory presented by Weinrich. A mixed-methods approach is used in the study to highlight the syntactic structure of embedded that-clauses in Pakistani academic writing. There are various indices that measure clausal complexity in Pakistani academic writing, including clauses per t-unit, coordination clauses, and embedded clauses. Most results show the complexity in the syntactic features of Pakistani academic discourse. Various clause structures are used, such as embedded clauses, nominal structures, and coordinating clauses. The long sentence structure showed that Pakistani English is a highly contact language. It is presented in the context of indigenous languages. Pakistani authors are multilingual. Multilingual authors provide the impact of other local languages. The study is highly significant in presenting the syntactic complexity in Pakistani academic discourse. Future dimensions suggest the comparative analysis of native and non-native varieties of English academic writing.

References

- Ahmad, A., Hussan, S., & Butt, A. (2020). English in the Domain of Legal Services: Problems of Access in Pakistan. *Pakistan Journal of Criminology*, 12(1), 109-127.
- Anwar, B., Ajmal, M., & Keezhatta, M. S. (2020). Nativization of grammatical features in Pakistani English. *Asian EFL Journal*, 27(42), 92-106.
- Anwar, B., Malik, N. A., & Shahbaz, M. (2020). Innovative syntactic features of Pakistan English newspapers. *Kashmir Journal of Language Research*, 23(1). <https://kjlr.pk/kjlr/article/view/36>
- Arif, M., Anwar, B., & Shahid, U. (2026). Functional variation in lexical bundles: A corpus-based comparative analysis of native and non-native web-based English. *Social Sciences Spectrum*, 5(1), 409-429. <https://doi.org/10.71085/sss.05.01.506>
- Baumgardner, R. J. (1993). *The English language in Pakistan*. Oxford University Press.
- Bernardo, A. S. (2023). Philippine English and global englishes-inspired grammar teaching. *Asian Englishes*, 26(1), 174-194. <https://doi.org/10.1080/13488678.2023.2174532>
- BIBER, D., GRAY, B., & POONPON, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35. <https://doi.org/10.5054/tq.2011.244483>
- Bjork, B. C., Roos, A., & Lauri, M. (2009). Scientific journal publishing: yearly volume and open access availability. *Information Research: An International Electronic Journal*, 14(1).
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. *Language Learning & Language Teaching*, 21-46. <https://doi.org/10.1075/llt.32.02bul>
- Butt, B., & Anwar, B. (2019). NDTs in non-native English: A syntactic analysis. *Global Regional Review*, IV(IV), 299-309. [https://doi.org/10.31703/grr.2019\(iv-iv\).33](https://doi.org/10.31703/grr.2019(iv-iv).33)
- Cargill, M., & O'Connor, P. (2006). Developing Chinese scientists' skills for publishing in English: Evaluating collaborating-colleague workshops based on genre analysis. *Journal of English for Academic Purposes*, 5(3), 207-221. <https://doi.org/10.1016/j.jeap.2006.07.002>
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66-79. <https://doi.org/10.1016/j.jslw.2014.09.006>
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474-509. <https://doi.org/10.1093/applin/amp042>
- Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes*, 7(2), 77-86. <https://doi.org/10.1016/j.jeap.2008.03.002>
- GRANGER, S., & TYSON, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17-27. <https://doi.org/10.1111/j.1467-971x.1996.tb00089.x>
- Huang, X., Li, J., Wu, J., Chang, J., Liu, D., & Zhu, K. (2024). Flexibly utilizing syntactic knowledge in aspect-based sentiment analysis. *Information Processing & Management*, 61(3), 103630. <https://doi.org/10.1016/j.ipm.2023.103630>
- Hyland, K. (2004). *Disciplinary discourses, Michigan classics ed.: Social interactions in academic writing*. University of Michigan Press.
- Hyland, K. (2016a). *Academic Publishing: Issues and Challenges in the Construction of Knowledge*. Oxford: Oxford University Press.
- Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, 31, 58-69. <https://doi.org/10.1016/j.jslw.2016.01.005>

- Imran, M., & Wahid, M. R. (2020). The syntactic variations in adjective phrase–double comparatives and double superlatives in Pakistani English (PakE): A corpus-based study. *American Journal of Applied Linguistics*, 8(1), 1-11.
- Ingvarsdóttir, H., & Arnbjörnsdóttir, B. (2013). ELF and academic writing: A perspective from the expanding circle. *Journal of English as a Lingua Franca*, 2(1), 123-145. <https://doi.org/10.1515/jelf-2013-0006>
- Kachru, B. B. (1986). *The alchemy of English: The spread, functions, and models of non-native Englishes*. University of Illinois Press.
- Lorés-Sanz, R. (2016). ELF in the making? Simplification and hybridity in abstract writing. *Journal of English as a Lingua Franca*, 5(1), 53-81. <https://doi.org/10.1515/jelf-2016-0003>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Mansoor, S. (2005). *Language planning in higher education: A case study of Pakistan*. Oxford University Press.
- Martínez, A. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing*, 35, 1-11. <https://doi.org/10.1016/j.asw.2017.11.002>
- Martínez, I. A. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing*, 14(3), 174-190. <https://doi.org/10.1016/j.jslw.2005.06.001>
- Mur-Dueñas, P. (2015). Looking into ELF variants: A study of evaluative it-clauses in research articles. *Journal of English for Specific Purposes at Tertiary Level*, 3(2), 160-179.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578. <https://doi.org/10.1093/applin/amp044>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518. <https://doi.org/10.1093/applin/24.4.492>
- Pallotti, G. (2014). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134. <https://doi.org/10.1177/0267658314536435>
- Shahid, U., Anwar, B., El-Dakhs, D. A., & Younas, M. (2026). Complexity in academic discourse: A corpus-based analysis of nominal groups in English research articles. *Social Sciences & Humanities Open*, 13, 102337. <https://doi.org/10.1016/j.ssaho.2025.102337>
- Swales, J. M. (1990). *Genre analysis*. Cambridge University Press.
- Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, 47(2), 420-430. <https://doi.org/10.1002/tesq.91>
- Trudgill, P. (1999). New-dialect formation and Dedialectalization. *Journal of English Linguistics*, 27(4), 319-327. <https://doi.org/10.1177/00754249922004741>
- Vygotsky, L. S. (1986). Thought and language (A. Kozulin, Ed. & Trans.; Original work published 1934).
- Weinreich, U. (2010). *Languages in contact: Findings and problems* (No. 1). Walter de Gruyter.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawai'i Press.
- Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a Lingua Franca academic writing. *Journal of English for Academic Purposes*, 43, 100798. <https://doi.org/10.1016/j.jeap.2019.100798>
- Yoon, S., & Park, S. (2021). Grammatical complexity of EFL learners' casual conversation at different proficiency levels. *Korean Journal of English Language and Linguistics*, 21, 599-616. <https://doi.org/10.15738/kjell.21..202107.599>