

A CORPUS-BASED COMPARATIVE ANALYSIS OF THE VERBS *START* AND *BEGIN* IN ACADEMIC WRITING

Afia Niaz

PhD Scholar English Linguistics, University of Sialkot

Email: afianiaz401@gmail.com

Ali Furqan Syed

Lecturer, Department of English, University of Sialkot, Pakistan

Email: alifurqan.syed95@gmail.com

Abstract

*This corpus-based study conducts a comparative analysis of the verbs *start* and *begin* in academic writing, focusing on their frequency, distribution, and contextual usage within Pakistani academic genres, including theses, essays, and newspaper articles from Dawn. Drawing from a dataset of 10 essays, 5 theses, and 2 Dawn articles, the research identifies 26 examples, revealing that *start* appears more frequently (20 instances) than *begin* (6 instances). Contexts show *start* often associated with dynamic, agentive actions such as beginning work, routines, or events, while *begin* is used in more formal or gradual initiations. A graphical comparison with the Michigan Corpus of Academic Spoken English (MICASE) highlights distributional differences, with *start* dominating in written Pakistani contexts. The findings underscore the role of synonyms in enhancing academic clarity and suggest implications for English language teaching in non-native contexts.*

Introduction

Corpus linguistics is the branch of linguistics that encompasses the compilation and analysis of collections of spoken and written texts as the source of evidence for describing the nature, structure, and use of languages (McEnery & Wilson, 2021). Corpus linguistics relies on naturally occurring data and language (Biber et al., 1999). It is systematic and real-life-based data. Corpus linguistics is not manual; it is not possible without computers, and it is objective (Nordquist, 2025). A method known as corpus linguistics makes use of enormous, electronically accessible collections of naturally occurring spoken language and written language (Sinclair, 1991). In this procedure, computers are a crucial and occasionally required tool (Kennedy, 1998). A synonym is a word that has the same meaning as another word or nearly the same meaning (Crystal, 2008). For example, *begin* and *start* (Merriam-Webster, n.d.-a; Merriam-Webster, n.d.-b).

The lectures start at noon.

I got the job and I begin work on Monday.

Synonyms should be used frequently since they improve writing quality and give readers a clear, distinct view of the material (Enago Academy, 2025). Furthermore, as will be discussed further below, it can enhance both written and speaking skills (Paperpal, 2024a). But employing complicated synonyms doesn't improve the text (University of Waterloo Writing and Communication Centre, 2025). As is common knowledge, academic writing is a formal type of writing that is employed in academic journals and publications (Scribbr, n.d.). In scholarly books and journal papers, you will find it (Paperpal, 2024b).

The written corpus includes authentic data; it is an objective and scientific approach (Biber et al., 1999). It contained far more words than the spoken corpus (McCarthy, 2012). Transcripts of planned or unplanned speech, including news broadcasts or elicited narratives and dialogues, can be found in corpora of spoken language (Xiao, 2008).

Research Objectives

- To identify the specific contexts in which *start* and *begin* are commonly used in academic writings
- To investigate the frequency and distribution of the words *start* and *begin* in academic texts

Research Questions

1. In what specific contexts are the words *start* and *begin* commonly used in academic writings?
2. What is the frequency and distribution of words *start* and *begin* in academic texts?

Significance of the Study

Analyzing the usage of words like *begin* and *start* can help identify whether they contribute to or detract from the clarity of the writing the factors affecting the choice of their complements, and the possible relation between the two choices in English Academic Writing. Researchers may explore alternative words or phrases that can better convey the intended meaning and enhance the overall quality of the text. Researching the usage of *begin* and *start* can assist writers in identifying suitable synonyms. By studying the contexts in which these words are used, writers can broaden their vocabulary repertoire and choose more appropriate alternatives.

Literature Review

Corpus linguistics has emerged as a pivotal methodology for examining language use through large-scale, empirical data, enabling objective insights into lexical patterns and synonymy (Biber et al., 1999). In academic writing, where precision and formality are paramount, the analysis of near-synonyms like *start* and *begin* reveals subtle distinctions that influence textual clarity and style. These verbs, classified as aspectual or inchoative, mark the inception of actions or states, yet their interchangeability is not absolute, often governed by syntactic, semantic, and contextual factors. Early corpus studies have demonstrated that such verbs exhibit preferences in complementation and register, providing a foundation for understanding their role in scholarly discourse (Leitner, 1994).

Aspectual verbs such as *begin* and *start* have been extensively studied for their syntactic behaviors across English varieties. In British, American, and Indian English, *begin* tends to imply gradual or predictable initiations, while *start* conveys sudden or dynamic onsets, with frequency variations tied to tense and text type (Leitner, 1994). For instance, *begin* shows a stronger past-tense orientation (up to 70% of instances), whereas *start* is more present-oriented, reflecting its use in imperative or subordinate structures. This distinction extends to non-native contexts, where cultural and linguistic influences may alter distribution, as seen in academic genres where formality favors one over the other.

Complementation patterns further differentiate these verbs, with corpus data indicating *begin* predominantly favors to-infinitive complements (e.g., 91.4% in written corpora), evoking stative or generic perspectives, while *start* leans toward -ing forms (57.6% in writing, rising to 75% in speech), aligning with dynamic or iterative events (Schmid, 1993). Using the Lancaster-Oslo-Bergen (LOB) and London-Lund corpora, analyses reveal that activities and processes influence these choices, with "*begin* + to" as the default in formal writing and "*start* + ing" in spoken or agentive scenarios. Such patterns underscore the *Horror aequi* principle, where structural avoidance impacts selection (Ruan, 2014).

Semantically, *begin* is associated with abstract, non-perceivable processes like cognitive or mental states, emphasizing intentionality and post-onset stability, whereas *start* links to concrete, human-instigated actions, focusing on transitional moments (Divjak & Gries, 2009). Behavioral profiling from corpora like ICE-GB shows collocations: *begin* with perception verbs (e.g., "realize"), and *start* with activity verbs (e.g., "talk"). Semantic prosody reinforces this, with *start* implying agentivity and abruptness, often in communicative or operational contexts, while *begin* suggests predictability in non-agentive events (Schmid, 1993).

Register and genre play crucial roles, as academic writing favors *begin* in fields like medicine or biology for its formal tone, while *start* prevails in fiction or spoken academic English (Ruan, 2014). Comparisons across registers, such as those in the Longman Grammar, indicate

begin is more intransitive and prevalent in written data, with *start* outnumbering it in spoken American English (Biber et al., 1999). Diachronic trends from Google Ngram Viewer show *start* increasing in academic prose over time, approaching parity with " *begin*," influenced by social factors like domain-specific conventions.

Variations across English varieties highlight sociolinguistic influences: in British English, *begin* dominates narratives; in American, *start* is favored in spoken and factual styles; and in Indian English, tense balance shifts toward present for *start* (Leitner, 1994). These differences extend to complementation, with nominal forms more common for *start* and verbal for " *begin*," reflecting notional properties like gradualness versus abruptness. Such findings inform cross-cultural analyses, particularly in non-native academic contexts where synonym choice affects readability.

The implications for academic writing are significant, as VARBRUL modeling predicts verb and complement choices based on syntactic tags, inflection, and social factors, aiding natural language processing and teaching (Ruan, 2014). By broadening vocabulary through synonym awareness, writers can enhance clarity, avoiding over-reliance on one verb. Future studies could explore non-Western corpora, like Pakistani academic texts, to assess how local contexts diverge from global patterns, potentially revealing hybrid usages influenced by L1 transfer (Divjak & Gries, 2009).

Methodology

The researcher collected data from several academic genres, such as thesis and essays, in this research report, the quantitative research design was applied. The researcher studied several thesis themes and articles in Dawn publications before analyzing all of the data to determine how frequently the terms *start* and *begin* are used in Pakistani academic genres. The researcher studied 10 English essays, 5 thesis/articles, and 2 Dawn newspapers for finding the two words " *start* and *begin* " but instead of these lemma words only second or third forms are used in academic writings like in thesis, essays, or Dawn essays. After studying so many academic genres the researcher collected these examples which are mentioned below in table 1.

Data Analysis

In this paper the researcher took examples from different academic genres and analyzed each of them with the Pakistani contextual references which is shown is Table 1.

Table 1

Examples and Contextual References of Words Start and begin

Examples	Contextual Reference
1. The job asked me how soon could I <i>start</i> and I wanted to make sure it was right.	In this sentence, "appears to be a statement made by an individual in the context of a job application or interview. The person is relaying that during the job application process, they were asked about their availability to <i>start</i> working.
2. Since the <i>start</i> of Syria's war and was expelled from the Cairo-based Arab League in 2011 over its violent crackdown on pro-democracy demonstrations.	The sentence refers to the events surrounding Syria's war and its expulsion from the Cairo-based Arab League in 2011. The sentence suggests that these two incidents are connected. Overall, the contextual reference of this sentence is the political and social turmoil in Syria particularly, regarding the war and the Arab League's response to the Syrian government's actions.
3. There is nothing like a morning walk to give a nice <i>start</i> to our day.	The contextual reference of the sentence suggests that the speaker is expressing the positive impact and benefits of taking a morning walk. The context could be a conversation or a written statement discussing the importance of morning

	routines or healthy habits.
4. That's why many doctors advise their patients to <i>start</i> morning walk because of their incredible result.	The sentence implies that doctors commonly recommend morning walks to their patients due to the positive outcomes or “incredible results”. The context could be a conversation or a written discussion about healthcare, exercise, or lifestyle choices.
5. A morning quick walk in the fresh air is the best way to <i>start</i> the day.	The sentence implies that the speaker considers taking a quick walk in the morning and breathing in fresh air to be the most optimal or beneficial way to <i>start</i> the day.
6. I was about to <i>begin</i> my first job, at sixteen, over 2000 miles from home, as a summer hire at the United States Embassy in Bogota.	The sentence indicates that the speaker, who is sixteen years old, is about to <i>start</i> their first job. Furthermore, it states that the job is located in Bogota, which is over 2000 miles away from their home. The job specifically mentioned is a summer hire position at the United States Embassy in Bogota.
7. Thus, it shows how through great work one can get the key to success. When you <i>start</i> working hard you will notice changes in your life.	The context could be a motivational speech, a self-help article, or a conversation focused on personal growth and success. The contextual reference of the sentence when you <i>start</i> working hard, you will notice changes in your life.” Suggests a general statement about the relationship between hard work and success.
8. Moreover, you will <i>start</i> seeing results within a short time.	The meaning of the sentence could vary depending on the topic or subject being discussed prior to this statement. However, in general the sentence implies that someone (referred to as “you”) will experience or observe some kind of desired outcome or progress in a relatively brief period.
9. There are many reasons why travelling abroad at an early age is crucial. For a <i>start</i> , it is the proper time to make life.	The sentence implies that there are multiple compelling reasons or arguments supporting the idea of travelling abroad at an early age. The context may involve a conversation, article, or essay centered around personal growth, educational opportunities or broadening one’s perspective through international travel. The contextual reference suggests that travelling abroad during one’s early years can have a significant and positive impact on personal development and life experiences.
10. It is no mystery that unattended food will eventually <i>start</i> to burn.	The sentence implies that it is widely known or expected that if food is left without supervision or attention for a significant amount of time, it will eventually catch fire or become burnt. The context could be a discussion about cooking, kitchen safety, or general awareness regarding the potential dangers of neglecting food on heat sources such as stovetops, ovens, or grills.
11. One may <i>start</i> from scratch and end up as a business tycoon.	The context may involve discussions on entrepreneurship, success stories of self-made business leaders, or motivational narratives emphasizing the potential for upward mobility and achieving significant financial success through one’s own efforts and skills that <i>starting</i> from humble <i>beginnings</i> does not necessarily limit one’s potential for attaining extraordinary success in the business world.
12. Then one can <i>begin</i> his	The context may involve discussions about career choices,

own workshop, factory or shop, or choose a career in police, films commercial art etc.	personal aspirations, or opportunities for self-employment. The context highlights the range of possibilities available to individuals beyond entrepreneurship and encourages exploring different career paths based on individual interests and talents.
13. Several bombs were reportedly defused before voting <i>begin</i> on Saturday morning, according to Al-Jazeera.	The context likely revolves around a specific election or voting process, where the presence of these bombs posed a potential risk to the safety and security of the voting procedure or those involved. The sentence suggests a situation where preemptive action was taken to neutralize a threat posed by several bombs before the voting process commenced, highlighting the significance of maintaining security during elections or other important democratic process.
14. Some parties <i>begin</i> campaigning in Pakistan as early as March 27, six weeks ahead of the May 11 election date.	The context likely revolves around discussions or reports on the electoral process, political strategies on the electoral process, political strategies, or the dynamics of political campaigning, building momentum, and engaging with voters in order to maximize support and secure electoral success. The sentence highlights the timeline and practices of political campaigning in Pakistan, specifically pointing out that some parties commence their efforts nearly two months prior to the election date.
15. They can choose the most suitable and comfortable time for them to <i>start</i> and finish their work with less stress.	The context may involve discussions or descriptions of flexible work arrangements, such as flexitime or remote work options. It could pertain to workplaces or occupations that prioritize employee well-being, productivity, and customization of work hours to fit individual needs promoting a more comfortable and stress-free working environment.
16. It is better to <i>start</i> with the description of the advantages of work at home to understand the disadvantages of working in an office.	The sentence implies that in order to gain a comprehensive understanding of the disadvantages associated with office work, it is preferable to <i>start</i> by highlighting the advantages of working from home. By establishing the positive aspects of remote work, one can then contrast them with the potential drawbacks or limitations of working in an office environment. The sentence suggests that by first exploring the benefits of remote work, one can gain a clearer understanding of the potential disadvantages or limitations associated with working in a traditional office setup.
17. It is simple rule, when the car ahead of you passes a permanent object on the side of road, <i>start</i> counting and at least 3 second passed before your car passes the same object and at night in bad weather condition this second can be doubled.	The context likely revolves around discussions or instructions related to defensive driving, safe driving practices, or road safety. It aims to promote a proactive approach to maintaining safe distances and reducing the risk of rear-end collisions especially in conditions that may require additional caution. The sentence provides guidance on maintaining a safe following distance while driving, highlighting the importance of adjusting the interval based on visibility and weather conditions.
18. From the <i>start</i> , I	The sentence conveys the idea that due to a lack of

<p>encountered several problems due to my lack of experience.</p>	<p>experience, the speaker or narrator faced challenges and encountered problems right from the <i>start</i> of their journey. It emphasizes the learning process and highlights the importance of overcoming initial hurdles to grow and develop proficiency in a given area.</p>
<p>19. I got up early that morning due to an improper sleep during the previous night. It was due to the anxiety to <i>start</i> fresh with my new life.</p>	<p>The context may involve discussions or reflections on personal transitions, such as <i>starting</i> a new job, moving to a different city, <i>beginning</i> a new relationship, or any significant life change. The mention of wanting to “<i>start</i> fresh” implies a desire for a new <i>beginning</i> or a fresh <i>start</i> in various aspects of their life. The sentence indicates that the person’s disrupted sleep was a result of the anticipation and nervousness surrounding their new life journey.</p>
<p>20. The mixture of excitement and nervousness <i>start</i> to fill within me.</p>	<p>The sentence conveys the emotional response or internal state of the individual as they navigate the anticipation and potential apprehension associated with the upcoming event or situation.</p>
<p>21. Could I <i>start</i>, and I wanted to make sure it was right at 6 am, I stood in front of my apartment building, squinting at the sun rising at the sun rising over the Andes Mountains.</p>	<p>In this sentence, the mention of squinting at the sun rising over the Andes Mountains provide a visual description of the scene. This implies that the person is experiencing the beauty and serenity of the sunrise, possibly to gather inspiration or to appreciate the natural surroundings before commencing their intended activity.</p>
<p>22. I <i>begin</i> my study which continues undisturbed till 8.00 p.m. From 8.00 p.m. to 9.00 p.m. the entire family watches two TV serials.</p>	<p>The sentence portrays a structured routine where personal study time occupies a significant portion of the evening, followed by a designated hour for family members to come together and enjoy television programs. It reflects a balanced approach to individual growth and familial engagement.</p>
<p>23. In fact, the entire family is addicted to these two serials and has been following up the story from the <i>start</i>.</p>	<p>The contextual reference of the sentence suggests that the speaker is referring to two specific television serials that the entire family is addicted to. The family has been closely following the story of these serials since the <i>beginning</i> or from the <i>start</i>.</p>
<p>24. I <i>begin</i> my day early, around 5 am. I spent the first hour of my day exercising with my brother.</p>	<p>The contextual reference of the sentence suggests that the speaker is describing their daily routine. They mention that they <i>start</i> their day early, around 5 am. They engage in exercise or physical activity with their brother.</p>
<p>25. I <i>start</i> with the daily newspaper, followed by my curriculum subjects.</p>	<p>The sentence indicates that the speaker is describing their approach to studying or their daily routine related to academics. They mention that they <i>start</i> their day by reading daily newspaper, likely to stay updated with current events and general knowledge. They proceed to study their curriculum subjects, implying that they allocate time for academic learning or studying specific subjects, relevant to their curriculum.</p>
<p>26. Early morning walk prepares the body physically and mentally for daily work.</p>	<p>The contextual reference of the sentence suggests that the speaker highlighting the benefits of morning walk. By engaging in an early morning walk, they believe that their</p>

Thus, we *start* or daily routine work early and with pleasant mood. It is due to morning walk.

physical and mental state is improved, leading to a positive and pleasant mood. They are able to *start* their daily routine work early and with a positive mindset, attributing this positive *start* to the morning walk.

Results

The findings of the current investigation are presented and discussed in this section. The findings will be presented in accordance with the preceding study's research questions. This section contains the findings that are intended to address the second research question: What is the frequency and distribution of words *start* and *begin* in academic texts? In this section there is a graph which showed the frequency of words *start* and *begin* in Pakistani context and in MICASE.

Conclusion

The corpus-based analysis reveals that in Pakistani academic writing, the verb *start* is more frequently employed than " *begin*," appearing in 20 of the 26 examined examples, often in contexts involving dynamic actions, daily routines, or sudden initiations, such as *starting* work or experiencing changes. In contrast, *begin* is used sparingly, typically for more formal or gradual *beginnings*, like commencing studies or campaigns. The graphical comparison with MICASE indicates higher *start* distribution in written Pakistani genres compared to spoken academic English, highlighting register-specific preferences. Overall, these findings affirm the non-interchangeable nature of these synonyms, contributing to clearer academic expression and underscoring the value of corpus methods in lexical studies.

Recommendations

Based on the findings, academic writers in Pakistani contexts should prioritize *start* for dynamic, agentive scenarios to align with prevalent usage, while reserving *begin* for formal, abstract initiations to enhance precision. Educators are encouraged to incorporate corpus tools in English language curricula to teach synonym distinctions, fostering vocabulary expansion and reducing redundancy. Further research could expand the corpus to include more diverse genres or compare with other South Asian varieties for broader insights. Additionally, integrating VARBRUL-like models into writing assistance software could aid non-native speakers in selecting appropriate verbs and complements.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell Publishing.

Divjak, D., & Gries, S. Th. (2009). Corpus-based cognitive semantics: A contrastive study of phasal polarity items in English and Russian. In K. Davidse, L. Vandelanotte, & K. Davidse (Eds.), *Recent advances in cognitive linguistics* (pp. 57–75). Mouton de Gruyter.

Enago Academy. (2025, January 28). How to use synonyms effectively in a sentence. <https://www.enago.com/academy/how-to-use-synonyms-effectively-in-a-sentence/>

Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.

Leitner, G. (1994). *begin* and *start* in British, American and Indian English. *Hermes - Journal of Language and Communication in Business*, (13), 103–122.

McCarthy, M. (2012, August 21). What is corpus? The 21st Century Text. <https://21centurytext.wordpress.com/home-2/special-section-window-to-corpus/what-is-corpus/>

McEnery, T., & Wilson, A. (2021). *Corpus linguistics* (2nd ed.). Edinburgh University Press.

Merriam-Webster. (n.d.-a). *begin*. In *Merriam-Webster.com thesaurus*. Retrieved December 28, 2025, from <https://www.merriam-webster.com/thesaurus/ begin>

Merriam-Webster. (n.d.-b). *start*. In *Merriam-Webster.com thesaurus*. Retrieved December 28, 2025, from <https://www.merriam-webster.com/thesaurus/start>

Nordquist, R. (2025, May 12). Definition and examples of corpus linguistics. ThoughtCo. <https://www.thoughtco.com/what-is-corpus-linguistics-1689936>

Paperpal. (2024a, March 8). What are synonyms and how do you use synonyms in academic writing. <https://paperpal.com/blog/academic-writing-guides/language-grammar/what-are-synonyms-in-writing-how-to-use-it-correctly>

Paperpal. (2024b, February 23). Characteristics of academic writing. <https://paperpal.com/blog/academic-writing-guides/language-grammar/characteristics-of-academic-writing>

Ruan, J.-C. (2014). A corpus study on *begin/start* in academic writing: A VARBRUL approach. *Open Journal of Modern Linguistics*, 4(2), 209–228. <https://doi.org/10.4236/ojml.2014.42017>

Schmid, H.-J. (1993). *Introspection and computer corpora: The meaning and complementation of start and begin*. Unpublished manuscript, Ludwig-Maximilians-Universität München. https://www.anglistik.uni-muenchen.de/personen/professoren/schmid/publ/start_and_begin.pdf

Scribbr. (n.d.). What is academic writing? Dos and don'ts for students. <https://www.scribbr.com/category/academic-writing/>

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

University of Waterloo Writing and Communication Centre. (2025, July 22). The power of synonyms: The good, the bad, and the in-between. <https://uwaterloo.ca/writing-and-communication-centre/blog/power-synonyms-good-bad-and-between>

Xiao, R. (2008). Corpus creation. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 147–168).