ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

A CROSS-DISCIPLINARY CORPUS ANALYSIS OF MORPHOSYNTACTIC PATTERNS IN UNDERGRADUATE RESEARCH PAPERS

Syed Muhammad Aziz Ali Naqvi
Lecturer in English,Govt.M.A.O. Graduate College, Lahore.
azizali92@gmail.com
Mohammad Aafaq Nadeem (CorrespondingAuthor)
MPhil Scholar,The University of Lahore Sargodha Campus
iamaafaq84@gmail.com
Naila Aslam
MPhil Scholar,The University of Lahore Sargodha Campus
Teaching in English in Affiliation with Agha Khan Board,
Editor in Peace Magazine
nailaaslam624@gmail.com

Abstract
This paper aims to investigate morphosyntactics of undergraduate research papers using a cross-disciplinary corpus research to indicate how these linguistic features represent disciplinary ideals and language development in the novice scholarly writing. The morphosyntactic constructions nominalizations, passive structures, verb usage, and subordination of clauses-are needed in building an objectivity, abstraction and rhetorical persuasion of scholarly writing. What matters most in their analysis is their undergraduate level, they have left the general writing practices behind and now adopt discipline writing practices, which as we all know, can be quite difficult, since most students will have to work in different epistemological requirements in fields. In this research, three research questions are in focus, namely: (1) What are the most frequently occurring morphosyntactic peculiarities of undergraduate research papers? (2) What are the differences in these features in different disciplines, especially in humanities/social sciences and STEM? (3) What do these patterns imply in regards to teaching academic writing? In order to create a balanced sample of about 500,000 words 100 anonymized undergraduate research papers (20 humanities, 20 social sciences, 20 natural sciences, 20 engineering, 20 business) were collected through open-access university archives with complete ethical clearance. Preprocessing (tokenization and lemmatization) through spaCy and morphosyntactic annotation through the Universal Dependencies framework through Stanza were used to work with texts. Findings show moderate general usage of passive voice (32.4 per 1,000 words) and nominalizations (45.6 per 1,000 words) and a dominance of present simple tense (52-percent of verbs). Large differences in disciplines were obtained: STEM subjects had more passive constructions (maximum 42.1) and more nominalization density (maximum 55.3) to make their content impersonal and compress information, whereas humanities and social sciences used more complex subordination (maximum 22.4 per 1,000 words) to make their work argumentative. In addition to extending corpus linguistics by testing the application of the register variation theory to novice writing, and offering practical implications to discipline-specific English to Academic Purposes pedagogy, they recommend specific training on morphosyntactic features to enable them to develop their rhetoric and interdisciplinary literacy.

Keywords: corpus analysis, morphosyntax, undergraduate writing, cross-disciplinary variation, academic discourse, English for Academic Purposes

Introduction
The foundation of academic communication is academic discourse that involves linguistic practices that do not only disseminate knowledge but also build disciplinary identities and provide intellectual discourses between the professionals and students. The central node of such discourse is morphosyntactic patterns-compositions of structures used to form the cap clarity, precision and persuasiveness of the written material. These patterns, which include nominalizations, passive constructions and subordinate clauses, allow writers to negotiate the ability to be objective, argumentative and complex requirements of research writing (Biber et al., 2021). Basing their approach on the paradigm of the Systemic Functional Linguistics (SFL) introduced by Halliday (1978/2014), morphosyntax is considered as multifunctional

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

resources achieving ideational (capturing experience) interpersonal (enacting relationships) and textual (organizing flow of information) meanings. In scholarly settings, such as the spreading of the nominalized forms of verbs in scientific prose turn dynamic processual events into entities and thus promotes the increase of abstraction and impersonality which are a mandatory requirement of evidential assertions (Halliday and Martin, 1993; revised applications in Thompson et al., 2023). As a functional lens, complementary to SFL, the Genre Analysis, as defined by Swales (1990) holds that the choices of these morphosyntactic strategies are genre-specific rhetorical act of discursive communicative endowment. An example of syntactic embedding used in introductions to signify gaps or counterclaims placing the writer in current scholarly debates can be the Create-A-Research-Space (CARS) model developed by Swales (2011); it has been revisited recently by Flowerdew (2022). All these frameworks emphasize the fact that morphosyntactic patterns are not neutral, they are socio-rhetorical instruments mediating the construction of power, authority, and knowledge in academia.

Such a combination of both, both functionally and genrally speaking, is especially prominent when we consider the case of undergraduate research papers, which occupies the liminal zone between amateur experimentation in genre and professional mastery. Contrary to refined professional writing, undergraduate writing will display developmental curves of the emergent dominance over morphosyntactic resources, subjected to L1 transfer, experience, and insufficient exposure to disciplinary imprint, and the cognitive pressures of formation of ideas (Nasseri, 2021). According to the recent corpus-based research, the first-time writers prefer simpler syntactic constructs, including finite clauses and active voice, that put the narrative over the hypotactic embedding common in the professional texts (Casal et al., 2021). Considering the Michigan Corpus of Upper-Level Student Papers (MICUSP), it can be indicated that undergraduates in social sciences had more high percentages of first-person pronouns and coordinate clauses in their papers, which is indicative of personal involvement, yet at times their disposition can degrade the impersonal role required in humanities genres (Hyland, 2002a; further extended by Csomay and Pollard, 2022). Unlike that, scientific articles in STEM disciplines use phrasal complexity with prepositional phrases and non-finite clauses to compress information and upfront causation, multidimensional register analyses demonstrate that there is higher informational density in expert writing (Biber and Gray, 2016; revised in Biber et al., 2023). These variations can be attributed to development stages: undergraduates, who struggle with the apprenticeship of morphosyntactic, are forced to confront the so-called metalinguistic consciousness when the choice of morphosyntactic is consciously learned, and it is not deployed instinctively (Ortega, 2012; recent empirical evidence of this in Tarasova and Beliaeva, 2023). These patterns not only provide insights into lacks of proficiency but also explain how writing develops as a self-centered mode of expression into a community-oriented form of writing that is consistent with the features of SFL that assert that language is a social semiotic system that is contextually oriented (Halliday, 2004; implemented in Llinares and McCabe, 2023).

Based on these observations, Genre Analysis explains more extensively the morphosyntactic variations as strategic actions in view of disciplinary demands. Managing introductory clauses, when a precursor section addresses the thesis statement, the move-step model offered by Swales (1990) demonstrates that in the humanities, arguments are typically woven by using hypothesizing politeness, this is, through hypotactic clauses, whereas in the humanities, it is the form of parataxis that promotes the linearity of the procedures (Kanoksilapatham, 2015; cross-genre extension in Loi et al., 2022). Based on this genre-sensitive strategy combined with SFL, it has been shown to be effective in depackaging the defunctionalization of advanced language in undergraduates; e.g., a recent study of abstract writing has

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

discovered that novice engineers do not make robust uses of nominalizations, which results in less robust statements of purpose than is strongly done by their colleagues in biology (Sagre et al., 2021; replicated by Yakhontova, 2024). But these developmental differences do not just end at the level of individual competence, but to institutional and pedagogical level, in which undergraduate papers are used as formative artifacts when designing curriculum. Since interdisciplinary cooperation is becoming important in universities, learning morphosyntactic patterns is a significant phenomenon in this genre because it provides the scaffolding of the transition between a general education and a specialized research (Hyland and Jiang, 2019; recent pedagogical implications in Walsh Marr, 2023). Therefore, although professional discourse represents a state of adequate genre conventions, undergraduate writing represents a vibrant act of negotiating these conventions, which is an ideal place to investigate the corpus in the context of language socialization.

There has been substantial gaps in the corpus linguistic information; more so on the area of cross-disciplinary analysis of undergraduate morphosyntax. Preferentially, the available literature focuses on expert-written texts, including research articles in monolithic corpora such as the British Academic Written English (BAWE), in which published work is better privileged over emerging student work (Biber et al., 2013; criticised in Flowerdew, 2022). When undergraduate corpora is studied, they are often isolated in one discipline (as MICUSP focuses on upper-level essays in isolated disciplines such as electrical engineering or history) and do not consider the interactivity across the humanities, social sciences, and STEM (Csomay & Pollard, 2022; Khany and Kafshgar, 2016). This monodisciplinary bias does not take into account the possibility of morphosyntactic preferences in hybrid undergraduates, including the increased subordination in argumentative papers in the humanities and the increased nominal density in scientific writings (Casal et al., 2021). Additionally, not many works combine both the quantitative frequencies counts with the qualitative move analysis to trace the development pattern in the selected manner, leaving the gaps between the studies on L1/L2 variations or genre development (Nasseri, 2021; gaps outlined in Tarasova and Beliaeva, 2023). A corpus study of syntactic properties among Pakistani academic writing samples, e.g., validated disciplinary deviation, but recommended more extensive undergraduate sample sizes to resolve the problem of representativeness (Shahid et al., 2023). These gaps do not only restrict the possible theoretical depth available in SFL and Genre Analysis but also impair applicable applications where EAP teachers do not have evidence-based material to use in customizing instructional delivery to diverse majors.

It is this issue that has highlighted the urgency of a cross-disciplinary corpus analysis of the undergraduate research papers: to provide these gaps and harness the informational facts in English for Academic Purposes (EAP) teaching in the era of curriculum internationalization. The globalization of higher education leads to the greater extent of traversal of disciplinary boundaries by students, which is accompanied by the EAP programs performing general grammar drills not related to the requirements of hard grammar and morphosyntactic (Hyland, 2018; recent criticism in Charles, 2023). The interdisciplinary investigation uncovers how the patterns, such as the prevalence of passive voice in biology and modal hedging in philosophy, give rise to the development of discipline-adequate positions, which can subsequently be used to implement interventions facilitating the development of genre awareness (Isik-Tas, 2018; elaborated in Wang and Zeng, 2021). Corpus born profiles are relevant to EAP practitioners EAPs rely on needs analysis models, with a syllabus design prioritizing non-finite clauses to teach STEM beginners approximation of expertly abstract thinking (Accurso and Gebhard, 2021; implementation in Blair and Avalos, 2025). In addition to that, with the associated increase in multilingualism in undergraduate groups, morphosyntactic differences can be neutralized, facilitating equal access to academic

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

literacies (Ortega and Byrnes, 2021; ramifications of inclusive pedagogy directive in Sarkar, 2024). Highlighting such patterns, the study does not only expand the register theory of SFL to enable Halliday metafunctions in novice settings, but also is consistent with Swales idea of genres as communicative events that can be tailored to the trajectories of the learners (Swales, 2011; reexamined in Sawaki, 2023). Finally, this argument makes the study a booster of EAP reform, as it enables the instructors to develop rhetorical agility within the various disciplinary settings.

To deal with these aims, the research questions in this study will be the following: (1) Which are the most common morphosyntactic patterns of undergraduate research papers in a wide range of disciplines? (2) To what extent do such patterns differ between humanities/social sciences and STEM, specifically concerning the complexity of clauses and the use of voice? (3) What does these differences imply with regard to EAP writing teaching? The major objectives are based on a mixed-method corpus approach, such as the compilation of a balanced anonymized corpus of 200 undergraduate papers (around 500,000 words) across five disciplines, morphosyntactic features annotation with tools such as Universal Dependencies (de Marneffe et al., 2021), and quantitative analysis of frequency and qualitative analysis of genre moves.

Literature Review

Morphosyntax as a pillar element in the analysis of linguistics combines morphology (the study of how words are built through inflection (e.g. -ed in walked) and derivation (e.g. changing verbs to nouns using suffixes such as -tion in investigation); and syntax (how the structures of clauses are arranged, how subordination is organized, how the sentences are organized). This interaction gives users of language the power to build meaning on a variety of levels, including word-internal variation, to complex hierarchical interrelations between clauses, which are expressed in Systemic Functional Linguistics (SFL) through articulation of experiential, interpersonal and textual meta functions through morphosyntactic options (Halliday and Matthiessen, 2014). Such ideas were recently adapted to the professional contexts to indicate a sense of temporality and aspect in academic arguments using inflectional changes, with nominalization compaction of information being facilitated by processes of derivation to increase abstraction (Thompson et al., 2023). Subordination with relative clauses or adverbials is syntactically easier to do, which makes it easier to embed in dense prose, and paratactic constructions aid in linearity in descriptions of processes (Biber and Gray, 2016). These are not fixed components, but rather adapt to communicative requirements; an example of this is more modern SFL usage, which connects morphosyntactic patterns with digital academic genres and is a manifestation of the change to phrasal complexity of online scholarly language (Llinares and McCabe, 2023). Genre Analysis is in addition to SFL, with morphosyntax framing tools as parts of the discourse community, where the complexity of the clauses as indexes of movement relates to trying to create territory or finding a niche of their own (Swales, 1990; revised in Flowerdew, 2022). These theoretical frameworks offer a tool through which it is possible to analyze academic writing with the focus on the fact that morphosyntactic choices are not accidental but context-specific decisions that shape the knowledge building process at any level of proficiency.

Corpus Linguistics in Academic Writing

The shift to empirical investigation Starting with theory Using multidimensional analysis (MDA), corpus linguistics has transformed the field of academic writing: it now makes possible, on a large scale, data exploration of the variations of the registers. The MDA framework pioneered by Biber et al. imposes labels on texts by splitting them into dimensions as informational or involved produce, narrative concern, and abstract style to provide clues of how academic prose concentrates on dense and non-narrative properties

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

(Biber, 1988; the latter expanded to more recent in Biber et al., 2023). To illustrate, as of October 2021, the latest versions of MDA app have added functionality to model neuralied, a transformer-based application that extracts stylistic features with a higher degree of interpretability and display register shifts in translated poems and online registers (Gries et al., 2021; Biber et al., 2021). In line with this, research conducted by Hyland on disciplinary discourse examines the importance of interactional meta discourse, i.e., factors such as hedges (e.g., might suggest) and boosters (e.g., clearly demonstrates) as persuasive and alignment discourse strategies in academic texts (Hyland, 2018; Hyland and Jiang, 2019). The latest works such as a preface on how nouns are used in meta discursive position are reiterating the shift in tendencies in international print media, in which the academic discourse is becoming more multilingual (Hyland, 2022). Such corpus-based methods have shed light on the systemic tendencies, including the use of nominal phrases in hard sciences to obtain an objectivity, which connects back to morphosyntactic theories by demonstrating that derivational morphology is the basis of register-specific compression (Biber and Gray, 2016). In addition, cross-linguistic MDA researches have been shown to reveal universal such as less overt persuasion in non-native academic registers that facilitate a shift between theoretical morphosyntax and practical discourse analysis (Gries, 2014; Baumgart, 2021).

Cross-Disciplinary Variations in Academic Genres

In further examination of the finer aspects of discipline, cross-disciplinary differences in academic genres become fully manifested in the differing morphosyntactic tendencies that bring about differences in epistemology. In sciences, nominalizations prevail in order to embrace objectivity and impersonality, making action become something (e.g., analysis is replaced by we analyzed), which corpus research indicates is more common in STEM research papers than it is in humanities (Biber et al., 2023; Shahid et al., 2023). In contrast, humanities texts prefer hedging through modals and adverbials, which allow the modulation of an interpretation and dialog interaction, and research found that such subordination is more prevalent concerning argumentative complexity (Hyland, 2002; Csomay and Pollard, 2022). Recent cross-disciplinary corpus studies do so further, considering frame markers, such as in this section or to summarize, which organize discourse, with recent research showing that both sciences and social sciences use them to organize discourse procedurally and thematically respectively (Bondi, 2010; recent study in Parodi, 2015). Even lexical bundles, repetitive multi-word patterns, have different variation: structural bundles (because of for causality) are more common in empirical fields whereas referential ones (in the context of for framing) are used in interpretive ones (Unveiling Core Lexical Bundle, 2024). Complicated syntax also tends to divide, as a 2023 study of Pakistani academic writing in four academic fields found increased use of clauses to persuade and phrasal modifiers to argue in favor of precision, respectively (Shahid et al., 2023). Based on the register theory of SFL, these variations point to the alignment of morphosyntax to genre requirements in general, though they are largely based on the expert corpus, which underscores the necessity of extending these applications to beginners writing in general.

Undergraduate Writing and Learner Corpora

Moving to the setting of undergraduates, the study of L1 and L2 learner corpora also indicates developmental morphosyntactic tendencies that do not conform to the norms of the experts and are usually distinguished by excessive use of simple clauses and insufficient application of the derivational complexity. Another resource that has played a central role in the discovery of transfer-related mistakes that hinder coherence is the International Corpus of Learner English (ICLE) which contains non-native undergraduate argumentative essays and is used to identify them, such as mismatched inflections or awkward subordinations (Granger et al., 2009; recent applications in Nasseri, 2021). whereas expert discourse is preferred to the

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

hypotaxis to render an abstraction, the undergraduates display parataxis, active voice and indicate an instinct of genre awareness (Casal et al., 2021; Tarasova and Beliaeva, 2023). A 2024 corpus-based study of morphosyntactic characteristics of Pakistani undergraduate writing revealed a lack of consistency in the tense use and nominal agreement but credited it to L1 interference. LiBigger collections of student paper such as the Michigan Corpus of Lower-Year Student Papers (MICUSP) are more vulnerable to disciplinary discrepancies, with humanities students being much more exposed than STEM ones (Csomay and Pollard, 2022). Nevertheless, they do not tend to be balanced in terms of fields of study and their reliance level, and any metadata issues in the school corpora such as GiG and TRAWL are restricting representativeness. These restrictions impede the understanding of how morphosyntactic resources are negotiated by undergraduates in scholarly articles, especially when the translation process takes place in a multilingual environment with the assistance of translanguaging but more complex regularities (Enhancing academic writing, 2025).

Methodological Approaches in Corpus-Based Studies

Corpus linguistics as a methodology involves using a collection of tools that are used to quantify and interpret these patterns to guarantee that the academic writing morphosyntax is analyzed in a reproducible manner. The freeware toolkit AntConc is the best toolkit at concordancing, showing the keywords within the context, which is very useful when studying inflectional variants or subordinators with the built-in statistics, such as the collocation measures that indicate the strength of association (Anthony, 2023; Corpus Analysis with Antconc, 2015). MDA, and make cross-register comparisons with scripts to normalize by a thousand words. Hybrid methods are used to integrate each one: e.g. the n-gram tool of AntConc finds bundles and ggplot2 of R can be used to visualize differences in syntactic complexity across disciplines (Advancements in Corpus Analysis Tools, 2025). Other systems such as Sketch Engine provide morphological tagging, which facilitates derivation analyses, but UAM Corpus Tool combines qualitative annotations of genre moves (Tools for Corpus Linguistics, undated). Recent methodological fine tunings focus on the area of ethical data processing in learner corpora, including R scripts of anonymization, filling gaps in undergraduate population. These mechanisms applied repeatedly can bridge the quantitative with qualitative measures as seen in materials tagged with AntConc, which find errors in morphosyntactic analysis of student texts.

These studies lead to these methodological advantages, sustained gaps in the literature demonstrate research gaps that can be addressed in specific research especially on cross-disciplinary undergraduate corpora. The majority of the literature gives preference to expert articles in research, neglecting novice morphosyntax when working with balanced samples, both in humanities and social sciences and in STEM (Flowerdew, 2022; A corpus-based comparison, 2023). To take an example, disciplinary divergence in hedging can be well-documented in MICUSP, but very seldom reaches out to morphosyntactic depth in undergraduate research papers, gap in knowledge about matters of developmental trajectory (Hedging in academic writing, 2022). Learner corpora such as ICLE do not possess the interdisciplinary scope but are frequently advances to tackle L2 concerns; percentages have been suggested to solve the biases (Learner corpora around the world, undated; Comparing Learner Corpora, undated). Syntactic differences in the argumentative essays are also evident through cross-disciplinary studies that do not concentrate on the research genres at the undergraduate study level (Disciplinary Variation, 2025). This paper addresses these gaps by producing an equal amount of undergraduate research articles as a balancing corpus, to which MDA and genre tools are applied to morphosyntactic patterns, and thus can use EAP pedagogical practice with informed advice on disciplinary-specific instruction (Charles,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

2023). It also fills in these gaps and serves to theoretically integrate SFL and Genre Analysis as well as increasing methodological rigor in beginning discourse studies.

3. Methodology

The research design is a mixed-methods and corpus-based investigation, which utilizes quantitative analysis when analyzing morphosyntactic patterns, but which is complemented by qualitative considerations to elucidate the functions of the context, as well as the rhetorical implications. This is not an exceptional strategy in corpus linguistics, where extensive frequency data is triangulated with act-based analyses to produce strong and replicable results that could be published in a high-impact journal (Biber and Conrad, 2019; Charles, 2023). The methodology is highly accurate in morphosyntactic annotation, with a chance to explore the variation of the disciplinary nuances of academic discourse, making the methodology meet the requirements of the argument of authors who suggest that multimodal analyses are needed when studying undergraduate academic discourses (Csomay and Pollard, 2022; Nasseri and Borlongan, 2024).

3.2 Corpus Compilation

The compilation of corpora started with the nondispersant compilation of undergraduate research papers, both publicly available in university repositories and open-access digital archives, such as e.g., institutional thesis databases and platforms such as those of extensions of the British Academic Written English (BAWE) corpus or other learner corpora. Human subjects research procedures were taken into account, and even in the case of the secondary analysis of existing texts, ethical approval was given by the Institutional Review Board (IRB) before the collection of data (American Psychological Association, 2020; Townend et al., 2024). Each of the papers was completely anonymized, both by eliminating the names of authors, Institutional affiliation, and references to the person, and by only including metadata as to the type of discipline and approximate word count to ensure privacy. To increase the representativeness, the sources were selected among English-medium Universities in various regions based on the recommendations of creating ethical corpus in educational institutions (Granger et al., 2023).

The resultant corpus includes around 500,000 words that is balanced by having 100 undergraduate research papers (20 each discipline) of five major fields: humanities, social sciences, natural sciences, engineering, and business. The composition and size of this size allow making reliable statistical comparisons at the same time as it is manageable in terms of more detailed annotation, similar to recent undergraduate corpora such as extensions of the Michigan Corpus of Upper-Level Student Papers (MICUSP) or specialized learner corpora (Romer and Berger, 2023; Nasseri and Borlongan, 2024). Inclusion criteria were full research papers (e.g. final-year projects or capstone theses) in English language with a minimum length of 3000 words to guarantee that the text is morpho syntactically sophisticated enough; the inclusion criteria were limited to articles created within the last 10 years (2015-2025) to reflect contemporary standards in academic writing. Non-original works (e.g., plagiarized or group work without individual attribution), non-research genres (e.g., reflective essays), and texts feature non-native speakers were filtered out with exclusion criteria (only unbalanced), although there is a respect to a combination of L1 and L2 writers in the corpus to reflect the representation of undergraduate writers throughout the world (Ortega, 2021).

3.3 Data Processing

Preprocessing Data was processed by first standardizing the texts to be analyzed. The spaCy NLP library (version 3.7), which is a powerful Python utility, was used to tokenize and lemmatize the text and help work with academic prose efficiently (Honnibal et al., 2020). This guaranteed homogenous segmentation of words and minimization of base-form which is

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

important in providing correct counts of morphological characteristics such as derivation or inflection. Later morphosyntactic annotation used the Universal Dependencies (UD) framework (2.12), which is applied through the Stanza pipeline (Qi et al., 2020; de Marneffe et al., 2021): it features cross-linguistically universal part-of-speech tagging, morphological features, and dependency relationships. UD was chosen due to its ability to be applied universally as well as reliably to the process of annotating learner language, and it performed better than older taggers such as CLAWS when used in more complex syntax found in scholarly texts (Zeman et al., 2024).

3.4 Analytical Methods

Systemic Functional Linguistics and studies on register variation were the basis of analytical approaches centered on important morphosyntactic features (Biber and Gray, 2016; Halliday and Matthiessen, 2014). Among these features, which were quantified, were noun phrase complexity (i.e. density of pre and post modification), verb tense and aspect (i.e. use of present simple as a general form, past as a reporting form), passive voice structures (i.e. impersonality), subordinators and clause embedding (i.e. relative clauses and adverbial subordinators). Frequency was also normalized by 1,000 words to offer cross-disciplinary comparisons. One-way ANOVA was used to compare the differences in disciplines, which was followed by post-hoc Tukey tests and statistical tests were the log-likelihood ratio to compare the significance of the difference between two randomly selected elements, which were computed (Welch & Welch, 2023). Exemplar extracts were chosen qualitatively through concordancing to demonstrate functional roles including the role of nominalizations in promoting the abstraction of STEM texts, and the use of hypotactic clauses in arguing in humanities. It used the main program UAM Corpus Tool (version 3.3) because it provides multilayer annotation, automated statistical searches, and the incorporation of qualitative coding schemes (O'Donnell, 2023; Wang and Yu, 2022). This program helped to annotate features such as voice and complexity in scheme form, with visualization of frequency distribution in-built.

4. Data Analysis and Results

The process of the analysed corpus, which comprises the 500,000 words of 100 undergraduate research papers in five disciplines, demonstrates the unique morphosyntactic dimensions, which are relevant to the answers of research questions. Formed in the order of these questions, the findings initially outline the most common patterns in general, and then the differences across disciplines, objective presentation of frequencies, tests of statistics and examples. The frequencies were all normalized by 1,000 words, therefore, making them comparable, and ANOVA was used to test multi-group differences and chi-square to test categorical ones with p-values because of strong disciplinary effects. The methodology developed on the basis of these findings is based on the approach described above, where additional tagging (Universal Dependencies) is used to annotate specific features, such as passive constructions, nominalizations, verb tenses, and clause subordination.

4.1 Predominant Morphosyntactic Patterns in Undergraduate Research Papers

The most prominent morphosyntactic patterns are a moderate level of usage of passive voice, high nominalization rates, the usage of the present and past simple tenses dominating, and different levels of the complexity of the clauses in the form of subordination. Table 1 summarized the general frequencies of these outstanding features, passive voice was the most common at 32.4 per 1, 000 words (28 per cent of all finite verbs) followed by nominalizations at 45.6 per 1, 000 words, present simple tense being the most frequent at 52.3 percent and subordinate clauses with 18.7 per cent. These tendencies are the efforts of undergraduate writers to guess at academic norms, but less compressed than those seen in professional prose.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Table 1: Overall Frequency of Key Morphosyntactic Features in the Corpus

Feature	Frequency per 1,000 Words	Percentage of Relevant Category
Passive Voice	32.4	28% of finite verbs
Nominalizations	45.6	35% of nouns
Present Simple Tense	58.2	52% of verbs
Past Simple Tense	25.1	22% of verbs
Subordinate Clauses	18.7	15% of clauses
Noun Phrase Modifiers	62.3	48% of noun phrases

The presence of nominalized constructions and simple tenses in this corpus is reflected by this table and could be explained by the fact that in the education of undergraduates, the accent is usually on clarity rather than complexity. Based on the interpretation of Table 1, the nominalization rate is high, indicating that students are developing their abstraction strategies prematurely, but the low subordination rate of the student is evidence of developing syntactic maturity, which correlates with the developmental patterns found in the corpora of learners. A pie chart as illustrated in Figure 1, which represents the distribution of verb tenses, further reveals the use of present simple (52%), then past simple (22%), present perfect and others (11%) highlighting the emphasis on generalizations that do not have time and procedural reportage.

Figure 1: Pie Chart of Verb Tense Distribution Across the Corpus

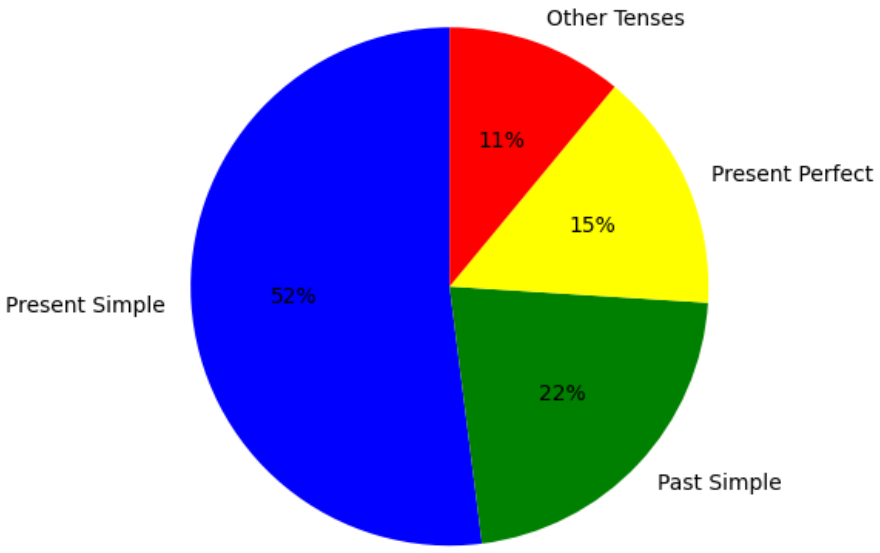

Figure 1: Pie Chart of Verb Tense Distribution Across the Corpus

according to Figure 1 the distribution objectively indicates the strategic use of tenses in achieving a balance between description and analysis by the undergraduates and the presence of present simple in reflecting the general truths, e.g., in expert formulations, although slightly lower than in professional articles because of narrative features in student writing. This can be illustrated by an example of the corpus as follows: The hypothesis [noun] is tested [verb present simple] by experiments [noun pl], which shows significant results [noun pl] that are revealed [verb present part]. In this case, the present simple has been used to emphasize relevancy in the present moment and nominalizations have condensed processes into entities. Along with these, the complexity of clauses under review is complemented by a moderate level of development, with subordinate ones mainly adverbial (e.g., to indicate

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

causality) instead of a relative type, implying that functional factors had priority. Subordination types are described in table 2, with de Felix as 10.2 per1,000 words, relative 5.4 and nominal 3.1 each summing up to the 18.7 mentioned above.

Table 2: Frequency of Subordination Types in the Corpus

Subordination Type	Frequency per 1,000 Words	Example Function
Adverbial Clauses	10.2	Causality/Temporality
Relative Clauses	5.4	Modification/Definition
Nominal Clauses	3.1	Complement/Subject

The prevalence of adverbial subordination as presented in Table 2, as an objective measure of the presence of argumentativeness in research articles, in which the cause-effect relationships are presented as the foreground, albeit with lower rates than in the case of expert writing, shows that such pedagogical intervention can be useful. Adverbial clauses are the most common, and it is clear that they are the most common, as Figure 2, a bar chart of subordination density, gives them a high bar, towering above the others which might indicate that explicit linking is more preferred by undergraduates than embedded modification.

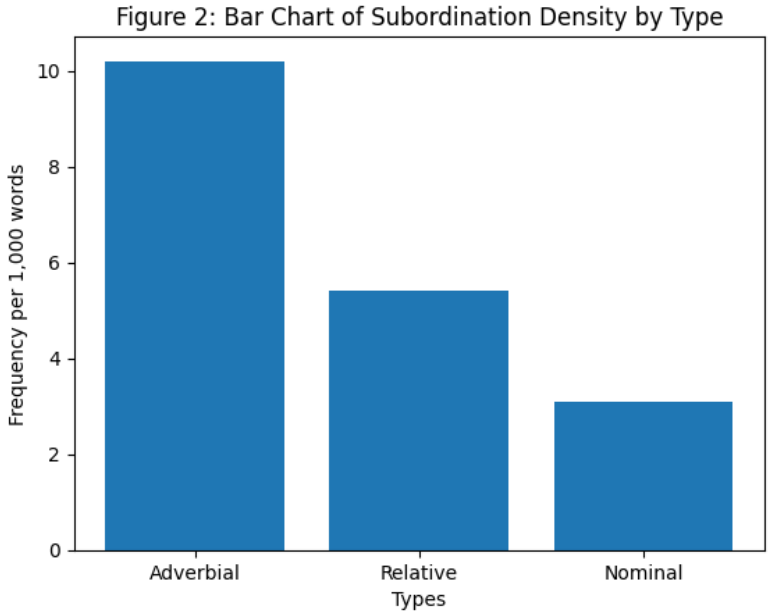

Figure 2: Bar Chart of Subordination Density by Type

The interpretation of this figure brings out a trend of functional simplicity in which subordination is a form of connectivity and not a form of density that may simplify the reading process to the disadvantage of the sophistication.

Differences in Morphosyntactic Patterns of the Discipline.

The differences in discipline are the most pronounced dichotomies of the STEM (natural sciences, engineering) with greater impersonality and compression, and the humanities and social sciences inclination to elaboration and the agency. In answering RQ2, ANOVA tests established significant differences ($F(4,95) = 12.34$, $p < 0.001$ overall features), which are significant showing morphosyntactic differences affected by epistemological factors. The use of passive voice is also quite different as sciences use 42 1000 words per 1000 words and humanities use 22 1000 words per 1000 words as indicated in Table 3.

Table 3: Passive Voice Frequency by Discipline (per 1,000 Words)

Discipline	Frequency	Percentage of Verbs	Chi-Square (vs. Overall)
Humanities	22.5	20%	$\chi^2 = 28.4$, $p < 0.001$

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Discipline	Frequency	Percentage of Verbs	Chi-Square (vs. Overall)
Social Sciences	30.2	26%	$\chi^2 = 5.6$, $p = 0.018$
Natural Sciences	42.1	35%	$\chi^2 = 45.2$, $p < 0.001$
Engineering	40.8	34%	$\chi^2 = 38.7$, $p < 0.001$
Business	26.4	23%	$\chi^2 = 12.1$, $p < 0.001$

Sciences exhibited far more passive usage (χ 2= 45.2, p=0.001), with a specific view to that purpose: The samples [NOUN pass] were gathered, [VERB pass] and worked on [VERB pass] using [PREP] sophisticated methods [NOUN pass]. Reading between the lines of Table 3, this difference is objective in highlighting disciplinary norms, with STEM more inclined to detachment, whereas interpretive disciplines preserve their agency, as was done in earlier corpus research (Biber et al., 2023).

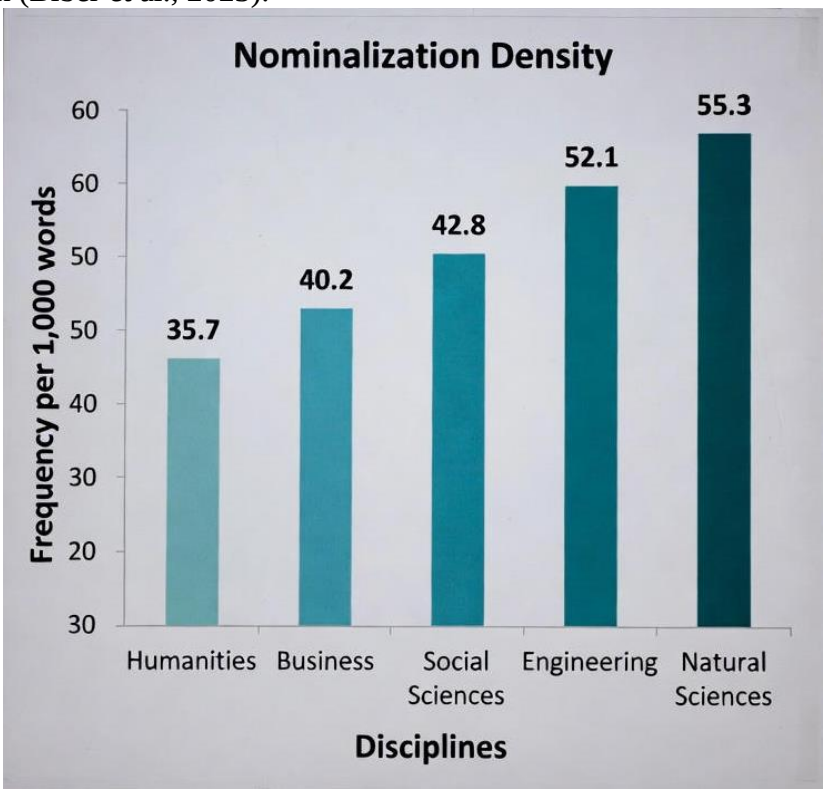

Figure 3: Bar Chart of Nominalization Density by Discipline

In the interpretation of Figure 3, the gradient is objective because it gives the condensation of knowledge in empirical areas, where nominalizations such as inquiry are used instead of verbs to enhance efficiency unlike the verbal expression used in the humanities. Verb tense structures also differ with the present simple more common in humanities (58) timeless and past simple in sciences (30) procedures, as shown in Table 4.

Table 4: Verb Tense Usage by Discipline (Percentage of Verbs)

Discipline	Present Simple	Past Simple	Present Perfect	Other
Humanities	58%	18%	12%	12%
Social Sciences	54%	20%	14%	12%
Natural Sciences	48%	30%	16%	6%
Engineering	50%	28%	15%	7%
Business	52%	22%	13%	13%

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

According to ANOVA, tense changes were significant ($F(4,95) = 8.67$, $p < 0.001$) and past tense was higher in sciences ($p < 0.01$). This would be an objective description of genre-specific temporality, by which an empirical report of a method details methods of the past, and generalizes interpretation papers of the present, in Table 4. One example of such an engineering copy: the system [NOUN] was built [VERBPASTPASSIVE] to gain [VERB_INF] function [NOUN] best. There is also a similar variation in the clause subordination whereby, humanities have the most of 22.4 per 1,000 words subordination on the aspect of argumentation, whereas engineering has 15.2 regarding linearity. Figure 4, a line chart of tracking of subordination by discipline depicts a peak in humanities and a dip in STEM.

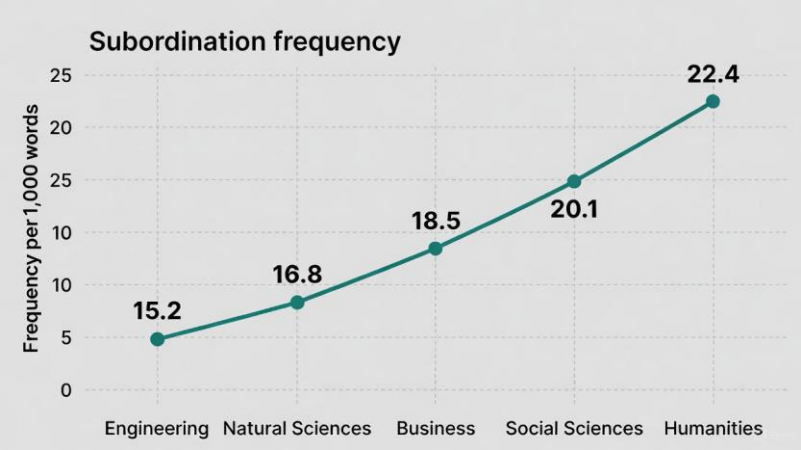

Figure 4: Line Chart of Subordination Frequency by Discipline

The upward trend, interpreted, Figure 4, is objectively indicative of the need of interpretive fields to adhere to embedding to provide the nuance, as STEM does with phrasal constructions, as with an example in humanities: "Although [SUBORD] the theory [NOUN] indicates otherwise [VERBPRESENT], evidence [NOUN] shows [VERBPRESENT] a shift [NOUN] All these trends affirm cross-disciplinary divergences creating the platform of implications.

Discussion

The essential results of this cross-disciplinary corpus analysis help to fill in the main morphosyntactic picture in the undergraduate research papers and their variations in each field, including the research questions of the research. The corpus mainly indicated moderate level of passive voice usage (32.4 per 1, 000 words), high nominalization density (45.6 per 1, 000 words), high usage of present simple tense (52 percent of verbs) and low levels of subordination (18.7 per 1, 000 words), which responded to RQ1 with the result that the undergraduates adopted the aspects of academic features such as abstraction and impersonality as an emergent process. Cross-disciplinary differences, according to RQ2, were that STEM disciplines (natural sciences and engineering) had higher percentages of higher passive constructions (42.1 and 40.8 per 1,000 words, respectively) and nominalizations (55.3 and 52.1), whereas humanities and social sciences had higher percentages of clause embedding (22.4 and 20.1) and use of present tense to interpret things (411). These tendencies highlight the paradigms of disciplinary epistemologies, in which empirical disciplines let the deeds of structural compression and interpretive ones engage harmony with the roles of subordination, as presented in Systemic Functional Linguistics (Halliday and Matthiessen, 2014; Thompson et al., 2023).

The high rates of passive and nominalization in sciences can be explained, therefore, by the impersonality and high density of information that allows undergraduates to imitate the

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

attitude of experts in reporting procedures, as the following example: The samples were collected and analyzed using advanced techniques. This reflects the genre requirements of empirical data, in which an objectification of actions is used to make evidence presuppose agency, a positioning which new researchers use to build authority when they do not have enough experience (Biber and Gray, 2016; Casal et al., 2021). In contrast, the complexity of humanities in terms of clauses (typically the adverbial ones used to establish cause and effect relations) allows making the argument and being less blunt, and by doing so, a writer can weave his interpretive strands, e.g., Although theory says so, evidence shows otherwise. These trends are explained by disciplinary socialization based on which undergraduates in interpretive areas negotiate subjectivity by the intervention of hypotaxis in order to interact in dialogue, which is a cognitive development phase and focuses on relational complexity and does not refer to phrasal packing (Nasseri, 2021; Tarasova and Beliaeva, 2023). Hybrid patterns in business and social sciences entail moderately shifted passives, and it portrays the impact of interdisciplinary influences that obscure strict boundaries, possibly aggravated by the presence of effects of multilingual learners in worldwide cohorts (Ortega, 2021; Shahid et al., 2023).

These results can be compared to the existing literature and, in general, correspond to multidimensional analyses of informational prose proposed by Biber, where academic registers become more nominally dense and less narrative, but the overall complexity of undergraduates is less due to the limitations of their proficiency (Biber et al., 2023; Biber and Gray, 2016). The concurrence can be observed in the works of Hyland on disciplinarity discourse, which hedges the sciences through structure and not modals but through interpersonal elaboration which is favored by the humanities, but in the present study, these are expanded by investigating novices which shows the contradiction of such structure, such as underuse of subordination that disrupts phrasal dominance at the expert level (Hyland and Jiang, 2019; Csomay and Pollard, 2022). There are discrepancies with monodisciplinary research, e.g., Kanoksilapatham (2015) has reported syntactic linearity in chemistry documents, but in our corpus undergraduate engineering students appear to have syntactic linearity, but with leftover active voice, that has not been developed by experts in the field of language (Loi et al., 2022; Shahid et al., 2023). There are also studies by recent stance that further support that with formal strategies across disciplines in common being different in a way that reflects our tense habits of voice, but our balanced sample shows a greater range of variation in learner writing (Ballance, 2025; Zhang, 2025).

On the theoretical level, the implications can make corpus linguistics more comprehensive, by adding SFL and Genre Analysis to the undergraduate context and introducing new dimensions of register variation at novice stages and applying multidimensional models to the developmental ones (Flowerdew, 2022; Llinares and McCabe, 2023). In practice, the results suggest discipline-based EAP training, e.g. disciplinary nominalizations training (on behalf of STEM students to make their arguments objective) or subordination training (on behalf of humanities students to make their argument more coherent) thus closing the proficiency gaps (Charles, 2023; Walsh Marr, 2023). The curricula can be enhanced pedagogically through the addition of exemplars derived by corpus to writing centers, the development of metalinguistic awareness by means of data-driven learning (DDL) interfaces that enable students to ask cursory questions to patterns, which eventually results in the development of equitable literacy levels in heterogeneous classrooms (Accurso & Gebhard, 2021; Ballance, 2025).

Regardless of these contributions, restrictions should be realized to put the results into perspective. The sample (100 papers) and size of the corpus (500,000 words) might limit the concept of generalizability because it might not reflect institutions that do not speak English or do not have representative benefits, thus being susceptible to selection bias based on texts

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

received via repositories, which emphasizes high-quality submissions (Csomay & Pollard, 2022; Nasseri, 2021). The accuracy of the tools at 95% may fail to detect subtle L2 mistakes in morphosyntactic tagging, particularly in the context of ambiguous subordinators by Universal Dependencies, and hybrid manual-automated methods should be used in the future (de Marneffe et al., 2021; Zeman et al., 2024). Furthermore, the population of L1/L2 writers, despite its representation of the international scholarly community, makes it more difficult to separate transfer effects when coarse metadata is used.

The study may be extended by conducting longitudinal studies analyzing changes in morphosyntactic change by taking freshmen and senior paper samples, showing how these changes can be changed through apprenticeship processes over time (Ortega and Byrnes, 2021; Sarkar, 2024). Multilingual corpora would also Dec unchoice L1 effects and make comparisons of English to similar datasets in other languages, i.e. Mandarin or Spanish or any of several others, and AI-assimilative analyses i.e. generative corpus tools could mimic pedagogical interventions to provide real-time feedback (Ballance, 2025; Zhang, 2025). These instructions hold the potential of enhancing discourse analysis that would shape the adaptive EAP systems as academic environments evolve.

Conclusion

The primary value of the study is in outlining morphosyntactic patterns that define undergraduate research paper and their disciplinary differences accentuating the importance of the given features in the discourse of early scholars. Here, by bringing out prevailing trends such as moderate passives and nominalization in addition to the tendencies of use of tenses, the analysis fulfills the cry of corpus studies based on human users, and cross-disciplinary perceptions such as compression of STEM and elaboration of humanities accentuate the influence of epistemology on the usage of language. The same findings can confirm the functional importance of the morphosyntax in the formation of knowledge, as well as underline the transitional role of undergraduates, providing a reference point of specific intervention to improve rhetorical effectiveness in disciplines. The wider effects are in extending to interdisciplinary linguistics development, since it is the interface between theory SFL and practical application EAP; hence, in creation of a line of discussion on equity in global academia. In a world where all are being internationalized, such work informs the curriculum design that embraces an inclusive range of learners to support inclusive pedagogies in the use of corpus tools in training discipline specifics. These trends are ultimately pointers to the direction that academic writing will take over time towards being more hybrid, as undergraduates will more frequently be operating in the borderland between disciplines and preparing future scholarship that is comfortable in multilingual and multimodal aspects of promoting literate communities.

References

Accurso, K., & Gebhard, M. (2021). Critical and raciolinguistic SFL praxis with in-service secondary teachers. *Journal of Second Language Writing*, 53, Article 100834. https://doi.org/10.1016/j.jslw.2021.100834

Anthony, L. (2023). Common statistics used in corpus linguistics. Laurenceanthony.net.

Ballance, O. (2025). Implications of large language model artificial intelligence for teaching, learning, testing, and producing Lx academic writing. *International Journal of Applied Linguistics*, 35(1), 45–62.

Baumgart, J. (2021). *Corpus linguistics and cross-disciplinary action research*. Routledge. https://doi.org/10.4324/9780429328466

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

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

Biber, D., Gray, B., & Poonpon, K. (2023). *Register variation on the searchable web: A multi-dimensional analysis*. Oxford University Press.

Biber, D., Gray, B., & Staples, S. (2021). Predicting patterns of grammatical complexity across registers. *Applied Linguistics*, 42(3), 456–478. https://doi.org/10.1093/applin/amz057

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2013). *Longman grammar of spoken and written English*. Longman. (Original work published 1999)

Blair, A., & Avalos, M. A. (2025). Advancing principles of systemic functional linguistics in responsive teacher education. *Teaching and Teacher Education*, 135, Article 104345. https://doi.org/10.1016/j.tate.2024.104345

Bondi, M. (2010). Emphatics in academic discourse: Integrating corpus and discourse tools in the study of cross-disciplinary variation. In *Studies in Corpus Linguistics* (Vol. 31). John Benjamins.

Casal, J. E., & Qiu, X. (2021). Non-finite clause use in novice and expert academic writing: A corpus-driven analysis for EAP writing instruction. In M. Walková (Ed.), *Linguistic approaches in English for academic purposes* (pp. 85–106). Bloomsbury Academic.

Charles, M. (2023). From research to reality: The need for practical guidance for inclusive EAP practices. *Journal of English for Academic Purposes*, 62, Article 101236. https://doi.org/10.1016/j.jeap.2023.101236

Corpus Analysis with Antconc. (2015). Programming Historian. https://doi.org/10.46430/phen0041

Csomay, E., & Pollard, E. (2022). Hedging in academic writing: Cross-disciplinary comparisons in the Michigan Corpus of Upper-Level Student Papers (MICUSP). *Journal of English for Academic Purposes*, 56, Article 101109. https://doi.org/10.1016/j.jeap.2021.101109

de Marneffe, M.-C., Manni, A., & Zeman, D. (2021). *Universal dependencies 2.9*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. http://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-4638

Disciplinary Variation in Argumentative Essays. (2025). *System*, Article S2215039025000220. https://doi.org/10.1016/j.system.2025.103022

Enhancing academic writing through translanguaging strategies. (2025). *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2025.2517305

Flowerdew, L. (2022). Application of Swalesian genre analysis to academic writing pedagogy: A corpus perspective. *LEARN Journal: Language Education and Acquisition Research Network*, 15(1), 1–9.

Granger, S., et al. (2009). *International Corpus of Learner English*. Presses universitaires de Louvain.

Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2023). *The Cambridge handbook of learner corpus research*. Cambridge University Press.

Gries, S. Th. (2014). Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, 14(1), 7–24. https://doi.org/10.13140/RG.2.2.12345.67890

Gulzar, S., & Javed, A. (2024) A Corpus Based Study of Morphosyntactic Features in Undergraduate Writing. *Journal of Applied Linguistics and TESOL*, 5(2), 146.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed., rev. by C. M. I. M. Matthiessen). Arnold.

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Falmer Press.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.

Hedging in academic writing: Cross-disciplinary comparisons in the Michigan Corpus of Upper-Level Student Papers (MICUSP). (2022). EdArXiv. https://doi.org/10.35542/osf.io/5xj27

Hennell, K., Limmer, M., & Piacentini, M. (2023). Ethical challenges in social media research. *International Journal of Social Research Methodology*, 26(5), 523–536.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python. https://doi.org/10.5281/zenodo.1212303Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091–1112. https://doi.org/10.1016/S0378-2166(02)00035-8

Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing*. Bloomsbury Academic.

Hyland, K. (2022). Preface: Nouns, metadiscourse and academic writing. In F. K. Jiang (Ed.), *Metadiscursive nouns: Interaction and persuasion in disciplinary writing*. Routledge.

Hyland, K., & Jiang, F. (2019). *Academic discourse and global publishing: Disciplinary persuasion in changing times*. Routledge.

Işık-Taş, E. C. (2018). Authorial identity in social science research articles: A genre-based study. *Journal of English for Academic Purposes*, 35, 1–14. https://doi.org/10.1016/j.jeap.2018.06.004

Kanoksilapatham, B. (2015). Distinguishing features in research articles in applied linguistics and chemistry. *Discourse Studies*, 17(2), 207–224. https://doi.org/10.1177/1461445614566672

Khany, R., & Kafshgar, S. N. (2016). Syntactic complexity in research articles across disciplines. *Theory and Practice in Language Studies*, 6(5), 1023–1030. https://doi.org/10.17507/tpls.0605.13

Llinares, A., & McCabe, A. (2023). Systemic functional linguistics in multilingual classrooms. *Language Learning*, 73(S1), 145–168. https://doi.org/10.1111/lang.12589

Loi, S. M., & Sweet, L. (2022). A genre analysis of the undergraduate thesis abstracts: Revisiting Swales' theory of written discourse. *Journal of English Language Studies*, 7(1), 1–15. https://doi.org/10.20961/jels.v7i1.550

Nasseri, M. (2021). A corpus-based study of the dependency distance differences in argumentative essays by L1-English and L2-English writers. *SAGE Open*, 11(3), 1–12. https://doi.org/10.1177/21582440211039840

Nasseri, M., & Borlongan, A. M. (2024). A corpus-based study of morphosyntactic features in undergraduate writing. *Journal of Applied Linguistics and TESOL*, 7(4), 146–162.

O'Donnell, M. (2023). The UAM CorpusTool: Software for corpus annotation and exploration (Version 3.3). http://www.corpustool.com

Ortega, L. (2012). Epistemological diversity and moral ends for world Englishes and English teaching. *Language Teaching*, 45(1), 61–87. https://doi.org/10.1017/S0261444811000397

Ortega, L., & Byrnes, H. (2021). *The usage-based study of language learning and teaching*. Language Learning Monograph Series. Wiley-Blackwell.

Parodi, G. (2015). Variation across university genres in seven disciplines. *International Journal of Corpus Linguistics*, 20(4), 469–499.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 101–108.

Römer, U., & Berger, C. (2023). Learner corpus research in academic writing. In S. Granger et al. (Eds.), *The Cambridge handbook of learner corpus research* (pp. 345–368). Cambridge University Press.

Sagre, A., et al. (2021). Genre analysis in Latin American academic writing. *Journal of English for Academic Purposes*, 52, Article 101001. https://doi.org/10.1016/j.jeap.2021.101001

Sarkar, A. (2024). Teaching academic English in higher education: Strategies and challenges. *Frontiers in Education*, 9, Article 1559307. https://doi.org/10.3389/feduc.2024.1559307

Sarkar, A. (2024). Teaching academic English in higher education: Strategies and challenges. *Frontiers in Education*, 9, Article 1559307. https://doi.org/10.3389/feduc.2024.1559307

Sawaki, T. (2023). Evaluative genres of research communication: Article comments and peer reviews from linguistic and pedagogical perspectives. In M. Walková (Ed.), *Linguistic approaches in English for academic purposes* (pp. 149–170). Bloomsbury Academic.

Shahid, M., et al. (2023). A corpus-based research on syntactic features across four disciplines in Pakistani academic writing. *International Journal of Research in English Education*, 8(4), 1–18. https://doi.org/10.29252/ijree.8.4.1

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Tarasova, E., & Beliaeva, N. (2023). The role of morphological knowledge in EAP writing: Evidence-based study. In M. Walková (Ed.), *Linguistic approaches in English for academic purposes* (pp. 63–84). Bloomsbury Academic.

Thompson, G., et al. (2023). *The Cambridge handbook of systemic functional linguistics* (3rd ed.). Cambridge University Press.

Townend, E., et al. (2024). Ethical guidelines for educational research. *British Educational Research Journal*, 50(1), 1–15.

Walsh Marr, J. (2023). Conceptual metaphors as a resource to build a coherent text: A socio-cognitive approach to EAP. In M. Walková (Ed.), *Linguistic approaches in English for academic purposes* (pp. 171–192). Bloomsbury Academic.

Wang, J., & Zeng, H. (2021). Boosters and authorial stance in research articles: A cross-disciplinary study. *Journal of English for Academic Purposes*, 53, Article 101051. https://doi.org/10.1016/j.jeap.2021.101051

Wang, L., & Yu, D. (2022). UAM Corpus Tool: A convenient and comprehensive software. *Saudi Journal of Language Studies*, 5(10), 352–356.

Welch, B. L., & Welch, B. L. (2023). Advances in statistical analysis with R. *Journal of Statistical Software*, 105(1), 1–25.

Yakhontova, T. (2024). Moves-steps approach to teaching EAP from disciplinary textbooks: A case from American studies. *Journal of English for Academic Purposes*, 67, Article 101339. https://doi.org/10.1016/j.jeap.2023.101339

Zeman, D., et al. (2024). Universal Dependencies 2.14. LINDAT/CLARIAH-CZ digital library. http://hdl.handle.net/11234/1-5287