LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 27, No. 1, April 2024, pp. 421-433

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

EXPLORING LEXICAL COMPLEXITY IN SCIENTIFIC WRITING: A CORPUS-BASED STUDY ACROSS DISCIPLINES

Mouna Ayadi
The Institute of Business Administration, University of Sfax, Tunisia
correspondence: mouna.ayadi@hotmail.fr
https://doi.org/10.24071/llt.v27i1.6790
received 6 July 2023; accepted 28 April 2024

Abstract
This study examines the lexical complexity of scientific writing. It specifically focuses on the degree of lexical density and lexical diversity in the results and discussion sections of research articles. Using a qualitative approach with quantification to back it up, this study scrutinized 60 results and discussion sections of research articles written by Tunisian linguistics and engineering scholars. The analysis was conducted using the lexical complexity analyzer developed by Lu (2010). Moreover, descriptive analysis and the independent T-test were conducted to ascertain the statistical distinctions between the discussion and results sections of linguistics and engineering disciplines. The findings suggest that engineering scholars wrote in a more concise and information-dense style, with a greater lexical density, while linguistics scholars used a broader range of linguistic forms, resulting in more diverse and richer expressions. The differences in writing style can be attributed to the nature of the disciplines and the types of research conducted within each field. The results obtained from this study may offer valuable implications for English for Academic Purposes (EAP) writing instructions.

Keywords: lexical complexity, lexical density, lexical diversity

Introduction
Scientific writing is critical for advancing knowledge and maintaining quality and credibility. It is an essential component of the scientific process, contributing to the expansion and dissemination of knowledge across disciplines. Furthermore, scientific writing enables researchers to communicate their findings to a broader audience. Gopen and Swan (1990) state that "the fundamental purpose of scientific discourse is not the mere presentation of information and thought, but rather its actual communication." (p.1). Hence, scientific writing has its own distinct and specialized way of describing processes and providing information and outcomes. According to Hyland (2000), texts reflect the construction and negotiation of knowledge in each discipline; thus, texts reveal characteristics projected by different disciplines. Halliday and Martin (1993) and Wignell (1998) found that the use of linguistic features varied across disciplines. In this respect, linguistic studies of research articles have primarily focused on genre analysis (Halliday & Martin, 1993; Swales, 1990). Researchers have studied

This work is licensed under CC BY-SA.
Creative Commons Attribution-ShareAlike 4.0 International License

various linguistic aspects of scientific texts, including lexical bundles (Adel & Erman, 2012), structural patterns (Lin & Evans, 2012), discourse moves (Upton & Connor, 2001), and interactive discourse structures (Camiciottoli, 2004). Studies on the linguistic features of research articles have yielded valuable insights into how academic writers engage in genre practices, as well as how academic text construction differs across academic disciplines (Hyland & Tse, 2007). The study of lexis played an important role in identifying variations in the hard and soft sciences. However, research on linguistic complexity in scientific writing has received little research interest to date.

Various aspects of linguistic complexity have been investigated from the standpoint of corpus linguistics. In terms of writing complexity, the literature has primarily concentrated on investigating syntactic complexity in scientific writing. Gray (2015) investigated linguistic complexity by focusing on disciplinary variation in the soft and hard sciences. The study focused on phrasal and clausal complexity in research articles across six disciplines: hard science, social sciences, and the humanities. Her study concluded that clausal complexity is more prevalent in the humanities and less relevant in the hard sciences. Furthermore, Gardner and Nesi (2019) concluded that clausal complexity is more prevalent in soft disciplines. Their study concluded that complexity was demonstrated by the use of epistemic adverbials and stance nouns.

While considerable attention has been focused on syntactic complexity in scientific writing, little is known about the use of lexical complexity in research article part-genres. Analyses of research article writing practices from phraseological and lexical perspectives have focused on examining them as a whole (Cunningham, 2017; Gilmore & Millar, 2018). Furthermore, there has been a significant focus on profiling the rhetorical construction of part-genres, revealing notable trends in disciplinary variation. Yet, there is a lack of understanding regarding the lexical complexity within the research article part-genre.

Previous studies were primarily concerned with the overall organization and movement analysis of discussion sections of articles written across a wide range of disciplines (Peacock, 2002; Rasmeenin, 2006). Recent studies have focused on the linguistic features and rhetorical strategies used in academic discussions to understand how authors interact with existing literature and present their arguments. For example, a study by Nizigama and Mahdavirad (2021) examined the use of hedging in the introduction and discussion sections of English research articles to determine how writers expressed certainty or uncertainty about their findings. This analysis illuminates how authors balance assertive claims with acknowledging research limitations. However, many researchers seem to overlook linguistic features like lexical complexity, which are among the most specified aspects of the results and discussion sections. Khany and Kafshgar (2016) note that the linguistic features of discussion sections, including lexical and syntactic complexity, often go unnoticed by researchers. This oversight is especially notable considering the significant implications of lexical choices and sentence structures on the overall quality of research findings.

Inspired by the claim that different scientific communities use linguistic features differently to express their ideas and knowledge (Halliday & Martin, 1993; Wignell, 1998), this study attempts to investigate linguistic variation in the

results and discussion sections of research articles across disciplines. Although previous research examined a wide range of linguistic features in a number of research articles from various disciplines, researchers rarely used lexical density and diversity to differentiate between soft and hard disciplines. It is still unclear how researchers implement these lexical complexity metrics in research articles across disciplines.

Lexical complexity is a component of the overall complexity parameters. The quantity and variety of vocabulary used in a sentence or text are referred to as its lexical complexity. Elements such as word repetition, word length, and the use of technical or specialized terms are taken into account. Dewi (2014) defines lexical complexity as "the features of language use found in the undergraduate students' research articles covering lexical density, lexical sophistication, and lexical variation" (p. 16). She states that lexical complexity is used to describe how writers communicate in written forms (Dewi, 2014, p. 2).

Multiple measures are used to provide a more comprehensive evaluation of lexical complexity. Two key measures, lexical diversity and density, are used to compare and contrast proficiency in scientific writing. These measures gauge the lexical complexity of texts. Within the aforementioned measures, lexical density refers to the percentage of lexical words in the text, i.e., nouns, verbs, adjectives, and adverbs (Laufer & Nation, 1995). On the other hand, lexical diversity represents the number of distinct words in any spoken or written text. Lexical diversity, also known as lexical variation or extent, is the range of a learner's vocabulary as demonstrated in his or her language use. The variety of lexical items in a text is measured using lexical diversity indices, which are thought to reflect the depth of the author's lexical knowledge (Kyle, 2020, p. 458).

Density can be used to convey information that is more detailed and precise. Johansson (2008, p. 65) argues, "By investigating this (lexical density), we receive a notion of information packaging; a text with a high proportion of content words contains more information than a text with a high proportion of function words (prepositions, interjections, pronouns, conjunctions, and count words)." The lexical density scale is used to rate the difficulty of a text. Denser texts are more difficult to comprehend. In addition, the denser the lexical density in the text, the more information it includes. According to Halliday (1985a, 1987), written language becomes complex by being lexically dense. In other words, complexity rises when lexical items outnumber grammatical words in a language. Lexical density is determined by the ratio of content words (nouns, verbs, adjectives, and adverbs) to the total number of words. (Lu et al., 2018, pp. 6–7). According to Halliday (1994), a text's information density is measured by its lexical density, which is "how tightly the lexical items have been packed into the grammatical structure" (p. 76). As a result, lexical density provides a measure of the information density in a text (Kyle, 2020, p. 457). Measuring lexical diversity involves various approaches. It is measured by the Type-Token Ratio (TTR) of each article, which describes the total number of unique words normalized by the length of the text. The Type-Token Ratio (TTR) calculates the percentage of different words in a text that make up the total number of words in that text. It is a useful tool for measuring language complexity. However, it is essential to combine it with other metrics to ensure a comprehensive evaluation of language proficiency. Examining TTR in conjunction with metrics like word frequency and

sentence structure offers a comprehensive insight into the nuances of lexical diversity.

The purpose of this research is to investigate the significance of lexical complexity in two core disciplines: linguistics and engineering. The analysis aims to identify and compare the approaches of researchers in linguistics and engineering towards lexical density and diversity in the results and discussion sections. As the most argumentative section of any article, the discussion section deserves scrutiny. Brett (1994) claims, "Initial, informal investigations of the corpus showed the Results section—conventionally the third section of the RA in an Introduction, Method, Results, Discussion structure—to possess characteristics that would merit further investigation and description. This is where writers articulate their new knowledge claims by presenting, explaining, and interpreting numerical data." (p.47). It is crucial to include the Results and Discussion sections in academic papers as they play a vital role in presenting and interpreting numerical data, thereby asserting new knowledge (Brett, 1994). This suggests that additional research and analysis of this section are required to fully comprehend its special features. Comparing the results and discussion sections of different disciplines is crucial for gaining insights into how each field presents and interprets its arguments. Conducting a comparative analysis could offer valuable insights into the distinct approaches adopted by writers in diverse academic disciplines.

The objective of this study is twofold: on the one hand, to investigate the complexity of the corpus analyzed using lexical density and, on the other, to determine diversity in the corpus using lexical variation. In fact, this research aims to explore lexical complexity measurements, namely lexical density and lexical variation, and evaluate them to answer the following research questions:

1. Are there any significant differences between the Linguistics and Engineering disciplines in terms of lexical density?
2. Are there any significant differences between the Linguistics and Engineering disciplines in terms of lexical diversity?

Method

This study uses the descriptive-quantitative research methodology founded on corpus linguistics to investigate the topic at hand. Khany and Kafshgar (2016) state that corpus linguistics "collects and stores authentic written and spoken data electronically, which is then used as the source of data for text analysis software in order to generate quantitative information about texts" (p. 283).

Quantitative analysis involves the examination of numerical data through statistical methods. This study employed descriptive statistics, mean, standard deviation, and the independent T-test for analysis. Quantification is employed to assess the lexical complexity of the results and discussions of research papers. The quantitative approach aims to offer a numerical representation of the lexical complexity of academic writing, facilitating a more objective analysis of the data. This method provides a comprehensive insight into how researchers utilize language to convey intricate ideas and findings in their research papers.

The corpus

This research focused on two disciplines: linguistics and engineering. The purpose of analyzing lexical complexity in the two disciplines is to gain insights into how language is used and structured in each field, offering a deeper understanding of the challenges and requirements of scientific writing in linguistics and engineering contexts. Comparing lexical complexity across disciplines enables the identification of patterns and trends that can enhance language teaching practices and communication strategies in specialized fields. These two disciplines were chosen based on the most practical method of categorizing disciplines into four major categories: sciences, social sciences, humanities/arts, and applied disciplines (Coffin et al., 2003). The terms 'hard' and 'soft', widely attributed to Storer (1967), are used to compare scientific fields based on perceived methodological rigor, exactitude, and objectivity. In summary, the applied, empirical, experimental, and natural sciences, such as astronomy, biology, mathematics, and physics, are labeled as 'hard', while the social sciences, including history, linguistics, literature, sociology, and political science, are termed 'soft' (cited in Pérez-Guerra & Smirnova, 2023, p. 157). This study comprises engineering articles in the hard-science sub corpus and linguistics research articles in the soft-science sub corpus.

The corpus for this study consisted of 60 academic research articles (30 in linguistics and 30 in engineering). The analysis concentrated on the results and discussion sections of research articles. An emerging line of research aims to explore writers' linguistic achievements by analyzing writers' changing use of complex structures across stages of research article (RA) writing. (Casal et al., 2021, p. 2). However, such research tends to emphasize RA introductions over other major part-genres, leaving questions about whether writers tend to produce complex structures over other major RA part-genres largely unanswered. Hence, this study focuses on the results and discussion sections of research articles to investigate lexical complexity and contribute to a more comprehensive understanding of linguistic features in academic writing beyond mere introductions. The data employed in this study were written by Tunisian scholars and downloaded from two Tunisian peer-reviewed journals dedicated to linguistics (TAYR) and engineering (IJETR). The two journals were chosen because they are well-known in their fields and publish high-quality research articles. Furthermore, focusing on Tunisian scholars allows for a more specific examination of academic writing practices in the country. It ensures a comprehensive examination of the language used in different fields of study within the Tunisian academic community.

Data analysis

The data analysis process consisted of three phases: The first phase involved counting lexical density. The second phase involved counting lexical variation in the corpus under scrutiny. In the third phase, the findings were consolidated into an Excel file for statistical analysis. Initially, a descriptive analysis utilizing means and standard deviation was conducted to compare the numerical differences in lexical measures across the two disciplines. Subsequently, an independent T-test was conducted to assess the statistical variances between the results and discussion sections of the linguistics and engineering disciplines.

Qualitative procedures were employed to categorize the lexical and functional words, along with ranking clauses in the texts. Quantification was employed to determine the lexical density and lexical variation indices.

An automation process using the Lexical Complexity Analyzer (LCA) was included to identify the types and tokens in the corpus. The L2 Lexical Complexity Analyzer (L2LCA) was developed by integrating findings from various studies on lexical complexity measures in English text. It covers the multiple facets of language use, including lexical complexity, lexical variation, and lexical density. The L2LCA is a comprehensive tool that takes into account a variety of lexical complexity factors in English text. By combining and analyzing the results of various studies, it offers a comprehensive analysis of language use, considering variables such as lexical density, sophistication, and variation. Researchers and educators interested in understanding and improving lexical complexity in second language acquisition can benefit from this software's multidimensional approach.

The researcher chose this software for its availability and the type of results it provides for the analysis and comparison of corpora such as the ones examined in this study. This software takes an English text as input and produces 25 lexical complexity indices. The researcher inputs the text into a web-based form. The Stanford POS tagger tags the text to identify parts of speech, and Morpha lemmatizes the text for analysis. Then, the lemmatized results are fed into the Lexical Complexity Analyzer, which produces 25 indices (Ai & Lu, 2010). The Lexical Complexity Analyzer is a dependable tool that determines text complexity by analyzing linguistic features like word length, sentence length, and vocabulary diversity. Lu (2010) has verified the reliability and validity of L2SCA, the Lexical Complexity Analyzer. According to Lu (2010), the system achieves a very high degree of reliability for identifying units and structures, with F-scores ranging from 0.846 to 1.000 for different types of data. Its application in this study will provide valuable insights into the lexical complexity of the corpora analyzed and allow for meaningful comparisons.

Findings and Discussion

Findings

This study examined the differences in lexical complexity in the results and discussion sections in two disciplines, namely linguistics and engineering. After downloading the results and discussion sections and counting lexical density and lexical diversity, all the data were fed into SPSS. The coding was followed by a quantitative statistical analysis to explore the lexical complexity differences between the two disciplines under scrutiny. The following results were obtained: The final dataset for the linguistics discipline contained 48748 words (n = 30, mean = 1624), while the engineering discipline had 48658 words (n = 30, mean = 1621). Tale 1 shows the results from the descriptive statistics for the overall performance of the mean, minimum, and maximum values.

Table 1. The corpus descriptive statistics

	N	Minimum	Maximum	Sum	Mean
Word count Linguistics	30	58.00	5694.00	48748.00	1624.933
Word count Engineering	30	184.00	3107.00	48658.00	1621.933

	N	Minimum	Maximum	Sum	Mean
Sentence count Linguistics	30	3.00	233.00	2109.00	70.3000
Sentence count Engineering	30	9.00	196.00	2482.00	82.7333
Valid N (listwise)	30				

Table 1 shows that the results and discussion sections in the linguistics discipline contained more words than those in the engineering discipline. However, the engineering discipline produced more sentences than the linguistics discipline.

As depicted in Table 2, the results of the lexical density of 60 linguistics and engineering results and discussion sections were shown. Accordingly, using Ure's formula, the table below shows that the lexical index of the two disciplines ranged from the minimum of 48.28% for the linguistics discipline to the highest of 62.81% and from the minimum of 50.38% for the engineering discipline to the highest of 63.59%. As a result, the total average of linguistics and engineering disciplines is 55.48% and 56.23%, respectively, which is above the average of 40%, as proposed by Ure (1971). Based on these results, it is evident that the results and discussion sections in the engineering discipline's research articles were more complex than those in linguistics. Therefore, to address the first research question, it is notable that the engineering discipline generally exhibits a higher average compared to the linguistics discipline.

Table 2. Lexical density means

	LD Linguistics	LD Engineering
Mean	55.48	56.23
N	30	30
Std. Deviation	3.30	2.82
Minimum	48.28	50.38
Maximum	62.81	63.59

An independent sample T-test with a significance level of 0.05 was used to investigate this difference further. This test sought to examine the differences between the two disciplines. According to the results reported in Table 3, the variance between the two groups is not significantly different ($p = .738 > 0.05$). Thus, one can assume that the two groups are equal. A t-test was conducted to compare the means of the two groups. The p-value from the test was 0.347, indicating that the difference between the means is not statistically significant at the alpha level of 0.05. Hence, $t(58) = -0.949$, $p = 0.347$.

Table 3. Independent samples test of lexical density

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
LD	Equal variances assumed	.113	.738	-.949	58	.347	-.75333	.79372	-2.34215	.83548
	Equal variances not assumed			-.949	56.634	.347	-.75333	.79372	-2.34296	.83629

In order to answer the second research question, the same procedures were used to determine the differences in means between the two disciplines. Table 4 showed that the mean Type-Token Ratio (TTR) in the linguistics disciplines was significantly higher (M = 3417; SD = 11) compared to the engineering discipline (M = 0.3043; SD = 0.98). In this vein, we can assume that researchers in the linguistics discipline employed more diverse linguistic forms than those in the engineering discipline. These results indicate a significant disparity in the utilization of diverse linguistic forms by researchers from the two disciplines to communicate their research inquiries and goals. This variation likely arises from the distinct research inquiries and methodologies employed within each discipline.

Table 4. Lexical diversity means

	TTR Linguistics	**TTR Engineering**
Mean	.3417	.3043
N	30	30
Std. Deviation	.11483	.09839
Minimum	.15	.18
Maximum	.69	.57

According to the results of the independent T-test reported in Table 5, the variance between the two groups is significantly different. The test indicated that the difference in means is statistically significant (t (58) = 2.258, p = 0.028). This indicates a meaningful difference between the two groups compared, suggesting that this difference is unlikely to be due to chance. The following table represents the statistics of the independent t-test.

Table 5. Independent samples test of lexical diversity

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
TTR	Equal variances assumed	3.147	.081	2.258	58	.028	.05633	.02495	.00640	.10627
	Equal variances not assumed			2.258	49.556	.028	.05633	.02495	.00622	.10645

Following the above analyses, one can conclude that both disciplines are statistically different. The engineering researchers wrote slightly denser texts than the linguistics researchers did. However, the latter used more diverse linguistic forms (i.e., higher means of TTR) than the engineering researchers did.

Discussion

The results of descriptive statistics indicated that researchers in the linguistics discipline wrote longer results and discussion sections than engineering researchers. This difference in length may be attributed to the nature of the research methodologies used in each discipline. In this regard, Becher and Trowler (2001) contended that the distinction between hard and soft science is supported by the various methodological and conceptual frameworks used in these disciplines. For example, the hard sciences frequently rely on quantitative data

and experimental methods, whereas the soft sciences may prioritize qualitative data and theoretical analysis. This diversity in approaches contributes to the varying lengths of research articles across disciplines. Furthermore, the complexities of language analysis may necessitate more detailed explanations and interpretations in linguistics research. Shahrokhi et al. (2013) highlighted that disciplines like sociology, psychology, and linguistics prioritize explicit interpretation over other fields, showcasing the need for in-depth analysis in these areas (p. 202). This suggests that linguistics research may require a higher level of analysis and interpretation than other disciplines.

Lexical density scored higher in the results and discussion sections written by engineering scholars. Researchers in the engineering disciplines wrote more complex and lexically dense discussion sections compared to those in the linguistics disciplines. The findings coincide largely with Jalilifar et al.'s (2017) study, which argued that academic researchers in the physics discipline tended to write in a more complex and lexically dense style. This tendency is likely due to the technical nature of their subject matter and the necessity for precision in their language use. According to Jalilifar et al. (2017), compared to writers in applied linguistics, academic writers in the field of physics have a tendency to (a) employ a more intricate and lexically dense writing style and package more information into compound nominal phrases by using a pattern where nominals are followed by strings of prepositional phrases (p. 1). The writing style of physics scholars may be influenced by the necessity for clarity and accuracy in conveying technical information, as suggested by these findings. Furthermore, contrasting with scholars in applied linguistics underscores the distinctive features of academic writing in physics. Academic writing in physics is typically more formulaic and technical, emphasizing precise data and analysis. In contrast, applied linguistics research often adopts a more theoretical and qualitative approach, unlike the formulaic and technical nature of academic writing in physics. The findings emphasize the significance of understanding disciplinary writing conventions and adapting one's style accordingly.

Contrary to the findings of this study, Gholami et al. (2012), investigating lexical complexity in the abstracts of ESP articles in soft and hard sciences, revealed that higher lexical density percentages were detected for psychology compared to engineering disciplines and tourism. They found that psychology had the highest lexical density, while the travel and tourism industry had the lowest (Gholami et al., 2012, p. 372). This variability indicates that lexical complexity differs among various fields in ESP.

The linguistics researchers used more diverse linguistic forms in their results and discussion sections. This is because linguistics, as a discipline, often promotes the use of diverse language forms to accurately convey complex linguistic concepts and theories. Therefore, writing styles may vary across academic disciplines based on their specific demands and expectations. For example, scientific writing is typically more objective and concise, whereas humanities writing may be more subjective and exploratory. In linguistics, using a variety of language forms enables researchers to effectively communicate their findings and interpretations to a larger audience. This diversity of writing styles reflects the interdisciplinary nature of linguistics, which draws on a variety of methodologies and theories.

Khany and Kafshgan (2016) discovered higher lexical complexity indices in humanities articles, which they interpreted as indicating a higher level of sophistication in soft science texts. Khany and Kafshgan (2016) revealed higher significant indices of lexical diversity in the soft sciences (humanities). They contended that their findings demonstrated that argumentation in humanities articles requires the use of a wider range of words (types). However, Khany and Kafshgan (2016) indicated higher significant indices for the humanities in lexical density and the ratio of subordinate structures compared with the other disciplines. The demonstrated differences attributed to the inherent nature of the disciplines can be seen in the study by Khany and Kafshgan (2016, p. 290). The finding suggests that authors who write in the humanities prefer to use words that have a specific meaning or content rather than more general words or filler words. The study's results emphasize the importance of considering the disciplinary context when examining linguistic complexity metrics in academic writing.

Conclusion

This study sought to identify the differences between the hard and soft sciences by comparing lexical complexity measures across the two fields. Overall, there was a statistically significant difference in the results of the lexical density in favor of the engineering scholars and the results of the lexical diversity, indicating a more diverse range of linguistic forms for the linguistics scholars. In summary, engineering scholars tended to write more complex and dense results and discussion sections, whereas linguistics scholars utilized a wider range of linguistic forms. These findings indicate that the hard sciences prioritize precision and technicality in their writing, while the soft sciences prioritize diversity and flexibility in language use. This highlights the distinct writing styles and priorities in each field, underscoring their contribution to the variations in lexical complexity measures.

Like any other study, this research has its own limitations. It only considers the differences in lexical complexity in the soft and hard sciences and ignores the differences in syntactic complexity in both disciplines. This hinders a comprehensive analysis of the characteristics of scientific writing, particularly in terms of structural intricacies and nuanced expression, limiting the depth of our insights. Therefore, future research could delve into the syntactic intricacies of both disciplines to enhance the elucidation of argumentation in the results and discussion sections of research articles. It would also be beneficial for future studies to consider the impact of language proficiency on syntactic complexity in scientific writing. This additional factor could provide a more holistic understanding of the nuances present in research articles written by non-native speakers.

Future investigations may require insights from other corpora, especially native scholars, or comparing corpora from ethnically different researchers to achieve a comprehensive understanding of the complexity of scientific writing. However, quantification is insufficient for understanding the characteristics of scientific writing style. Thus, future research could include a qualitative analysis of the various strategies used in scientific writing. For instance, this could involve analyzing the rhetorical use of metaphor and analogy, as well as examining how scientists establish their authority and credibility through citation practices and

language choices. Further studies could look into the impact of disciplinary and cultural differences on scientific writing styles.

The realized differences in the deployment of lexical complexity in the hard and soft sciences can be pedagogically inspiring. The findings of this study imply that ESP and EAP teachers should consider their students' major fields of study and the potential writing conventions of those disciplines when implementing classroom practices.

Acknowledgments

Special thanks to Prof. Mounir Triki for his valuable help in proofreading the manuscript.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81–92. https://doi.org/10.1016/j.esp.2011.08.004

Ai, H., & Lu, X. (2010, June). A web-based System for automatic measurement of lexical complexity. *27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10), Amherst* (pp. 8-12). http://dx.doi.org/10.13140/RG.2.2.16499.07208

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Buckingham and Philadelphia: The Society for Research into Higher Education and Open Up.

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes*, *13*(1), 47–59. https://doi.org/10.1016/0889-4906(94)90024-8

Camiciottoli, B. C. (2004). Interactive discourse structuring in L2 guest lectures: Some insights from a comparative corpus-based study. *Journal of English for Academic Purposes*, *3*(1), 39–54. https://doi.org/10.1016/s1475-1585(03)00044-4

Casal, J. E., Lu, X., Qiu, X., Wang, Y., & Zhang, G. (2021). Syntactic complexity across academic research article part-genres: A cross-disciplinary perspective. *Journal of English for Academic Purposes*, *52*, 100996. https://doi.org/10.1016/j.jeap.2021.100996

Coffin, C., Curry, M., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2003). *Teaching academic writing: A tool kit for higher education*. London, UK: Routledge.

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, *25*, 71–83. https://doi.org/10.1016/j.jeap.2016.11.005

Dewi, R. (2017). Lexical and syntactic complexities in undergraduate students' research articles and their correlations to their quality. *Jurnal Pendidikan Humaniora*, *5*(3), 123–133. https://doi.org/10.17977/um030v5i32017p12

Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, *40*(4), 646–674. https://doi.org/10.1093/applin/amy005

Gholami, J., Mosalli, Z., & Bidel Nikou, S. (2012). Lexical complexity and discourse markers in soft and hard science articles. *World Applied Sciences Journal*, *17*(3), 368–374.

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, *51*, 1–17. https://doi.org/10.1016/j.esp.2018.02.002

Gopen, G. D., & Swan, J. A. (1990). The science of scientific writing. *American Scientist*, 1–16.

Gray, B. (2015). On the complexity of academic writing: Disciplinary variation and structural complexity. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 49–78). Amsterdam: John Benjamins.

Halliday, M. A. K. (1987). Spoken and written modes of meanings. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 55–82). San Diego, CA: Academic Press.

Halliday, M. A. K. (1993). Some grammatical problems in scientific English. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 69–85). London, UK: Falmer.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London, UK: Edward Arnold.

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

Halliday, M. K. (1985). *An introduction to functional grammar*. London, UK: Edward Arnold.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London, UK: Longman.

Hyland, K., & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL Quarterly*, *41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Jalilifar, A., White, P., & Malekizadeh, N. (2017). Exploring nominalization in scientific textbooks: A cross-disciplinary study of hard and soft sciences. *International Journal of English Studies*, *17*(2), 1-20. https://doi.org/10.6018/ijes/2017/2/272781

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers*, *53*, 61–79.

Khany, R., & Kafshgar, N. B. (2016). Analyzing texts through their linguistic properties: A cross-disciplinary study. *Journal of Quantitative Linguistics*, *23*(3), 278–294. https://doi.org/10.1080/09296174.2016.1169848

Kyle, K. (2020). Measuring lexical richness. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 454–458). New York, NY: Routledge.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.26686/wgtn.12560396

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, *31*(3), 150–160. https://doi.org/10.1016/j.esp.2011.10.002

Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V., Schnaars, M., & Zhang, C. (2018). Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology*, *70*(5), 462–475. https://doi.org/10.1002/asi.24126

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.15.4.02l

Nizigama, E., & Mahdavirad, F. (2021). Hedging and boosting in the introduction and discussion sections of english research articles: A cross-cultural study of papers written by native and non-native academics. *Iranian Journal of English for Academic Purposes*, *10*(1), 108–123. https://dorl.net/dor/20.1001.1.24763187.2021.10.1.7.1

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, *30*(4), 479–497. http://dx.doi.org/10.1016/S0346-251X(02)00050-7

Pérez-Guerra, J., & A. Smirnova, E. (2023). How complex is professional academic writing? A corpus-based analysis of research articles in "hard" and "soft" disciplines. *Vigo International Journal of Applied Linguistics*, *0*(20), 149–183. https://doi.org/10.35869/vial.v0i20.4357

Rasmeenin, C. (2006). *A structural move analysis of MA thesis discussion sections in applied linguistics* (Doctoral dissertation). Mahidol University.

Shahrokhi, M., Sadeghi, S., & Dehnoo, M. A. (2013). lexical cohesion patterns in research articles: Hard science vs. soft science disciplines. *International Journal of Social Science & Education*, *4*(1), 196–204.

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

Upton, T. A., & Connor, U. (2001). Using computerized corpus analysis to investigate the text-linguistic discourse moves of a genre. *English for Specific Purposes*, *20*(4), 313–329. https://doi.org/10.1016/s0889-4906(00)00022-3

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics: Selected papers of the second international congress of applied linguistics* (pp. 443–452). Cambridge: Cambridge University Press.

Wignell, P. (1998). Technicality and abstraction in social science. In J. R. Martin & R. Veel (Eds.), *Reading science. critical and functional perspectives on discourses of science* (pp. 297–326). London, UK: Routledge.