A COMPARATIVE GENRE ANALYSIS STUDY OF SCIENTIFIC ARTICLES ABSTRACTS AND AI-GENERATED ABSTRACTS

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Arts
In the Department of Linguistics
University of Saskatchewan
Saskatoon

By

ABDELRAHMAN ALQUDAH

Permission to Use

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Linguistics,
Arts Building, 9 Campus Drive,
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A5 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

This study investigated and compared the rhetorical moves and steps employed in human-written abstracts of published articles with those generated by ChatGPT, 3.5 free version. The grammatical features and move patterns are also analyzed. The data for this study was collected by compiling corpus of 25 original research papers' abstracts and 25 ChatGPT-generated abstracts. The ChatGPT abstracts were based on the titles of the 25 original abstracts collected prior, which were all published in Q1 journals from 12 different disciplines. Those disciplines are Psychology, Economics, Biology, Physics, Geology, Artificial Intelligence, Linguistics, Sociology, Agriculture, Mechanical Engineering, Management, and Sports Medicine. UAM CorpusTool (version 2.8), a corpus annotation tool, was used to annotate the texts based on a scheme developed by the researcher. Adapted from the IMRD (Introduction-Methods-Results-Discussion model for abstract writing, this scheme contains four moves, namely, Introduction-Methods-Results-Conclusion or IMRC, and a total of 21 steps. This adaptation reflected the rhetorical structures identified in the corpus. Four moves (Introduction, Methodology, Results, and Conclusion) were found to be employed in the abstracts written by human authors and those generated by ChatGPT. It was found that the Results move is the most frequent move in ChatGPT abstracts, while the Introduction move is the most frequent one in original abstracts. Significant differences were found in the frequency of the Introduction and Conclusion moves, and key literature findings step between the ChatGPT-generated abstracts and original abstracts. Such differences are likely due to the nature of the training data used for the AI model. As for the linguistic features, a high tendency was found in both ChatGPT and original abstracts to use past tense and simple present tense, content words, and active voice. However, significant differences found between ChatGPT and original abstracts indicated that simple present was employed much more in ChatGPT abstracts than in original abstracts, while the opposite was the case for simple past tense. Significant differences in the use of pronouns also showed that this part of speech is used significantly more in original abstracts. The move sequence patterns found showed that the patterns of ChatGPT is similar to that of typical research publications, while human abstracts display a variety of move sequence patterns. The findings of the study show that ChatGPT abstracts mimic original abstracts in terms of the rhetorical move pattern and in most of the steps used in each move. The researcher concludes the study with recommendations for future researchers based on the findings of the current study.

Acknowledgements

Praise be to the Almighty one for giving me the strength and patience to finish this thesis. This thesis was finished in a social context and was not an individual experience. For that I sincerely thank everyone who helped me with this.

To my supervisor Li Zhi, words cannot describe how grateful I am for the guidance and help you provided throughout my study. When I had any issues with anything you were always there. I will be eternally grateful for the opportunities and assistance that you were more than happy to give.

To my committee, Professor Spreng and Professor Makarova, who helped me make this thesis presentable with their rich contributions. I have both enjoyed being a student in both of your courses and hope to make you proud with this final version of my thesis.

And finally, my family that I miss dearly every day. My mother that raised me alone, my sisters that I played with every day, and my life partner that I will always have everlasting true love to. I will always be yours.

TABLE OF CONTENTS

PERMISSION TO USE I
ABSTRACT II
ACKNOWLEDGEMENTS IV
TABLE OF CONTENTS V
LIST OF TABLES VII
LIST OF FIGURES VIII
1. INTRODUCTION 1
  1.1 Background 1
    1.1.1 ChatGPT 1
    1.1.2 Genre Analysis 3
  1.2 Gap and Aims 4
  1.3 Significance of The Study 5
  1.4 Hypothesis 5
2. LITERATURE REVIEW 7
  2.1 Overview 7
  2.2 Abstract Genre 7
  2.3 Genre Analysis 8
    2.2.1 Swales' CARS Model 10
    2.2.2 Bhatia's Move Analysis Framework 12
    2.2.3 Hyland's Five-Move Model for Abstracts 13
  2.3 Genre Analysis Studies 14
    2.3.1 Genre Analysis of Academic Writing 14
    2.3.2 Genre Analysis of Abstracts 16
  2.4 Linguistic Features of Abstracts 18
  2.5 AI-Generated Content 20
  2.6 Synthesis of The Literature 23
3. METHODOLOGY 25
  3.1 Overview 25
  3.2 Materials 25

3.3 Analytical Procedures 27
4. RESULTS AND DISCUSSION 37
4.1 General Texts Statistics 37
4.2 Genre Analysis 39
4.2.1 The Introduction Move 42
4.2.2 The Methodology Move 46
4.2.3 The Results Move 49
4.2.4 The Conclusion Move 53
4.3 Grammatical Features 56
4.4 Pattern Analysis 65
5. CONCLUSION 68
5.1 Summary of Findings 68
5.2 Limitations 72
5.3 Recommendations 72
APPENDIX A 74
APPENDIX B 75
REFERENCES 78

LIST OF TABLES

Table 1. Corpus Description 26
Table 2. Examples on the steps shown in the scheme from the study's corpus 31
Table 3. Comparative analysis of the general text statistics of ChatGPT and original abstracts 38
Table 4. Comparative analysis of the moves of ChatGPT abstracts and original abstracts 40
Table 5. Comparison of the Introduction steps in ChatGPT abstracts and original abstracts 43
Table 6. Comparison of the Methodology steps in ChatGPT abstracts and original abstracts 46
Table 7. Comparison of the Results steps in ChatGPT abstracts and original abstracts 49
Table 8. Comparison of the Conclusion steps in ChatGPT abstracts and original abstracts 54
Table 9. Comparative analysis of the grammatical features of ChatGPT and original abstracts 57
Table 10. Most frequent nouns, adverbs, adjectives, and pronouns in ChatGPT and original abstracts 61
Table 11. Pattern analysis of moves in original and ChatGPT abstracts. 65

LIST OF FIGURES

Figure 1. Scheme for annotating human-written and ChatGPT-generated abstracts 29

1. Introduction

1.1 Background

This chapter provides a background on the subject and acts as an introduction to the thesis. It states the significance of the study, examines its limitations, identifies the research gaps in the field, provides the research questions to be addressed. Furthermore, it offers a summary of the present study's structure and organization. By addressing these topics, this chapter clarifies the context and goals of the study for the readers as well as preparing them for the subsequent sections of the thesis. Overall, the current thesis is constructed to have five chapters. The current chapter, Chapter one, is an introduction to the study. Chapter two will present a thorough review of the related literature. Moreover, Chapter three will present and explain the methodology of data collection and analysis. The findings and the discussion of them will be presented in Chapter four. Finally, Chapter five will conclude the study with a summary and some recommendations for future researchers.

1.1.1 ChatGPT

According to Copeland (2023, Artificial Intelligence section, para. 1), the term 'AI' is commonly used to refer to the capability of creating artificial intelligence systems that perform human-like cognitive functions, for example, the ability to find meanings, reason, generalize, and learn from previous experiences. Copeland mentioned that, with the development of digital computers in the 1940s, computers can perform extremely complicated tasks with high proficiency, such as playing chess or finding proofs for mathematical theorems. However, Copeland argues despite constant upgrades in computer memory and processing power, there is yet no program that can fully match the human ability to adapt over a wider range of fields or in tasks requiring a great deal of everyday knowledge.

AI has revolutionized many industries, including education. AI is able to automate administrative tasks and provide personalized learning practices. One area where AI has shown potential in education is personalized learning, which is defined as "instruction that is paced to learning needs, tailored to learning preferences, and tailored to the specific interests of different learners" (U.S. Department of Education [USDE], 2019, p. 12). AI algorithms have the potential to analyze student information, for example their style of learning, pace, and preferences, which provides tailored learning experiences. According to the above definition, these details (i.e., style, pace, and preferences) are used to personalize the learning experience of students. This can be effective in the improvement of retention, engagement, and attainment (Alhadlaq et al., 2021).

Another employment of AI in the process of education is the utilization of chatbots. AI-generated texts are becoming increasingly common in education. For instance, AI-powered writing tools can assist students in developing their writing skills by suggesting improvements in grammar, syntax, and style. Chatbots can provide students with instant feedback and support, such as answering questions, providing study resources, and even grading assignments (Chu, 2021). These tools can also provide feedback on the readability and coherence of the text (Cohen & Ellis, 2021). Some studies investigated the use of ChatGPT to grade and provide feedback on students' academic writing (Makarova et al., 2024).

OpenAI is a leading company in the field of AI that was established in 2015 (Fitria, 2015). Its main interest since then has been developing generative language models, which are used to create novel content (Kalyan, 2023). OpenAI developed Generative Pre-trained

Transformer (GPT), which is its own generative language model. GPT model relies on pre-training on large language models to produce human-like content. Its first version (GPT-1) was launched in 2018 and after achieving great success in generating human-like content, new versions were developed (Saka et al., 2023). Those versions are GPT-2, GPT-3, ChatGPT (GPT-3.5), and GPT-4 which is the latest version launched in March 2023. ChatGPT is basically a chatbot that utilizes generative AI to provide detailed and accurate answers even to very complicated queries (Lund & Wang, 2023). ChatGPT was recently introduced by OpenAI in November 2022, and since its launch, it has become the center of interest for many learners and educators, not to mention other industries and fields. According to Lund and Wang (2023), this worldwide interest is a result of the high abilities of ChatGPT to write sophisticated paragraphs, or even whole articles, in any field with a nearly perfect language, at least in English, and with non-plagiarized content. Roumeliotis and Tselikas (2023) stated that ChatGPT is programmed to be able to perform different tasks. For example, it can answer questions, recognize language, and complete paragraphs. Moreover, they mentioned its ability to create chatbots and other conversational interfaces.

ChatGPT focuses on producing and understanding language, allowing it to engage in different conversations and deal with various inputs. It can generate texts in numerous genres, including academic, literary, and scientific ones. It is worth mentioning that these texts are well-constructed and bear a strong resemblance to human writing (Roumeliotis & Tselikas, 2023).

1.1.2 Genre Analysis

In applied linguistics, the term 'Genre' is defined as the utilization of language that adheres to specific established patterns and conventions in accordance with the norms of a particular group engaged in communication (Widdowson, 2007, p. 129). Based on this working definition, each genre may have its own specifications and styles. This led researchers to propose 'Genre analysis', which is a concept that pertains to a field within the study of discourse analysis, which examines particular instances and applications of language (Hyland, 2004, p. 195). Based on Mirador's (2000, p. 47) work, a 'move' can be defined as a notion used to describe the strategic use of words or expressions by individuals engaging in written or spoken communication. According to Swales (2004, p. 228), a 'move' is a textual unit which provides the text's communicative aims as well as maintaining coherence in the text's genre. Moreover, Mirador (2000) suggested that each move can be achieved through minor steps.

1.2 Gap and Aims

Although the problem of finding methods to identify ChatGPT-generated texts, especially in academic settings, is in the center of interest of academic institutions, few studies on the genre and style of ChatGPT-generated academic texts have been conducted. Thus, the current study aims to fill these gaps through addressing the following questions:

1. How do the human-authored abstracts differ from the ones generated by ChatGPT in terms of the rhetorical moves and steps?
2. What are the linguistic features, namely part of speech, tenses, and voices, used to achieve the moves in human-authored scientific article abstracts and ChatGPT-generated abstracts?
3. To what extent do the abstracts generated by ChatGPT mimic the genre and linguistic features of scientific article abstracts?

1.3 Significance of The Study

The study's findings will enhance understanding of ChatGPT's text generation and its credibility in academic contexts. It will shed light on the quality and suitability of AI-generated content in academic settings. This is very important for educators, researchers and institutions that benefit from such tools. Moreover, analyzing the moves and linguistic features of AI-generated texts offers insights into whether or not this content conforms to existing conventions or academic writing which help educators in integrating AI writing tools appropriately into the learning process and assist students to develop their writing skills.

In addition, it may help researchers to identify areas where ChatGPT may struggle, excel or have some shortcomings. This plays a significant role in the constant improvement of this model that leads to more accurate and appropriate responses in different contexts. Moreover, this study opens the door for future research to examine the texts of ChatGPT from different points of view as well as studying other chatbots or AI writing services and compare their results to get a better understanding of their working process.

1.4 Hypothesis

The researcher, based on the reviewed literature (see Chapter 2), expects that there is a certain fixed pattern or arrangement of the moves in the academic abstracts generated by ChatGPT. All in all, the reasons behind the fixed order and moves might be the way that ChatGPT is programmed based on the general form of the database available for the AI chatbot. The findings of the current study will be of substantial aid to detect ChatGPT-generated research articles. In addition, the researcher expects that the order of the moves in the abstracts of the human-written research articles is not fixed since different humans and thoughts generated it. The researcher anticipates that the findings of the current study can contribute to an evaluation of the quality of ChatGPT-generated academic content. The study can also help in finding out what are the areas of development for such AI models. This might be achieved by taking the limitations and biases found in ChatGPT-generated abstracts into consideration. As for the linguistic features examined, it is expected that ChatGPT-generated abstracts will have a more consistent usage of some features like tense and voice, while human authors might tend to have a varied writing style.

2. Literature Review

2.1 Overview

This chapter offers an in-depth review of the literature related to the current study. The genre of abstracts will first be defined. An overview of how the term genre came into use in discourse analysis will be presented, and the theories developed to studying genre analysis are introduced. Moreover, various studies that have examined different types of texts and of AI-generated content are reviewed.

2.2 Abstract Genre

Abstracts are an independent section used to summarize a research project, a practice or policy implementation project, a report, or as an induction of a conference or a symposium (Drury et al., 2023, p. 1; Nundy et al., 2022, p. 179). An abstract is an integral part of academic writing, as it provides an overview with the most essential details of what it summarizes. The most common types of abstracts are research study abstracts, implementation abstracts, and review articles abstracts (Drury et al., 2023). A good abstract should show the characteristics of being condensed, clear, brief, and accurate (Cross & Oppenheim, 2006, p, 432; Nundy et al., 2022, p. 180).

The key details that a research abstract must contain are background, previous literature, purpose/rational of the study, methodology, major results, conclusions, and implications (American Psychological Association, 2020; Drury et al., 2023; Kaplan et al., 1994). Other guidelines for abstracts vary according to journal guidelines and/or writing manual style (e.g., American Psychological Association). Some of these guidelines are word limit or inserting keywords after the abstract. Word limit can vary up to a maximum of 500 words (Nundy et al., 2022; PLOS Medicine, 2024). A different classification of abstracts divides them to: descriptive (i.e., for social studies), informative (i.e., scientific studies), critical, highlighting, or structured (Nundy et al., 2022, p, 181).

Studies that tackle abstracts investigate them from various approaches such as writing guides (Drury et al., 2023; Nundy et al., 2022), genre analysis (Al-khasawneh, 2017; Hyland, 2000a), linguistic analysis (Jiang & Hyland, 2017; Jiang & Hyland, 2023) and thematic analysis (El-Dakhs, 2018). Some studies used multiple approaches; for instance, Cross and Oppenheim (2006) studied the thematic, linguistic, and generic structures of abstracts.

2.3 Genre Analysis

The notion of genre analysis was proposed by the work of Halliday (1978). Martin (1992, p. 142) stated that the study of genre in linguistics is primarily based on the work of Halliday on functional grammar and his propositions of analyzing language as a social semiotic. Since this introduction of genre into linguistic studies, many researchers suggested different models or frameworks of genre analysis (Bhatia, 1993; Swales, 1990).

Within the domain of applied linguistics, researchers have adopted the term 'Genre' to make distinctions among different types of texts. Genre, in this context, can be defined as the utilization of language that follows particular schematic and textual conventions, as accepted by a specific discourse community (Widdowson, 2007, p. 129). This definition implies that each genre possesses its own set of characteristics and styles that are aligned with the expectations and norms of the respective discourse community.

Genre analysis can be defined as a branch of linguistics that concerns the systematic study of different genres, or types of texts, in order to identify their specific linguistic features, communicative purposes, and social contexts (Dudley-Evans, 1989, p. 72). It involves examining the structure, content, language use, and rhetorical strategies employed within various genres, such as academic papers, newspaper articles, emails, advertisements, and more. The aim of genre analysis is to figure out how genres are constructed, how they function within specific communities or domains, and how they shape and are shaped by social interactions (Bhatia, 1993; Martin, 1992; Swales, 1990).

Hyland (2004, p. 195) argued that genre analysis is a part of discourse analysis that tries to look into and comprehend the distinctive language patterns found within genres. In order to recognize how language is used in various genres, one must first analyze the patterns, structures, and communicative goals that are specific to each genre. Genre analysis is used by discourse analysts for their investigation of language from cognitive and social perspective, including both spoken and written discourse (Hyland, 2004).

In genre analysis, the idea of a 'move' was put out. A move stands for an important communicative element within a genre and denotes a specific objective or goal in the use of language. A genre's discourse construction is sequential and ordered since each action can be further divided into smaller phases or subunits, which can be viewed as the moves and the 'steps' used to achieve these moves (Mirador, 2000). According to Casal and Kessler (2024), steps are limited functional units that constitute a move (p. 83). The move-step analysis helps in understanding the structure of the genre and the coherence of the text.

As genre related studies have been expanding since its introduction into applied linguistic research in the 1980s, different approaches were used to study the notion of genre (Bhatia, 1993; Halliday, 1978; Swales, 1990). One of the important initial approaches was Halliday's use of the notion of register (1978), which was employed by many Australian systemic functional linguists (Martin; 1992; Eggins, 1994). According to (Biber & Conrad, 2009; Frow, 2005;), the notions of register and genre are approximately alike. Biber and Conrad (2009) suggest that the main difference is that the study of register focuses on the linguistic features that can be associated with a certain situational context (for example the register of doctors in hospital settings), while genre analysis examines the conventional structure of texts (p. 2).

In general, genre analysis offers a useful framework for comprehending how language is used within diverse genres, illuminating the rules, structures, and communication functions that support various textual forms. Researchers learn more about the complex mechanisms and methods that shape effective communication within particular discourse communities by looking at the moves and steps within genres.

2.2.1 Swales' CARS Model

Genre analyses can be carried out on any section of a research article to find the model/models that researchers follow. Swales (1990), being one of the leading researchers in the field of genre analysis, suggested a model for the Introduction section, which he describes as possibly the most perplexing part to understand in a research article. Initially, Swales (1990) argued that each group of communicative events must have some common communicative drives as a principal criterion to be characterized as a genre. Furthermore, Swales (1990) claimed that each group of communicative events defined as a genre is further composed of minute units that he defines as 'moves'. A 'move', according to Swales, is a rhetorical and discursive segment that, in a written or verbal discourse, serves as a cohesive communicative unit (Swales, 2004, p. 228).

Before reaching his conclusive view of genre analysis in 1990, Swales had multiple studies on different genres that led to his final model. At first, Swales and Najjar (1981) studied 48 research article (RA) introductions in several fields of science and proposed a four-move model of article introductions:

Move 1: Establishing the field

Move 2: Summarizing previous research

Move 3: Preparing for present research

Move 4: Introducing present research.

Swales and Najjar (1981) suggested that the structure of introductions of research articles mostly follows a certain typology of moves. Authors adhere to this typology of moves, which consists of several steps used in a predictable order. According to Swales, a ‘step’ in this regard means a lower text level, which offers various alternatives that authors can employ in the moves of the introduction. Many researchers found it problematic to distinguish between Moves 1 and 2, as reviewing literature (Move 2) was in many cases part of establishing the field (Move 1) (Dudley-Evans, 1986). In other words, many authors tend to have the literature review as a step in Move 1 (Establishing the field). In 1990, the difficulty in distinguishing between Move 1 and Move 2 led Swales to modify his original model which is called the Create a Research Space (CARS) model. In this new model, Move 2 was merged to be a step in Move 1:

Move 1: Establishing the field (topic significance – definition/overview – summarizing previous research)

Move 2: Preparing for present research (identify/indicate gap/questions/building on previous studies)

Move 3: Introducing present research (purpose – findings – research outline)

Various studies used this model for their analysis of academic content such as Jogthong (2001) who studied research articles introductions, or Candarh (2012) who studied abstracts. Those studies found that Swales' model accounts for the majority of the analyzed academic texts.

2.2.2 Bhatia's Move Analysis Framework

One of the pioneers of genre analysis frameworks is Bhatia (1993). Bhatia argued that the communicative purpose is the main factor in genre identification. His work concentrated on the benefits of genre analysis studies for the field of English for Specific Purposes (ESP). For Bhatia (1993, 2004), both students and teachers of ESP can utilize the findings and suggestions of genre analysis studies. Such studies provide a background of the moves and steps used by native users of English, especially as most of the genre analysis studies focus on analyzing English discourse or comparing a specific genre in English with another language. According to Bhatia such pre-knowledge of the techniques of language use can offer a great help in the learning process of ESP students and aid them in learning the linguistic resources that they need. Bhatia also recognizes that the content, form, intended audience, method or channel do have a role in defining a genre's nature and construction, but to a lesser degree. This suggestion of Bhatia is attributed to the fact that the communicative purposes shared within a genre is what defines it and its internal structures. In other words, a significant change in any of the communicative purposes may result in a modification in genre. Nevertheless, slight changes or adjustments do not result in such a change in genre, but help in distinguishing sub-genres. Bhatia discovered that specialists understand the communicative purposes and the construction of the genres in which they work in their daily life.

Bhatia (2004) presented a seven stages model for practical genre analysis. Those seven stages are:

1. Placing the given genre-text in a situational context
2. Surveying existing literature
3. Refining situational/contextual analysis
4. Selecting corpus
5. Studying the textual, intertextual and interdiscursive perspective
6. Conducting ethnographic analysis
7. Studying institutional context (Bhatia, 2004, pp. 189,193)

This model can serve as a roadmap for genre analysis studies by offering a practical guide for researchers. This guide starts from deciding the topic that will be studied, going through reviewing the literature and collecting the data, and continues until detailed data analyses. However, it does not offer a specific scheme for move-step analysis.

2.2.3 Hyland's Five-Move Model for Abstracts

Hyland (2000a) built his five-move model based on a review of research article abstracts from eight different disciplines, which are Molecular Biology, Mechanical Eng., Electronic Eng., Magnetic Physics, Sociology, Philosophy, Marketing, and Applied Linguistics. The five-move model includes Introduction, Purpose, Method, Product, and Conclusion (IPMPC). The model proposed by Hyland differed from the conventional IMRD model in three main areas. First, the move Purpose was separated from the Introduction move and became Move 2. This new move contained two steps: purpose and hypothesis. Second, Results and Discussion, which are Moves 3 and 4 in the IMRD model, were merged to become a single move in Hyland's model and he named it Product. Finally, a Conclusion move (Move 5) was added to the new model. These adjustments that Hyland made were based on his corpus. Hyland's (2000a) method was to build his own model to analyze the corpus of his study based on his examination of that corpus. An advantage of creating a corpus-based model is that it will certainly account for the data, unlike using an existing model, where the data is molded to fit into the model.

2.3 Genre Analysis Studies

Reviewing diverse genre analysis studies aids in understanding how each genre, field or linguistic background has its own conventions. For example, many studies have examined the differences between native and non-native English speakers across different genres, such as job applications (Al-Ali, 2004; Connor, Davis, & De Rycker, 1995), research article abstracts (Al-Khasawneh, 2017; Behnam & Golpour, 2014; Martin, 2003), and business emails (Malyuga & Mccarthy, 2018; Mehrpour & Mehrzad, 2013). These studies show that each genre has its own rhetorical moves and that the linguistic or cultural backgrounds of the writers affect their use of moves or steps.

2.3.1 Genre Analysis of Academic Writing

Some studies provide evidence that academic texts constitute a distinct genre or genres that differ from other types of language use (Biber et al., 2004; Hyland, 2000b; Swales, 1990).

For example, according to Biber et al. (2004) and Hyland (2000b), the use of hedging and nominalization was found to be a distinct characteristic of academic research articles. Hedging is employed to signal the writer's uncertainty or caution and nominalization help in being concise (Biber et al., 2004; Hyland, 2000b). Such features, hedging and nominalization, were used to develop a classification system for distinguishing academic research articles from other types of texts.

When reviewing literature, a knowledge gap can be noticed about non-IMRD 'Introduction, Methods, Results, and Discussion', research articles. Yang and Allison (2003) claim that previous research on research articles has mostly concentrated on papers with a defined framework of IMRD, ignoring the reality that many research articles do not follow this particular pattern, which resulted in this knowledge gap. Additionally, the majority of investigations have been restricted to certain fields, particularly linguistics and experimental scientific study, with few multidisciplinary studies (Jogthong, 2001; Martin, 2003; Maswana et al., 2015; Posteguillo, 1999; Soler, 2011).

In light of Yang and Allison's (2003) above suggestion about the gaps in literature, Maswana et al. (2015) conducted a study that focuses on move analysis in research articles from five different engineering domains. The study discovered that the frequency and order of moves varied between engineering domains. The domains also shared traits, indicating that there are some recurring patterns in how research publications are put together. The study contributes to the understanding of the discipline heterogeneity in move structures. Their study revealed that, despite the fact that moves and steps vary greatly within subdisciplines, some sections, like Introduction and Results, are common to all of them.

Other studies tried to conduct genre analysis on research articles from other disciplines or genres like research article introductions and titles (Jogthong, 2001; Posteguillo, 1999; Soler, 2001). Posteguillo (1999) conducted a linguistic analysis of the schematic structure used in the research articles from the field of Computer Science. The researcher selected three different academic journals to collect and analyzed 40 research articles on computing research. The findings of the study revealed that the IMRD (introduction-methods-results-discussion) pattern was not systematically used in the computer science research articles. However, the introductory and concluding sections are employed more frequently.

Another related example is the study of Jogthong (2001) who analyzed the research article introductions (RAIs) that were written by Thai academic writers in Thai language. The researcher, based on Swales' (1990) CARS model, investigated rhetorical features of 40 Thai RAIs extracted from different Thai journals specialized in educational and medical domains and compared them to the findings of Swales's (1990) study on English RAIs. Generally, the findings of the study revealed that Swales' framework goes in line with the pattern of examined RAIs. Compared to their English counterparts, Thai writers did not present their study's structure and results in the introduction section. They ended their introductions by introducing their present study, highlighting objectives of their research, and outlining implications of their research. Those differences are results of socio-cultural factors and differences in cultural-linguistics and research environments.

2.3.2 Genre Analysis of Abstracts

Another genre that attracted researchers is abstract writing in academic papers. Al-Khasawneh (2017) compared English native speakers' to non-native speakers' introduction, discussion and conclusion within abstracts using Hyland's (2000a) five-move model. To carry out the study, Al-Khasawneh (2017) compiled a corpus of 20 abstracts from two journals. The collected abstracts were equally divided between those written by English native and non-native authors. The results of Al-Khasawneh (2017) showed that native authors follow IMRD model and write conclusions in abstracts more than non-native authors. Specifically, the frequency of the conclusion move in native authors abstracts was over twice as much in non-native authors abstracts. These findings align with Liu (2015)'s study of Chinese students' conclusions that were unorganized. Liu attributed that to the lack of pragmatic awareness that non-native speakers of English possess.

Other languages were also contrasted with English in terms of the features of abstracts written in each language (Candarh, 2012; Martin, 2003). Martin (2003) compared the specifications of the genre of abstracts for scientific articles in English and Spanish. The results of Martin (2003) revealed that although the writers in both languages shared similarities in general rhetorical pattern, such as the presence of purpose, methods, and results moves, there were also differences in the frequencies and linguistic realizations of these moves. Furthermore, the finding of the study showed that the rhetorical structure of abstracts written in Spanish in the field of experimental social sciences generally mirrors the conventions of using IMRD model.

Candarh (2012) conducted a genre analysis study comparing Turkish and English research abstracts published in the field of education. The analysis was based on Swales' CARS model (2004). The corpus comprised 40 abstracts equally divided between the two languages. Overall, the findings showed that abstracts from both languages are highly similar in their selection and employment of moves, just as found by Martin (2003) above regarding Spanish and English. However, significant differences were found between English and Turkish abstracts in indicating the gap/problem of the study in Move 2, as Turkish authors did not use it at all.

Apart from language-based comparisons, Bonsu and Adusei (2023) compared how abstracts from different disciplines would approach the same topic, in terms of move-step structure. They collected 72 multidisciplinary abstracts of research articles addressing ChatGPT. Using Hyland's (2000a) IPMPC model, the study found that Product and Purpose moves were obligatory and the most frequent. Various rhetorical moves sequences were also identified such as PMP and IPPC.

Studying abstracts from only one field was also an approach found in literature. Cross and Oppenheim (2006) examined 12 abstracts from the discipline of protozoology. They followed Bhatia's seven stages model as a basis for their practical genre analysis of the abstracts. They found that the abstracts followed a five-move model, (Related research, Purpose, Methodology, Results, Discussion). Cross and Oppenheim (2006) found that Related research move was in many cases merged within Results move. Moreover, the Discussion move contained two steps, conclusion and recommendations.

2.4 Linguistic Features of Abstracts

Various studies were conducted on different linguistic features that could be observed in abstracts (Candarh, 2012; Cross & Oppenheim, 2006; El-Dakhs, 2018; Jiang & Hyland, 2017). Upon comparing English and Turkish abstracts, Candarh (2012) reported that differences were found in the lexico-grammatical choices, such as the higher tendency to use present simple and passive voice in Turkish abstracts compared to English ones. English abstracts, on the other hand, had a mixed use of present simple and past simple, and more employment of the active voice. Candarh (2012) suggested that cultural variations cause these differences. As for Cross and Oppenheim (2006), the tense/voice choices were based the rhetorical moves. This finding is consistent with Bhatia's suggestion that the rhetorical pattern can affect tense choices (1993, p. 6). Cross and Oppenheim (2006) found that while present tense and active voice are associated with Purpose and Discussion (including conclusion and recommendations steps) moves, past tense and passive voice are used in Methodology and Results moves.

Other studies focused on comparing abstracts from different academic Journals. For example, El-Dakhs' (2018) focused on the variations between the abstracts in highly prestigious journals and less prestigious journals. Journals that are not Scopus indexed were considered less prestigious ones. The study shows the differences between the two sources and offers insights into the causes of these discrepancies by studying a corpus of abstracts. El-Dakhs' (2018) found that abstracts of less prestigious journals tend to have much more details and longer sentences in the Introduction, Purpose, and Method moves. Moreover, the findings revealed that abstracts of less prestigious journals have significantly more frequent use of transition devices (e.g., 'however', 'in addition') and evidentials, which are used to refer to literature. The study clarifies the structural and linguistic variations in abstracts, highlighting the ways in which these variations are driven by the communicative goals and contextual elements unique to each genre.

As for studying more specific linguistic units, Jiang and Hyland (2017) examined how metadiscursive nouns are employed in research paper abstracts and how they help to promote interaction and cohesiveness among various abstract moves. Metadiscursives are nouns that are used to refer and describe elements of discourse itself or its participants. For example, using 'the study', or 'this introduction' to refer to the text the utterance is within. The distribution, collocations, and patterns of metadiscursive nouns across the abstract moves—such as announcing the purpose, outlining the procedures, outlining the findings, and drawing conclusions—were analyzed by the researchers. The researchers investigated a sizable collection of 240 research paper abstracts from diverse fields using a corpus-based technique. They discovered that these nouns were key in establishing ties between various moves in the abstract and connecting information. Furthermore, it was discovered that metadiscursive nouns aid in establishing a feeling of authority and impartiality in academic discourse, as they can be used to refer to the author. The study underlines the importance of taking into account the function of metadiscursive nouns in academic writing.

2.5 AI-Generated Content

AI-generated texts have recently gained much attention. ChatGPT in particular is creating new content and stirring up a lot of controversy. Wach et al., (2023) listed some of these 'darkside' controversies such as AI plagiarism, personal data violations and generating biased content. This AI program is creative and uses natural language models, which are computational models that are trained on a huge number of texts to be able to create content that mimics human language abilities and knowledge. It creates content from scratch that flows naturally in conversations. It can instantly respond to inquiries and pen poems, fan fiction, and picture books. Even without extra training or extensive years of medical schooling, ChatGPT has passed the theory portion of the United States Medical Licensing Examination. Language-based AI is already being studied by scientists. Four preprint manuscripts list ChatGPT as an author, according to Nature. Additionally, it was revealed in an article that an academic paper was produced using AI (Anderson et al., 2023).

Some studies tackled the challenges that face AI language generating models. According to Haroon and Azzam (2023), AI-generated contributions are not constrained by any contextual understanding nor affected by any sentience for the consequences that the generated language or the (re)shaped narratives might have on its receivers. This makes AI-generated narratives different from human inputs, which are still subject to their spatiotemporal situatedness, socio-cultural settings, and individual preferences. Moreover, the fact that AI models are exposed to a huge amount of data causes problems related to their outputs. For example, AI models provide answers without mentioning their sources. In addition, some of the outputs were produced by copying texts without paraphrasing (Allam et al., 2023). In addition, it is possible for AI models to produce wrong information because of their large amount of stored data.

AI language models have attracted the attention of educators, students and basically everyone in academia. This attention, coupled with the great potential that AI chatbots offer, led to some fears and recommendations that need to be addressed to reach conclusions about the guidelines of using such chatbots. AlAfnan and MohdZuki (2023) studied the writing style of ChatGPT-4. This study explored the stylistic merits of case study, business correspondence, and academic writing generated by ChatGPT-4. The study found that responses generated by the ChatGPT-4 case study are composed of two to three paragraphs with a word count of 16 to 18. Furthermore, the second-person possessive determiner 'your' and the second-person pronoun 'you' are frequently used. The study also showed that the ChatGPT-4 generated responses for academic writing are given in paragraphs of three to four sentences, with a word count of sixteen to nineteen. The majority of the sentences in academic writing are produced in a declarative mood with various present tenses and active voice with high lexical and keyword densities. Nonetheless, passive construction was used in some instances. Similar findings of AlAfnan and MohdZuki (2023) regarding the linguistic features of GPT-4 were also reported by Guo at al.'s (2023) study of the ChatGPT-3.5 answers to a diverse set of 40,000 questions.

Some recent studies compared human and ChatGPT-generated content, in a small scale (Johansson, 2023) or a large-scale comparison (Herbold et al., 2023). While Johansson's (2023) study compared one essay written by a student for an English Literature class with a one generated by GPT-4 on the same topic, Herbold et al. (2023) made a large-scale comparison of 270 essays written by students, ChatGPT-3.5, and GPT-4. Johansson's comparison of the 2000 words essay was based on two criteria: Voice Intensity Rating Scale (Helms-Park & Stapleton, 2003), and fulfilling the instructed requirements of the essay (covering the main areas of the topic, word limit, and referencing). Voice Intensity Rating Scale (VIRS) includes the following parameters: assertiveness (i.e., hedges and boosters), self-identification (i.e., personal pronouns), authorial presence (i.e., showing countervoice by contrasting own opinion with other cited work). Johansson (2023) reported that some of the references in the ChatGPT essay were 'ghost references', meaning that they do not exist, or they represent quotes that cannot be found in the original referenced study. Furthermore, the essay written by the student had significantly higher and more diverse use of boosters and hedges.

ChatGPT's element of individuality or agency was also tested by Johansson (2023). The content generated by ChatGPT was found lacking the element of agency or individuality, as evidenced by the rare meaningful use of hedges or boosters, the constant repetition of standard phrases and the use of active voice only. This last finding was also reported in other studies (Guo at al., 2023; van Woudenberg et al., 2024).

Herbold et al.'s (2023) study found that ChatGPT-generated essays, whether the 3.5 or 4 version, received much higher ratings based on parameters like cohesion, coherence, vocabulary, logic, fluency, and comprehensiveness. The ratings were obtained from 111 grammar teachers.

Also, they examined some linguistic features: Lexical diversity, sentence complexity, nominalization, presence of modals, epistemic and discourse markers. It was found that ChatGPT essays have more diversity, complexity, and higher use of nominalization. On the other hand, modals, epistemic and discourse markers were significantly more frequent in human authored essays. Interestingly, Herbold et al. (2023) reported that ChatGPT tends to exceed any word limit that human users may specify in the query. For example, when asked to write a 200 words essay, ChatGPT might write up to 300 words.

The findings of both Johansson (2023) and Herbold et al. (2023) demonstrate that ChatGPT-generated texts have better language mastery (i.e., fluency), but lack the element of individuality. The absence of individuality is caused by the lack of hedges, boosters, epistemic and discourse markers and rhetorical questions. ChatGPT texts also rarely link statements to previous literature or contrast own opinion to other opinions. Similar findings were outlined by other researchers who suggested that ChatGPT can assist in summarizing, proofreading, suggesting implications, and brainstorming; however, it fails in providing sufficient, authentic, or reliable review of literature, and it cannot offer credible answers to specific research questions (Ariyaratne et al., 2023; Buruk, 2023; Golan, 2023; Gupta, 2022; Semrl et al., 2023).

2.6 Synthesis of The Literature

Reflecting on the research questions of the current study in light of the literature reviewed, it is evident that no studies tackled the issue of comparing abstracts generated by AI models to those written by human authors from genre analysis perspective. Some studies were conducted on other abstracts comparison areas, for instance, Al-Khasawneh's (2017) comparative analysis of the abstracts written by native and non-native English authors. Others examined the differences between ChatGPT and human essays (Herbold et al., 2023; Johansson, 2023). Moreover, even the issue of move-step analysis of AI abstracts was not addressed in the literature. Subsequently came the gap that needs to be filled by attempting to answer the following first research question of this study: 'How do abstracts differ between published scientific articles and ChatGPT-generated abstracts, in terms of the moves used?'

The second research question of this study attempts to find out the linguistic features used in scientific article abstracts and ChatGPT-generated abstracts. The above review of the current literature reveals that some studies investigated some linguistic features in certain academic content. For example, Omidian et al. (2018) investigated the use of multi-word expressions (MWEs) in research articles abstracts, and Candarh (2012) studied the voice and tense of English and Turkish abstracts. However, specifying distinctive linguistic features related to academic writing and comparing them between human written and Chat-GPT generated abstracts has not been tackled in the existing body of literature.

Finally, an examination of the AI-generated language's lexical content demonstrates how a non-human agent skillfully crafts stories on complex human themes like gender performativity, human feelings, and psychological complications (Haroon & Azzam, 2023). In an attempt to address the emerging phenomenon of a nonhuman AI agency capable of freely, meaningfully, and effectively collaborating with its human users, this study addresses the issue of ChatGPT-generated academic content, specifically abstracts, using genre analysis framework. As the existing body of literature does not offer such insights based on genre analysis, the current study investigates the extent to which abstracts generated by ChatGPT mimic the genre and linguistic features of scientific article abstracts.

3. Methodology

3.1 Overview

This chapter outlines the study's methodology for analyzing abstracts produced by the artificial intelligence platform 'ChatGPT' and contrasting them with abstracts of scientific articles gathered from Q1 scientific journals that are globally recognized. These journals, according to their Impact Factor, are categorized within the first quartile (Q1) of their relevant discipline, meaning that the journal is in the top 25% of journals in its field. This chapter aims to offer a general account of the resources used, the analytical techniques employed, and the statistical analysis performed to assess how closely the generated abstracts adhered to the conventions of scientific writing. The chapter also presents the measures taken by the researcher to ensure the reliability and validity of the analysis.

3.2 Materials

The researcher collected 25 scientific articles, including empirical, position and literature review papers, from diverse 'Q1' scientific journals. The aim of choosing diverse journals is to get a multi-disciplinary representative corpus so that the findings of the study can be generalized. The collected scientific articles were chosen from issues published in 2023. The corpus includes 12 different disciplines. Those disciplines are Psychology, Economics, Biology, Physics, Geology, Artificial Intelligence, Linguistics, Sociology, Agriculture, Mechanical Engineering, Management, and Sports Medicine. Table 1 below illustrates the description of the collected corpus including journals' names from which the abstracts were extracted and disciplines, and important general corpus stats. The titles of those articles were also used to collect the ChatGPT abstracts. One abstract/title was collected from each journal, except for *Memory and Cognition* and *Reviews of Modern Physics*, from which two abstracts/titles were collected.

Table 1

Corpus Description

No.	Discipline	Journal Name
1	Psychology	*Memory & Cognition*
2		*The Journal of Creative Behavior*
3		*Educational Psychologist*
4		*Child Development Perspective*
5	Economics	*The Quarterly Journal of Economics*
6		*Journal of Political Economy*
7		*Journal of Finance*
8		*Journal of Consumer Research*
9	Biology	*Nature Reviews Molecular Cell Biology*
10	Physics	*Reviews of Modern Physics*
11		*Advanced Materials*
12	Mechanical Eng.	*IEEE*
13	Management	*Academy of Management*
14	Geology	*Geology*
15	Artificial Intelligence	*Artificial Intelligence Review*
16		*Science Robotics*
17	Linguistics	*Language*
18		*Applied Linguistics*
19	Sociology	*Human Relations*
20		*Social Media and Society*
21		*Crime Science*
22	Sports Medicine	*Sport Medicine*
23	Agriculture	*Agriculture Ecosystem and Environment*
	Total No. of Texts	50
	Word Count Total	14,154

The researcher asked the AI platform 'ChatGPT' to generate abstracts for the 25 scientific articles, separately, using the same titles of the collected scientific articles. The version used to extract the data is ChatGPT 3.5, which is a free version. The abstracts were extracted in June and July 2023. The abstracts of both collected articles from ChatGPT and the Q1 scientific articles were analyzed based on genre analysis framework. The query to ChatGPT to generate the abstracts was: 'Write an abstract for a scientific article about X', where X represents the title of the paper collected from one of the Q1 journals. The query is structured to be general to extract texts without limitations that can affect the generalization of the extracted data. The use of specific detailed instructions can restrict ChatGPT responses. On the other hand, if any details are specified or if too many instructions are given, the response of ChatGPT can be limited to certain aspects of the subject, or to a specific method of writing. Using this query would eliminate any chance for biases or limits on the subject, word count, style of writing, or the details of the generated abstract such as the results found, or the method used.

3.3 Analytical Procedures

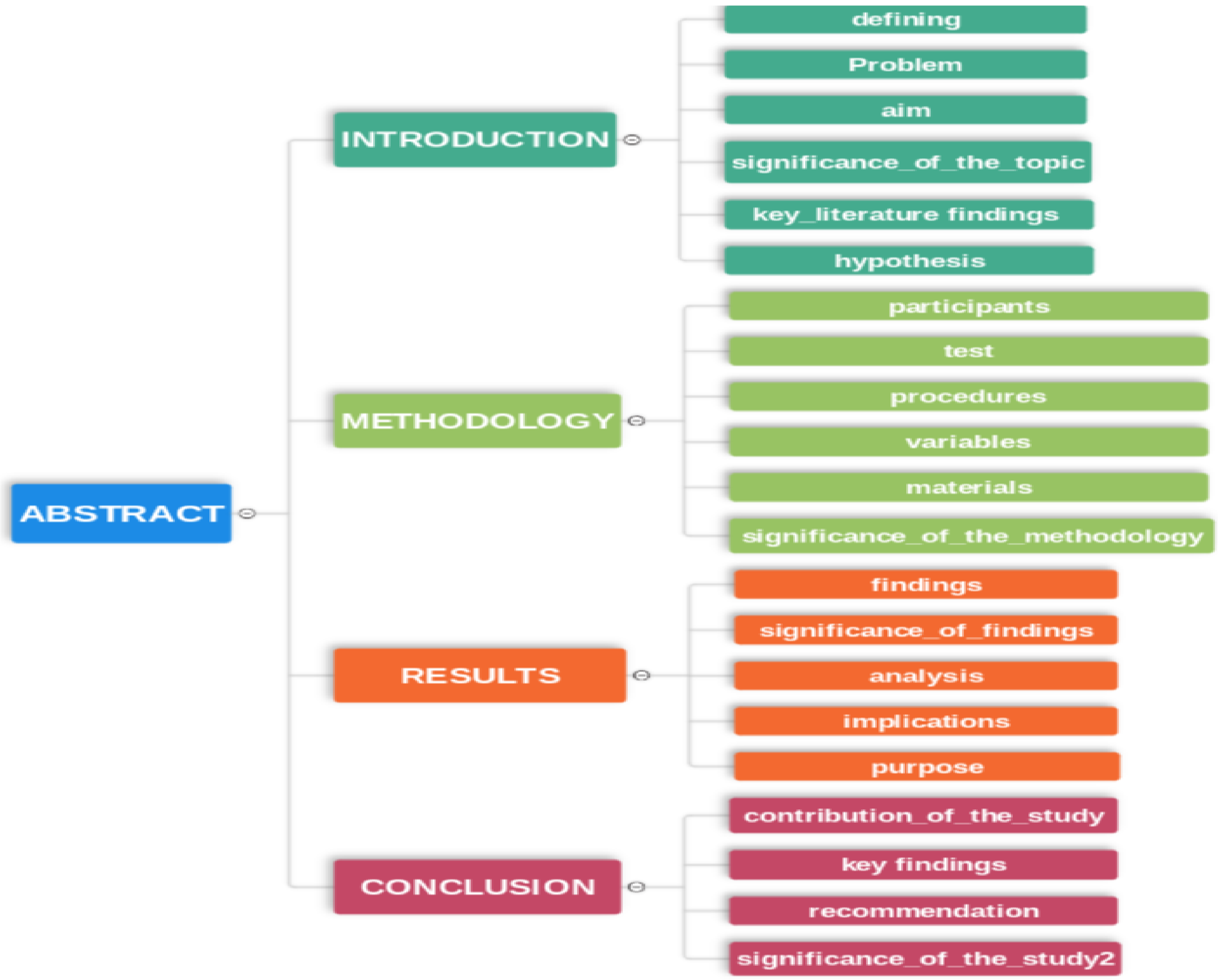
UAM CorpusTool version 3.3 (O'Donnell, 2008) was used for annotating and analyzing the collected abstracts. UAM CorpusTool is a powerful tool for performing genre analysis on different types of texts. It provides users with a range of features that help to analyze the language used in a particular text and identify patterns and structures that are typical of a particular genre. The tool has built-in schemes to annotate texts based on any level, whole texts, sentences, phrases. Moreover, users can develop their own scheme and annotate texts based on it. The tool also contains built-in statistical analysis features to analyze the annotated corpus (i.e., number of texts, words in each text/category, texts complexity, annotation comparisons). This tool is used to find the pattern (order) in the generated abstracts by ChatGPT, and to annotate the moves and minor steps in these abstracts. Figure 1 below demonstrates the scheme developed by the researcher based on the data. This scheme was developed based on the corpus of ChatGPT-generated and human-authored abstracts to analyze the collected data. The IMRD (Introduction, Methods, Results, and Discussion) model and Hyland's (2000a) IPMPC (Introduction, Purpose, Method, Product, and Conclusion) model were initially taken as a guide, but they were found to be not suitable for the collected corpus. Subsequently, the researcher made a new scheme from scratch based on the observations of the collected data.

The newly developed annotation scheme contains four moves, each of which contains various steps (21 in total). The steps are not obligatory, as some texts might contain them while others might not. Table 2 below shows examples on each step from the study's corpus. During the development of the scheme, the moves were first identified, taking the IMRD model into consideration and as a guide.

The researcher started by reading each of the collected texts and taking notes regarding the possible moves. It was first noticed that all of the collected abstracts from both ChatGPT and from the original sources start with the Introduction move, which is used to provide a background of the study and presents the aims and significance of study, along with a review of the most important relevant literature findings. Then, the Methodology move, which showcases the participants, procedures and the materials used to collect the data, was identified as most of the abstracts in the corpus contain descriptions of data collection and analysis procedures. It may also present the studied variables and the rationale behind the data collection methodology. At this point, the IM (Introduction-Methodology) pattern was noticed to be very frequent for the first two moves. However, the collected corpus showed that the four identified moves are Introduction, Methodology, Results and discussion, and Conclusion (IMRC).

Figure 1

Scheme for annotating human-written and ChatGPT-generated abstracts



The two main differences between this IMRC model and the IMRD model are that 'results and discussion' were coupled in one move, as the data in the corpus clearly showed that stating the results and providing interpretations for them were mostly presented in a one-by-one matter, in which a result is stated and then its interpretations are provided, for example:

> '*The results revealed that affective interference had a significant impact on working memory performance. Specifically, emotionally valenced stimuli, particularly negative*

stimuli, led to increased interference effects, manifested as reduced accuracy and slower response times **(findings)**. *These findings suggest that the presence of emotional stimuli can impede working memory processes, resulting in compromised cognitive performance* **(analysis)**. *Furthermore, mnemonic load exerted a significant influence on the dynamic adjustment processes in working memory. As mnemonic load increased, participants demonstrated decreased accuracy and prolonged response times* **(findings)**. *This suggests that the cognitive resources required to maintain and manipulate information in working memory become more limited as the mnemonic load intensifies* **(analysis)**.' (GPT 2).

Also, in some instances some results are stated, then some interpretations are provided, and then the text goes back to other results, for example:

'*Results revealed that seasonal N2O emissions from RR-NP and RW-NP treatments were 1.83 ± 0.18 and 1.43 ± 0.25, and 1.41 ± 0.09 and 0.88 ± 0.08 kg N ha−1, respectively, during two consecutive rice-plantation seasons, which were significantly higher than those from RR-CP treatment (1.38 ± 0.16 and 0.91 ± 0.15 kg N ha−1) and RW-CP treatment (0.95 ± 0.08 and 0.50 ± 0.10 kg N ha−1)* **(findings)**. *Higher N2O emissions from RR rotation than from RW rotation, regardless of rice planting* **(findings)**, *indicated that upland cultivation affected soil N2O efflux* **(analysis)**. *For rice planting treatments, strong positive relationships between N2O fluxes and soil-dissolved organic carbon (DOC), ammonium (NH4+), nitrate (NO3-), and functional genes (nirS, nirK, and nosZ genes) were observed* **(findings)**, *implying that soil available C and N, and related functional genes are key regulatory factors controlling N2O emissions* **(analysis)**. *Compared with the RW-CP treatment, the higher abundance of the nirK gene and lower nosZ/(nirS + nirK) ratio from the RR-CP treatment may facilitate greater N2O production, while reducing conversion of N2O to N2, resulting in increased N2O emissions* **(findings)**. *Furthermore, structural equation model (SEM) showed that field flood depth (FD), soil available C and N, and the abundance of nirS gene, together, displayed more than 70 % impact effect on N2O emission for RR-CP treatment, while, FD, soil available N, and nirS and nirK genes, together, displayed more than 60 % impact effect on N2O emission for RW-CP treatment* **(findings)**.' (Original 25).

Moreover, in some cases the purpose of the study is restated as part of the results section. The second difference is the addition of the ‘Conclusion’ move. This move was clear, especially in the ChatGPT texts. This move was found to be mainly used to restate the main findings of the study, the contribution of the study to its field, and provide recommendations. After finalizing the moves, the researcher revised each move for further classification into the steps employed to achieve each move. Table 2 below demonstrates the steps in each move with examples from the collected corpus. The ‘Source’ column provides a link to the source text. For each source link, the text source (i.e., ChatGPT or Original) and text number are displayed for reference. For example, ‘ChatGPT 3’ refers to the third abstract extracted from ChatGPT. To check the whole corpus, refer to Appendix A, which provides a link to the whole corpus, and a list of articles’ titles.

Table 2

Examples on the steps shown in the scheme from the study’s corpus

Move	Steps	Example	Source
Introduction	Defining	*‘Revenge ideation, a cognitive process involving fantasies and thoughts of seeking retribution against perceived wrongdoers,’*	GPT 3
	Problem	*‘Research on the internet and social media commonly focuses on publicly available data and therefore does not examine the more private online interactions that can take place.’*	Original 20
	Aim	*‘This study examines the phenomenon of output order effects in autobiographical memory in older adults, with a specific focus on the role of emotional organization.’*	GPT 1
	Significance of the topic	*‘Due to the field coupling between velocity and temperature, convection provides a new knob for controlling heat transfer beyond pure conduction, thus allowing active and robust thermal modulations.’*	Original 12

Table 2 (continued)

Move	Steps	Example	Source
	Key literature findings	*'Previous studies have suggested that emotional experiences are key determinants of memory organization and retrieval processes.'*	GPT 1
	Hypothesis	*'In addition, we expected group and differentiated individual-focused transformational leadership to buffer the negative indirect effect of team social capital on team innovation via team proportional task conflict.'*	Original 18
Methodology	Participants	*'Using a sample of adult participants (n = 200, mean age = 35 years),…'*	GPT 2
	Tests	*'… completed a delayed recognition working memory task with mnemonic load (High vs. Low) and affective interference (Negative vs. Neutral) parametrically manipulated.'*	Original 2
	Procedures	*'Over the course of the intervention, students are exposed to new words using both methods, and their retention and recall abilities are measured through vocabulary quizzes and recall tasks.'*	GPT 15
	Variables	*'… three conditions for presentation of word meanings: (a) definitions in the L2 (English), (b) translation equivalents in the shared school and majority language (Swedish), and (c) translation equivalents in the shared school and majority language plus other prior languages among the learners (Swedish and other).'*	Original 15
	Materials	*'… through a comprehensive analysis of primary sources, including government documents, public opinion surveys, and media representations.'*	GPT 4
	Significance of the method	*'This hybridization approach was proposed to reinforce the diversity ability of CSA and balance its search abilities for promising solutions to achieve robust search performance.'*	Original 14
Results	Findings	*'The findings reveal a significant positive impact of the education reform on educational attainment.'*	GPT 6
	Significance of findings	*'Based on the premise that the output order of AMs informs about the organisation of autobiographical memory, our*	Original 1

Table 2 (continued)

Move	Steps	Example	Source
		results highlight the role of emotional associations among AMs in old age.'	
	Analysis	*'These contrasting ages suggest a complex tectonic evolution of the subduction zone during the transition from the Permian to the Triassic.'*	GPT 13
	Implications	*'These results combined with previous data are used to present a new model for the tectonic evolution of the distal Austroalpine unit associated with the Meliata Ocean in a Wilson cycle.'*	Original 13
	Purpose	*'The research further examines the interplay between origin and substance in disruptive duplication technologies.'*	GPT 17
Conclusion	Contribution of the study	*'These techniques contribute to a deeper understanding of actinide behavior and provide critical insights for the development of advanced materials, nuclear waste management strategies, and environmental remediation technologies.'*	GPT 7
	Key findings	*'The findings highlight the diverse roles of each platform in political communication and emphasize the need for a nuanced understanding of politicians' social media practices.'*	GPT 20
	Recommendations	*'We conclude with a call for educational psychologists to turn toward critical frameworks, to center equity and justice in their work, and to honestly and intentionally grapple with our collective racist history.'*	Original 22
	Significance of the study	*'The review aims to obtain a better understanding of the applicability of the different techniques to various stacks and of the origin of apparent disagreements among literature values.'*	Original 8

To ensure the reliability and validity of the annotation, inter-rater reliability was used. In this regard, another rater, with proper experience in discourse analysis and using UAM CorpusTool and has an MA in Linguistics, participated in. The researcher explained the created scheme and the data, then trained the second rater by going through analyzing two texts with him. After that, the second rater was asked to independently analyze two texts, one from ChatGPT abstracts and the second from human-authored abstracts, under the monitoring of the researcher. Finally, when the second rater completed the analysis, the annotation data of each rater was compared for each text by comparing each step annotation. These results were tabulated after being extracted from the UAM CorpusTool tables. Then, a Python code was developed and used to calculate Unweighted Cohen's Kappa. Unweighted Cohen's Kappa was chosen as the appropriate method to conduct the inter-rater reliability because the data of the current study is categorical. The results of the test gave a score of 0.95, which indicates that there was excellent agreement. According to Landis and Koch (1977), a kappa score of ≥ 0.81 suggests an almost perfect agreement status between the compared coders.

The researcher compared the adherence of the collected abstracts to the conventions of scientific writing. Using the above scheme, this comparison was implemented to evaluate the collected abstracts based on the following aspects, aspects a-d are used to help answer research questions 1 and 3, regarding the moves and steps used in the collected abstracts and their comparison to original abstracts, while aspects e and f are used to help answering the second research question about the linguistic features.

a) Adherence to IMRD: Test the abstracts' adherence to the IMRD framework of Introduction, Methods, Results, and Discussion. Make sure the abstracts have a clear introduction, a description of the methods used, a summary of the main findings, and a discussion of the implications or importance of the results.

b) Context and background: Evaluate whether the abstracts contain enough context and background data to place the research within the field being studied. Check to see if they successfully draw attention to the area of need or issue.

c) Research goals: Look at how concisely the abstracts describe the goals or objectives of the research. Check to see if the goals are precise, quantifiable, and related to the stated research issue or problem.

d) Key findings: find out if the major conclusions or results of the research are summarized in the abstracts. Verify that the results are presented clearly and without ambiguity. Look for statements that describe statistical significance or trends in a clear and concise manner.

e) Linguistic features (grammatical features): Evaluate the use of academic language based on parts of speech, the frequency of using passive form, and the tense used. These features were chosen as they can help in investigating the stylistic differences between the compared texts. Moreover, there are conventions in academic writing related to these features (Hinkel, 2004). For example, present and past simple tenses are widely used in abstracts to describe the study (Chalak and Norouzi, 2013).

f) General text statistics: such as number of words, sentences, lexical density, and average word length.

The last two points, linguistic features and general text statistics, will be analyzed based on built-in automatic features in the UAM CorpusTool. For linguistic features, an automatic coding scheme that is built-in the UAM CorpusTool was used to find the grammatical features of the texts. Systemic functional linguistics (SFL) Mood, which is an annotation scheme that can be applied automatically, is highly appropriate for fulfilling the same aim, as its scheme already has the required linguistic features targeted to be investigated in the current study. It should be noted that this annotator provides various results, but only the illustrated ones above will be checked. As for general statistics, UAM also provides a built-in feature by which such aspects related to general text statistics, as the ones mentioned above, can be automatically calculated.

By comparing the adherence to these scientific writing conventions, one can gain insights into how well the abstracts generated by ChatGPT align with the standards of scientific writing according to the assigned scheme and based on IMRD (Swales, 1990).

The move and step variables were analyzed and compared using chi-squared (χ^2) tests to find out their use and differences between the two types of texts, the AI-generated abstracts and the original ones. The variables that were studied are the numbers and percentages of moves overall, the number of sentences used to achieve each move, and the frequency of IMRC. Moreover, the frequencies of other sequence patterns were also analyzed.

4. Results and Discussion

4.1 General Texts Statistics

This chapter, based on UAM CorpusTool statistics, presents the results of the study by highlighting the findings of genre analysis of the ChatGPT-generated texts and original texts extracted from Q1 journals. Moreover, general linguistic statistics of the texts will also be presented, such as number of words, average words length, and lexical density of the texts. Finally, the findings of the analysis of the academic language (grammatical features) of the texts will also be demonstrated. Furthermore, in the current chapter, each of these results will be presented in a separate section and thoroughly discussed. When comparing the two types of texts, ChatGPT and original abstracts, *p*-values will be used in order to discover whether the differences between them are statistically significant. The significance level is set to .05.

Table 3 below provides the general statistical findings of the texts collected for the present study, and their comparison according to the genre of the text. As demonstrated in the table below, the corpus of this study was compiled of 50 scientific articles' abstracts divided equally between two sources, ChatGPT-generated abstracts and original abstracts extracted from Q1 journals. The table shows that ChatGPT-generated texts are more lexically dense and complicated as shown with the number and percentage of lexemes per text. Also, they are generally much longer texts. These aspects can be shown in the fact that ChatGPT abstracts are composed of about double the number of words in the original abstracts, as ChatGPT collected abstracts are composed of (9,770 words), compared to (4948) words in original abstracts. Subsequently, the tokens number in ChatGPT abstracts (10841) was much higher than the number of tokens in the original abstracts (5700). Moreover, ChatGPT abstracts have a higher level of text complexity than original abstracts, this is because the average word length and segment length are higher in ChatGPT abstracts. Finally, the abstracts generated by ChatGPT are lexically denser, meaning that the percentage of lexemes per text is higher in them than in the original abstracts; however, the difference is not very big, 65.68% compared to 63.16%. Other studies on ChatGPT-generated content found similar results (AlAfnan & MohdZuki, 2023; Herbold, 2023). On the other hand, contradictory findings were observed by Guo et al. (2023), which revealed that human experts' answers to diverse questions were lexically denser than the answers generated by ChatGPT. For abstracts, human-authors might be intentionally resorting to less complex and lexically dense abstracts to make them readable and understandable since lexically dense texts tend to be harder to read and understand (Kaplan et al., 1994).

Table 3

Comparative analysis of the general text statistics of ChatGPT and original abstracts

Feature	ChatGPT abstracts	Original abstracts
Length		
Number of texts	25	25
Tokens in texts	10,841	5700
Words in texts	9597	4557
Text Complexity		
Av. Word Length	6.49	5.92
Av. Segment Length	382.2	182.2
Min. Segment Length	340	100
Max. Segment Length	506	472
Lexical Density		
Lexemes per text	256.68	125
Lexemes % of text	65.68%	63.16%

Although the corpus is equally distributed between the two types of texts and they address the same subjects, ChatGPT-generated abstracts are much longer than original abstracts. This indicates that ChatGPT's language model attempts to offer more details about the study in the abstract, while original abstracts are mostly less detailed. A possible argument that explains this finding can be that authors of original abstracts tend to offer a general view of the study without going into too many details. The general tendency of ChatGPT to generate longer content, including academic content, than humans was found by multiple studies (Cai et al., 2023; Guo et al., 2023; Herbold et al., 2023). Another interpretation might be that there are no length constraints governing the abstracts generated by ChatGPT. However, even when conditioned by a certain word limit, ChatGPT tends to exceed the set limitI. Similar observations were found by Herbold et al. (2023). On the other hand, original texts may be governed by a specific number of words. Thus, academic writers attempt to provide a general overview of their articles without offering excessive details.

4.2 Genre Analysis

One of the core objectives of the current study is to examine and analyze the texts based on genre analysis measures, namely the moves and steps employed. Table 4 below presents a comparative analysis of the moves and steps of ChatGPT-generated abstracts and original abstracts based on the annotation scheme which includes 4 moves and 21 steps.

The first table shows the moves first, namely, Introduction, Methodology, Results, and Conclusion, with a presentation of the numbers and percentages for each move for ChatGPT-generated abstracts and original abstracts, respectively, and then presents the results of chi-square tests and the associated p values. Furthermore, each of the following tables in this section then present the percentages and numbers of the steps for each move separately. The *p*-value is also used to compare the steps in each move.

As shown in Table 4, the third move, Results, was the most employed move in the ChatGPT-generated abstracts with a percentage of 39.79%, and the least employed move by the AI chatbot was Methodology, 16.28%. On the other hand, the Introduction move was the most used by the authors in the original abstracts (43.63%), while the Results move was the second most used (30.88%). Contrary to ChatGPT-generated abstracts, the least frequent move in original abstracts written by human authors was conclusion (5.39%). Similar findings for human-written abstracts were also indicated by multiple studies (Al-Khasawneh, 2017; Bonsu and Adusie, 2023; Candarh, 2012; Hyland, 2000a). Furthermore, it is shown in the table that there were significant differences between ChatGPT and original abstracts in the Introduction and Conclusion moves. In the following subsections, the results of each move will be presented separately.

Table 4

Comparative analysis of the moves of ChatGPT abstracts and original abstracts

Moves	ChatGPT abstracts N=387		Original abstracts N=204			
	N	Percent	N	Percent	χ^2	*p*-value
Introduction	98	25.32%	89	43.63%	20.69	P < .001
Methodology	63	16.28%	41	20.10%	1.34	.719
Results	154	39.79%	63	30.88%	4.57	.206
Conclusion	72	18.60%	11	5.39%	19.32	P < .001

According to the current study's findings, the Introduction move accounted for 43.63% of the content in original abstracts written by humans, making it the most common move overall. This complies with accepted conventions for academic writing, pointing out the aims of the study, defining the research context, and summarizing the study's significance. Such conventions are proven by the findings of most studies that some Introduction steps are obligatorily present in abstracts (Al-Khasawneh, 2017; Bonsu & Adusie, 2023; Hyland, 2000a). In contrast, the original abstracts made the least use of the Conclusion move. This may be explained by the general emphasis on abstract shortness, the ranking of important findings, and the custom of saving in-depth conclusions for the paper's main body. Moreover, significant differences were found in two moves between ChatGPT and human written abstracts, Introduction and Conclusion, while no such significant differences were found in the Methodology moves and the Results moves. This raises concerns regarding the model's compliance with accepted academic writing rules as it implies that ChatGPT's abstract creation deviates methodically from human-authored patterns. While the pattern mostly identified in human-authored abstracts is more frequent Introduction moves than Conclusion moves (Al-Khasawneh, 2017; Candarh, 2012; El-Dakhs', 2018; Jiang & Hyland, 2017), ChatGPT shows the opposite pattern (Johnson, 2023). The findings of Johansson (2023) show that the human essay had longer introduction, but shorter conclusion than the ChatGPT-generated essay.

Although ChatGPT did not conduct substantive research, the abstracts it generated had Methodology moves, which ideally contain a detailed descriptions of data collection and analysis procedures. However, the frequency of Methodology move was higher in original abstracts than in ChatGPT abstracts. An explanation can be that ChatGPT abstracts are not based on an actual study where there was a data collection process, making ChatGPT brief about such details.

Several studies found that ChatGPT can limitedly help in developing or suggesting research methodology (Khalifa and Albadawy, 2024). On the other hand, the Result moves and the Conclusion moves can be based on more abstract details that the model may extract based on its training data.

The study's subsequent promise to provide comprehensive findings for every step individually shows that it is dedicated to elucidating the subtle differences between the rhetorical structures used by ChatGPT and human writers. These findings underscore the multifaceted influences, such as training biases, the phrasing of the query, or the genre being generated, that shape the outputs of language models and highlight the need for a comprehensive understanding of their training dynamics. The training biases are caused when the AI algorithms are trained on imbalanced or unrepresentative datasets leading to skewed outcomes in their generated content (Brown et al., 2020; Buruk, 2023; Dergaa et al., 2023).

4.2.1 The Introduction Move

The first move found in the corpus is Introduction, which was found to have six possible steps, those are: defining, problem, aim, significance of the topic, key literature findings, and hypothesis. As table 5 above shows, the aim step was the most employed step in both ChatGPT and original abstracts with percentages of 36.73% and 37.08% respectively. Also, the hypothesis step was not present in any of the collected abstracts from ChatGPT, but there were two instances of using this step in the original abstracts. Among the steps that were present in the ChatGPT-generated abstracts, key literature findings step was the least frequent with a percentage of only 6%. Also, there were significant differences between ChatGPT and original abstracts only in reviewing the literature in the Introduction move.

Table 5

Comparison of the Introduction steps in ChatGPT abstracts and original abstracts

	ChatGPT abstracts		Original abstract			
Introduction	N=98		N=89			
Steps	N	Percent	N	Percent	χ^2	*p*-value
Defining	31	31.63%	13	14.61%	7.51	.185
Problem	11	11.22%	8	8.99%	0.26	.998
Aim	36	36.73%	33	37.08%	0.00	1
Significance of the topic	14	14.29%	3	3.37%	6.72	.242
Key literature findings	6	6.12%	30	33.71%	22.83	< .001
Hypothesis	0	0.00%	2	2.25%	2.23	.816

The aim step was found to be the most used step in both ChatGPT and original abstracts. This finding shows that both ChatGPT-generated and human-authored abstracts emphasize the need of properly conveying the purpose or objective of the research. This is a result of many factors; for example, during the training process, ChatGPT collects patterns from a variety of data sources, including scholarly publications. The terms 'aim', 'goal', 'seek', or any other constructions conveying the aims of any article are frequently used in different texts. Most abstract analysis studies indicated similar findings showing that the aim/purpose was the most common step/move (Al-Khasawneh, 2017; Bonsu & Adusie, 2023; Candarh, 2012; Hyland, 2000a; Jiang & Hyland, 2017; Martin, 2003). Consequently, when ChatGPT is asked to summarize an article or generate an abstract, the model will learn to provide responses that contain the phrase 'aim' or any of its synonyms if it is frequently used in the context of study objectives. In addition, a clear, succinct, and useful summary of the research emphasis is provided by including aims in original abstracts, which fulfills a crucial function. It contributes to the success of scholarly interaction by improving the way readers understand the goal and aim of the study. Furthermore, having the study's objectives stated in the abstract makes it easier for readers -including researchers and peer reviewers- to determine if the study fits with their needs or interests. Additionally, adhering to accepted academic writing conventions, such as mentioning objectives in abstracts, ensures uniformity throughout scholarly papers. It enables researchers to follow accepted guidelines and procedures.

One noteworthy finding is that, in contrast to two instances where it was included in the original abstracts, the hypothesis step is absent from all the gathered abstracts from ChatGPT. This mismatch might indicate a bias or limitation in ChatGPT's handling of this particular abstract content element by highlighting differences in the modeling of hypotheses within the language model's outputs. Furthermore, this also indicates that ChatGPT lacks the ability to decide what hypotheses can be expected by only relying on the title of the study. Such inability of ChatGPT was observed in many studies that found the academic content generated by ChatGPT to be superficial (Ariyaratne et al., 2023; Buruk, 2023; Semrl et al., 2023). Moreover, the absence of the hypothesis in ChatGPT's abstracts and the rare use of it in the original abstracts might be a result of the fact that, in different academic fields, abstracts usually summarize the key points of the study, which are the aims, methodology, results, and conclusions, rather than stating the hypotheses. Because abstracts need to be brief, while additional significant details may take precedence in the available space.

With a percentage of just 6%, the analysis of every step included in ChatGPT-generated abstracts revealed that the least repetitive step was key literature findings. This implies that, in contrast to other steps of the introduction, ChatGPT tends to present key literature findings less repeatedly. The researcher argues that since ChatGPT relies on a limited set of data, although it is a very large-scale data, the ability to cite the findings of other studies would be harder for ChatGPT. Furthermore, the research findings revealed highly significant differences between ChatGPT and the original abstracts in key literature findings step. Certain variances might represent changes in the way ChatGPT produces material on certain introduction-related elements, possibly resulting from subtleties or biases in the training set. Multiple studies found matching results demonstrating that ChatGPT fails to provide sufficient or accurate citations (Ariyaratne et al., 2023; Buruk, 2023; Johansson, 2023; Semrl et al., 2023).

For further understanding of the Introduction move, Examples 1 and 2 below showcase two Introduction moves examined in the study and that were extracted from ChatGPT and Original abstracts respectively:

<u>Example 1 (GPT 21) :</u>

*'This study investigates the factors influencing people's propensity to legitimize and cooperate with the police in Quito, Ecuador, focusing on the impact of procedural justice (***aim***). Legitimacy and cooperation with law enforcement are essential for effective policing and maintaining social order (***significance***). However, little is known about the specific mechanisms that foster public support and cooperation with the police, especially in contexts characterized by diverse social and cultural norms (***problem***).*'

<u>Example 2 (Original 12):</u>

*'Convective thermal metamaterials are artificial structures where convection dominates in the thermal process (***defining***). Due to the field coupling between velocity and temperature, convection provides a new knob for controlling heat transfer beyond pure conduction, thus allowing active and robust thermal modulations (***significance***). With the introduced convective effects, the original parabolic Fourier heat equation for pure conduction can be transformed to hyperbolic (***literature***). Therefore, the hybrid diffusive system can be interpreted in a wave-like fashion, reviving many wave phenomena in dissipative diffusion (***literature***). Here, recent advancements in convective thermal metamaterials are reviewed and the state-of-the-art discoveries are classified into the following four aspects, enhancing heat transfer, porous-media-based thermal effects, nonreciprocal heat transfer, and non-Hermitian phenomena (***aim***). Finally, a prospect is cast on convective thermal metamaterials from two aspects. (***aim***).'*

To sum up, the examination of the Introduction move offers insightful information on the subtle distinctions between abstracts written by humans and those created by ChatGPT, pointing out both parallels and discrepancies in the application of steps in the introduction. The fact that ChatGPT-generated abstracts lack the hypothesis step and that some steps differ significantly between the two types of texts highlights the need for a comprehensive knowledge of the model's capabilities when it comes to replicating similar texts to those written by humans.

4.2.2 The Methodology Move

Describing the methodology was the second specified move found in the collected corpus. This move was carried out in both ChatGPT and original abstracts by employing six steps: participants, test, procedures, variables, materials, and significance of the methodology. Table 6 below demonstrates a comparison of the Methodology steps between ChatGPT and original abstracts.

Table 6

Comparison of the Methodology steps in ChatGPT abstracts and original abstracts

	ChatGPT abstracts		Original abstract			
Methodology	N=63		N=41			
Steps	N	Percent	N	Percent	χ^2	*p*-value.
Participants	5	7.94%	7	17.07%	2.03	.944
Test	3	4.76%	4	9.76%	0.99	.963
Procedures	21	33.33%	14	34.15%	0.01	.999
Variables	13	20.63%	3	7.32%	3.38	.641
Materials	17	26.98%	8	19.51%	0.76	.979
Significance of the methodology	4	6.35%	5	12.20%	1.07	.956

.

For ChatGPT abstracts, the most employed step in the methodology move was procedures, 33.33%, and least one was (test, 4.76%). Similar to ChatGPT-generated abstracts, the most employed step in original abstracts was (procedures, 34.15%); however, the least one was variables, 7.32%, noting that test step was the second least used step, with a percentage of 9.76%. Moreover, no significant differences were found between the two datasets.

The Methodology move findings in the gathered corpus shed light on how information on research procedures is organized and presented in abstracts created by ChatGPT and original ones alike. The Methodology step that was used the most often in ChatGPT abstracts was procedures, indicating that in ChatGPT created abstracts the actions and processes involved in the study procedures are explained in depth. This step is employed to indicate how the data was collected making it more important and an integral part of a Methodology move. Similarly, Alshater (2022) and Dergaa et al. (2023) suggested that ChatGPT can be used to develop research methodology and design. Other steps in Methodology can be neglected in the abstract. For example, there must be a certain procedure according to which the data was collected, these procedures are not obliged to contain a test or certain materials or even without the need for participants (Drury et al., 2023). On the other hand, the test step was the least used, accounting for 4.76% of all steps. This suggests that the importance of clearly outlining the tests carried out during the research process is somewhat diminished.

Similarly, the procedures step, which make up 34.15% of the content in original human-authored abstracts, have been identified as the most often used step in the Methodology move. This is consistent with the need to give a thorough description of the study methods. However, at 7.32%, the least utilized step was variables step, and at 9.76%, the second least used step was test. This implies that human authors, similar to ChatGPT, also have a tendency to include less specific information on the variables and tests used in the methodology move in abstracts.

Hyland (2000a) indicated that the main aim of the Methodology move in abstracts is to outline the procedures.

The biggest difference revealed was in the variables step, which was not significantly different between ChatGPT and the original abstracts within the Methodology move. This suggests that although there were some discrepancies in how the variables were presented between the two datasets, these differences were only marginal. The use of the Methodology move was not significantly different between the types of texts, as the way it was employed by academic authors and by ChatGPT did not differ significantly, indicating that the interplay or options available when producing Methodology move are not of a wide range and are strict to limited choices. Examples 3 and 4 below, extracted from the corpus of the study from ChatGPT and Original abstracts respectively, can aid in understanding how the steps of the Methodology move are employed. While example 3 shows how only three steps were used to accomplish the Methodology move in this abstract, example 4 shows how a single step can occur several times within the same move.

<u>Example 3 (GPT 10):</u>

*'Using a comprehensive dataset of loan-level information from Chinese banks (***materials***), this study analyzes the relationship between monetary stimulus measures and infrastructure investment outcomes (***procedures***). The analysis focuses on the effects of interest rate adjustments, reserve requirement ratio changes, and credit supply expansions on the financing and implementation of infrastructure projects (***variable***).'*

<u>Example 4 (Original 24):</u>

*'Design: Umbrella review (***procedure***). Twelve electronic databases were searched for eligible studies published from inception to 1 January 2022 (***materials***).*
*Eligibility criteria for selecting studies: Systematic reviews with meta-analyses of randomised controlled trials designed to increase physical activity in an adult population and that assessed depression, anxiety or psychological distress were eligible (***procedures***). Study selection was undertaken in duplicate by two independent reviewers (***procedures***). Ninety-seven reviews (1039 trials and 128 119 participants) were included (***participants***). Populations included healthy*

*adults, people with mental health disorders and people with various chronic diseases (***participants***).'*

The analysis of the Methodology move showed that both ChatGPT-generated and human-authored abstracts place an equal priority on describing the study methodologies. Minor discrepancies in the ways that the language model and human authors describe different parts of the study technique are highlighted by the differences in the least used steps, for instance, test and variables. The slight significance of the variations in variables points to some agreement between ChatGPT-generated and the human-authored abstracts in handling this move in the dataset under study.

4.2.3 The Results Move

The most overall used move in all of the collected corpus was Results. This move was employed using five steps: findings, significance of the findings, analysis, implications, and purpose. Table 7 below also shows that there were no significant differences in any of the steps.

Table 7

Comparison of the Results steps in ChatGPT abstracts and original abstracts

	ChatGPT abstracts		Original abstract			
Results	N=154		N=63			
Steps	N	Percent	N	Percent	χ^2	*p*-value
Findings	63	40.91%	39	61.90%	7.91	.094
Significance of findings	16	10.39%	1	1.59%	4.80	.308
Analysis	34	22.08%	19	30.16%	1.58	.812
Implications	36	23.38%	4	6.35%	8.62	.071
Purpose	5	3.25%	0	0.00%	2.09	.719

As the table shows, findings was the most frequent step for both ChatGPT (40.91%) and original abstracts (60.9%), and purpose step was the least frequent one for both. Moreover, only one instance of significance of the findings step was found in the original abstracts' corpus.

Before going into the discussion of the findings demonstrated in table 7, the examples provided below can help to put the results of this study in context regarding the use of the Results move. Example 5, which is ChatGPT-generated, shows how the Results move usually has more details in ChatGPT abstracts.

<u>Example 5 (GPT 16):</u>

*'The findings reveal that Chinese learners often struggle with the use of English greetings' fixed expressions and idiomatic phrases, leading to literal translations and inappropriate usage (***findings***). Additionally, the pragmatic aspects of greetings, such as register, tone, and politeness levels, pose challenges for Chinese speakers (***findings***), as the cultural norms surrounding greetings in English differ from those in Chinese (***analysis***). The study also identifies areas of improvement in English language teaching for Chinese learners concerning greetings (***findings***). Based on the analysis of learner errors, recommendations are provided for instructors to address specific difficulties and promote cross-cultural understanding in greetings (***implication***). The research contributes to the broader field of second language acquisition and intercultural communication by highlighting the intricacies of greetings as a linguistic and cultural phenomenon (***significance of the findings**). *Understanding the challenges faced by Chinese speakers in greeting in English sheds light on the importance of teaching culture-specific language aspects to enhance learners' communicative competence in cross-cultural contexts (***significance of the findings***).'*

Example 6 *(<u>Original 25</u>)*

*'Results revealed that seasonal N2O emissions from RR-NP and RW-NP treatments were 1.83 ± 0.18 and 1.43 ± 0.25, and 1.41 ± 0.09 and 0.88 ± 0.08 kg N ha−1, respectively, during two consecutive rice-plantation seasons, which were significantly higher than those from RR-CP treatment (1.38 ± 0.16 and 0.91 ± 0.15 kg N ha−1) and RW-CP treatment (0.95 ± 0.08 and 0.50 ± 0.10 kg N ha−1) (***findings***). Higher N2O emissions from RR rotation than from RW rotation, regardless of rice planting (***findings***), indicated that upland cultivation affected soil N2O efflux (***analysis***). For rice planting treatments, strong positive relationships between N2O fluxes and soil-dissolved organic carbon (DOC), ammonium (NH4+), nitrate (NO3-), and functional genes (nirS, nirK, and nosZ genes) were observed (***finding***), implying that soil available C and N, and related functional genes are key regulatory factors controlling N2O emissions (***analysis***). Compared with the RW-CP treatment, the higher abundance of the nirK gene and lower nosZ/(nirS + nirK) ratio from the RR-CP treatment may facilitate greater N2O production, while reducing conversion of N2O to N2, resulting in increased N2O emissions (***finding***). Furthermore, structural*

*equation model (SEM) showed that field flood depth (FD), soil available C and N, and the abundance of nirS gene, together, displayed more than 70 % impact effect on N2O emission for RR-CP treatment, while, FD, soil available N, and nirS and nirK genes, together, displayed more than 60 % impact effect on N2O emission for RW-CP treatment (***finding***).'*

The findings about the Results move in the gathered corpus demonstrate how often this move is and offer insights into how study findings are presented and discussed in both human-written and ChatGPT-generated abstracts. Results move occupies the most space of abstracts in both ChatGPT and original abstracts. This finding is consistent with Hyland's (2000a) finding that abstracts contain complex and long Results moves.

The step in the Results move that was used the most frequently in both the original abstracts and ChatGPT was the findings step. This finding implies that there is a widespread focus on clearly outlining the main conclusions of the study in the abstracts' results section. The higher percentage of the findings step in original abstracts might indicate a more detailed or extensive presentation of findings by human authors. The main conclusions and outcomes of a study are usually presented in the 'results' section, which is of great importance for readers. Additionally, presenting the results in the abstract makes it easier for readers to extract the study's impact and contribution without having to read the entire study. Moreover, ChatGPT is exposed to various studies from the internet during training, where abstracts with an emphasis on results are common. With the employment of these data, ChatGPT gains knowledge, and the patterns it finds during training may be mirrored in its output as summarizations and implications of its training data. Conversely, the purpose step was found to be the step that was least repeated in the original abstracts and ChatGPT. This shows that there may be less of a tendency to repeat study-related information within the results move since there are more important details to be mentioned than the repetition of the purpose of the study. Those findings are reflected by the assertion by multiple studies that the content generated by language models is based on the training data (Bender & Friedman, 2018; Van Dis et al., 2023) and that summarizing content is one of ChatGPT's main points of strength (Brown et al., 2020; Buruk, 2023).

Interestingly, there is only one occurrence of the significance of the findings step and it was found in the corpus of original abstracts. This step was uncommon in both ChatGPT and original abstracts. This could mean that authors, both human and AI-generated, may not always state the importance of the results in their abstracts. In common, abstracts are brief, thus writers have to focus on delivering the most important details so that readers may quickly get the overall idea of the study in a concise way (Kaplan et al., 1994). Abstracts commonly follow a traditional format that highlights crucial components such as background, goals, procedures, findings, and conclusions (Hyland, 2000a). In the introduction, discussion, or conclusion sections of the whole paper, where they have more space to provide context and address implications, authors frequently go into greater detail about the significance of their work. Consequently, during training, ChatGPT was likely exposed to abstracts that rarely mentioned the significance of the study. Accordingly, it produces abstracts with rare or not mentioned significance. Authors may tend to spare this step to the Conclusion move.

The statistical analysis showed that ChatGPT and the original abstracts did not differ significantly in any of the steps. This implies that there are no significant differences in the way these procedures are used when presenting the findings. However, the disparities in the findings step suggest potential differences in the concentration on clarity, detail, or emphasis between the two sets of abstracts.

The frequency of the implications step differs noticeably between abstracts written by humans and those created by ChatGPT. The implications step appears 36 times in the collection of abstracts produced by ChatGPT, suggesting a comparatively frequent inclusion of thoughts regarding the wider meaning or ramifications of the study's findings. This may suggest that ChatGPT frequently highlights or goes into further detail on the implications, real-world uses, or prospective paths forward that result from the research it summarizes. Golan (2023) and Gupta (2022) provided similar arguments that ChatGPT can impressively help in suggesting research implications. On the other hand, the implications step was found only four times in the human-authored abstracts from the same sample. This decreased frequency could indicate that human authors in this sample either prefer to express implications more succinctly or incorporate them more sparingly. It might also suggest a shift in focus when it comes to talking about the wider significance or implications of the study's findings. However, the collected data might also be biased if studies are argued to have no proper inducted implications that can be suggested by the authors. The analysis of the Results move emphasizes how important it is to present findings in abstracts, and both ChatGPT and human-authored abstracts place a high value on this process. The statistical analysis of these variances, as well as the variations in repetition and the inclusion of specific processes, offer important insights into the subtle differences between ChatGPT and human authors when it comes to the presenting of the studies' findings in the abstracts. It can also be noted that the analysis step is more frequent in original abstracts. Most likely, although ChatGPT can make up findings based on its data, tailoring explanations for those findings might be harder for the AI model (Buruk et al., 2023; Semrl et al., 2023).

4.2.4 The Conclusion Move

The last identified move in the collected corpus was Conclusion. This move is achieved through four steps: contribution of the study, key findings, recommendations, and significance of the study. In both ChatGPT and original abstracts, contribution of the study was the most employed step, with percentages of 58.33% and 63.64% respectively. However, while the least used step in ChatGPT abstracts corpus was significance of the study (5.56%), key findings step (0%) was the least used one in original abstracts. Moreover, the findings revealed that there were no significant differences between ChatGPT-generated abstracts and original ones in all four steps. The results of the comparison between ChatGPT and original abstracts regarding the last move are shown in table 8 below.

Table 8

Comparison of the Conclusion steps in ChatGPT abstracts and original abstracts

	ChatGPT abstracts		Original abstracts			
Conclusion	N=72		N=11			
Steps	N	Percent	N	Percent	χ^2	*p*-value
Contribution of the study	42	58.33%	7	63.64%	0.11	.990
Key findings	19	26.39%	0	0.00%	3.76	.288
Recommendation	7	9.72%	3	27.27%	2.77	.428
Significance of study	4	5.56%	1	9.09%	0.21	.975

The findings concerning the Conclusion move in the collected corpus shed light on how conclusions are organized and communicated in research abstracts by ChatGPT-generated and human-authored abstracts. Contribution of the study step was the most frequent step in both ChatGPT and original abstracts. This implies that both ChatGPT and human writers emphasize underlining the study's contribution or significance in the Conclusion move, demonstrating a common rhetorical technique. On the other hand, differences occurred in the last step of the Conclusion move, significance of the study, which was the least used step in ChatGPT-generated abstracts. Conversely, the key findings step was never utilized in the original abstracts. This suggests that human authors might be less likely to specifically highlight important findings in this section, even though ChatGPT tends to place less emphasis on describing the study's significance in the conclusion. Academic authors seem to rely on providing the key findings in the results section and focus on the overall contribution of the study along with the recommendations in the Conclusion move (Al-Khasawneh, 2017; Bonsu & Adusie, 2023; Candarh, 2012; Hyland, 2000a). Conversely, as part of ChatGPT strategy of excessive detailing in the abstract, the key findings are more likely to be used in ChatGPT-generated abstracts. This can be part of ChatGPT's tendency to write long texts whatever is the content (Cai et al., 2023; Guo et al., 2023; Herbold et al., 2023).

Some variations were found in key findings and recommendations steps between the original abstracts and those generated by ChatGPT, according to the statistical analysis. This implies that although there are slight distinctions, they are not significantly different from one another. The Conclusion move's presentation of the recommendations by ChatGPT and human authors does not differ statistically significantly, as indicated by the *p*-values.

In conclusion, the analysis of the last move emphasizes a common focus on highlighting the study's contribution from both ChatGPT and human writers. Variations in the least used steps point to subtle differences in the conclusion's clear discussion of the study's importance and important findings. The moderate variations in how ChatGPT-generated and human-authored abstracts handle these steps show no significant differences. However, comparing the frequency of the Conclusion move itself in the whole abstracts to other moves reveals significant differences, which indicate that the area of variation here between ChatGPT and human authors is in employing the move in the first place. ChatGPT-generated abstracts had more Conclusion moves that occupy more space in the abstract. This finding might indicate a consistent trend for ChatGPT as similar findings were reported by Johansson (2023) for essays. However, such deduction would require further investigation to be assured. This tendency by ChatGPT can be attributed to its ability to generate abstract conclusions with vague recommendations and repeated findings, while in actual human authored articles, authors tend to have a short section for conclusion. The two examples below demonstrate this variation and how ChatGPT, as in Example 7, offers very general conclusions based on its data.

Example 7 (GPT 22):

'In conclusion, this article presents a comprehensive analysis of the racisms of commission and omission in educational psychology through a historical lens and systematic review (**contribution of the study**). *The research underscores the importance of acknowledging and confronting historical biases and inequalities in the field* (**key findings**). *By shedding light on the presence and impact of racisms in educational psychology, this study calls for transformative efforts to promote racial equity and social justice within the discipline and the broader educational context. (***recommendation***).'*

Example 8 (Original 24):

*'Physical activity is highly beneficial for improving symptoms of depression, anxiety and distress across a wide range of adult populations, including the general population, people with diagnosed mental health disorders and people with chronic disease (***contribution of the study***). Physical activity should be a mainstay approach in the management of depression, anxiety and psychological distress (***recommendation***).'*

4.3 Grammatical Features

As clarified earlier, the linguistic features employed in the texts are one of the aspects of comparison between the groups in the corpus is. The focus was on parts of speech, tense-aspect, voice, and the use of contractions. Regarding the latter, no use of contractions was observed in both ChatGPT and original abstracts because the employment of contractions is not favored in academic writing.

Table 9 below shows that nouns are the most present part of speech in both ChatGPT abstracts (38.3%), and original abstracts (36.03%), with no significant differences between them.

Similarly, adjectives came second, and verbs came third in both without any significant differences between both sources.

Table 9

Comparative analysis of the grammatical features of ChatGPT and original abstracts

	GPT		Original			
	N	Percent	N	Percent	χ^2	*p*-value
POS	N=9597		N=4557			
Noun	3676	38.30%	1642	36.03%	6.79	.559
Verb	1261	13.14%	617	13.54%	0.43	.999
Adjective	1449	15.10%	712	15.62%	0.66	.999
Pronoun	83	0.86%	92	2.02%	33.70	< .001
Determiner	994	10.36%	424	9.30%	3.80	.874
Adverb	185	1.93%	163	3.58%	35.04	< .001
Conjunction	574	5.98%	233	5.11%	4.33	.826
Preposition	1375	14.33%	674	14.79%	0.54	.999
Tense-aspect	N=538		N=275			
Simple present	423	78.62%	161	58.55%	36.26	< .001
Simple past	64	11.90%	73	26.55%	27.87	.005
Simple future	0	0.00%	0	0.00%	0.00	1
Simple modal	33	6.13%	22	8.00%	1.00	.999
Present perfect	18	3.35%	18	6.55%	4.40	.975
Past perfect	0	0.00%	0	0.00%	0.00	1
Future perfect	0	0.00%	0	0.00%	0.00	1
progressive	0	0.00%	1	0.36%	1.96	.999
Voice	N=1162		N=515			
Active clause	1105	95.09%	457	88.74%	22.58	< .001
Passive clause	57	4.91%	58	11.26%	22.58	< .001

Among functional words, prepositions were the most repetitive part of speech in both types, followed by determiners. For pronouns, it is noticed that they are more frequent in original abstracts, with highly significant differences. The third feature examined in the study is the voice employed, whether active or passive. As shown in Table 9 below, most clauses in both ChatGPT and original abstracts were active, 95.09% and 88.74% respectively, with significant differences. Conversely, passive clauses were more employed in original abstracts, with highly significant differences.

It is no surprise that nouns are the most frequent part of speech in both ChatGPT and original abstracts since nouns are the most used word category in language. This prevalence is evident as Burkette and Kretzschmar (2018) stated that about 50% of the words in English are nouns. Similar findings indicating that nouns are the largest POS in abstracts were reported by multiple studies (Jiand & Hyland, 2017; Kretzenbacher, 1990). Furthermore, the findings showed that this applies for both ChatGPT and original abstracts, indicating that such finding is not specific to one of the two types of texts. Similarly, Guo et al. (2023) found that ChatGPT-generated content contains more nouns than any other POS. The differences between the two types of texts, where ChatGPT texts contain a higher percentage of nouns, was also found in the corpus of Guo et al. (2023).

Moreover, no significant differences in the percentages of verbs and adjectives were found. It can be argued that the contextual limitations of using verbs and adjectives, which are used to convey information or provide descriptions, restricts the potential variations, especially since the context examined is research articles' abstracts where brevity is essential (Nundy et al., 2023) and there is no room for circumlocutions or excessive literary language. The POS distribution illustrated in Table 9 above abides to the conventional features of academic writing (Schleppegrell, 2004).

As for pronouns, significant differences were found in the percentage of pronouns employed between ChatGPT and original abstracts, where pronouns were used more frequently in original abstracts. Guo et al. (2023) also reported that pronouns were more frequent in human-authored content than in ChatGPT-generated content. Conversely, Johansson (2023) found that the essay generated by ChatGPT had more pronouns than the human written one. However, the findings of Johansson (2023) were based on a small-scale data set of only two essays.

An argument can be made that the higher frequency of pronouns in human-authored abstracts is caused by the higher frequency of (key literature finding) and (participants) steps in both Introduction and Methodology moves. This argument is based on the fact that when reporting about the literature and the participants of the study, pronouns need to be used to refer to participants and the researchers of the reviewed studies. Moreover, the employment of some pronouns, such as ‘I’, ‘We’ or ‘Our’, denote the presence of a human author and facilitate a direct line of communication between the author and the reader.

Pronouns can be used by human writers to clearly identify themselves or their research team in abstracts, indicating ownership and accountability for the work. In contrast, ChatGPT lacks a unique identity, ideas, or personal experiences because it is an artificial intelligence language model that lacks individuality and agency (Guo at al., 2023; van Woudenberg et al., 2024). It does not have a personal opinion or claim to the content it creates; instead, it functions based on patterns and information found in its training data. Therefore, pronouns may be used less frequently in ChatGPT-generated abstracts because the information does not have to be attributed to a particular person or group.

Adverb usage is more frequent in human-written abstracts than in abstracts produced by ChatGPT. Guo et al. (2023) reported similar findings. Adverbs of frequency or manner bring some nuance to verbs, adjectives, and other adverbs, providing them with more depth. They are useful for adding more information, describing the way or ferocity of an action, and enhancing the writing's general tone and style.

Adverbs are frequently used by human authors to provide nuance, express confidence levels, or expound on the circumstances surrounding the study or its conclusions. This may include going into more depth on the methods used in the studies, giving a more vivid description of the results, or highlighting particular steps in the research process. Unlike human-written abstracts, ChatGPT does not have a sophisticated grasp of context or personal experiences. Instead, it writes text based on patterns observed from a variety of data sources writing (Roumeliotis and Tselikas, 2023). Consequently, ChatGPT can produce abstracts that are cohesive and informative, but it might not use adverbs as often or as subtly as writing by humans. Adverbs are a natural choice for stylistic and communicative purposes in human-written abstracts because of their rich contextual awareness and subjective nuances.

For better understanding of the use of different parts of speech in ChatGPT and original abstracts and the differences between them, Table 10 above demonstrates the top 10 most frequent nouns, adverbs, adjectives, and pronouns from the ChatGPT and original abstracts.

Table 10

Most frequent nouns, adverbs, adjectives, and pronouns in ChatGPT and original abstracts

POS /No.	Nouns		Pronouns		Adjectives		Adverbs	
	ChatGPT	Original	ChatGPT	Original	ChatGPT	Original	ChatGPT	Original
1	Memory	Memory	Their	We	Social	Social	More	How
2	Study	Effects	Its	Their	Educational	Thermal	Furthermore	More
3	Research	Team	It	Its	Public	Emotional	Additionally	When
4	Findings	Research	They	It	Economic	Cognitive	Moreover	Not
5	Language	Study	Our	Our	comprehensive	Mnemonic	Also	as
6	Intervention	Results	We	I	Thermal	Physical	Significantly	also
7	Implication	Control	There	Who	Significant	General	When	Especially
8	Infrastructure	Task	-	They	Valuable	Negative	Where	Therefore
9	Effects	English	-	There	Diverse	Educational	Particularly	Well
10	Innovation	Symptoms	-	-	Positive	Private	How/Often	Then

There was also a notable difference in the prevalent use of tense between abstracts written by humans and those generated by ChatGPT. The simple past tense, a style used frequently in academic writing, was more common in abstracts authored by humans. Here, human writers often use the past tense to describe processes, events, and results that have already happened, giving their stories a distinct chronological flow. This decision is consistent with the standard procedure, which is to present research events in the past tense in order to create a logical and chronological narrative (Graetz, 1985). However, both abstracts produced by ChatGPT and human-authored showed a clear inclination towards the simple present tense. Similar findings for abstracts were reported by other studies (Candarh, 2012; Chalak & Norouzi, 2013; Cross & Oppenheim, 2006). Due to ChatGPT's lack of contextual awareness and personal temporal viewpoint, it often generates text in the present tense, presenting facts in a more objective and timeless way. The tendency to use present simple tense was also reported by AlAfnan and MohdZuki's (2023) study of GPT-4 model. The default mode of ChatGPT, which conveys information objectively without a specific temporal orientation, may be reflected in the usage of the simple present tense in its outputs. However, Bhatia suggested that tense choices can be affected by the rhetorical moves along with the syntactic requirements (1993, p. 67). Accordingly, it was found that the present tense is mostly used in the Introduction moves and the Conclusion moves, while past tense is employed in the Methodology moves and the Results moves. This finding is consistent with previous studies (Candarh, 2012; Cross & Oppenheim, 2006). The examples below show the use of tense in pairs from ChatGPT and original abstracts, and how there can be a tendency toward using certain tenses in certain moves, with ChatGPT using more present tense.

Example 9 (*Original 2*)

'*The current study attempted to replicate prior findings that suggest upregulation may occur following higher load trials or affective interference during a working memory task. Further, the study examined anxiety, depressive symptomatology, working memory capacity, mood, and dispositional mindfulness and possible moderators for upregulation of cognitive control (***Introduction***). Participants (N = 150) completed a delayed recognition working memory task with mnemonic load (High vs. Low) and affective interference (Negative vs. Neutral) parametrically manipulated. Participants completed measures of the individual difference factors (***Methodology***).*'

Example 10 (*GPT 2*)

*'This study investigates the role of affective interference and mnemonic load in the dynamic adjustment processes of working memory (***Introduction***). ... Using a sample of adult participants (n = 200, mean age = 35 years), this study employed a series of working memory tasks that involved the presentation of emotionally valenced stimuli with varying mnemonic loads. Participants were required to perform a variety of cognitive tasks, including memory encoding, maintenance, and retrieval, while simultaneously managing affective interference (***Methodology***).'*

Example 11 (*Original 1*)

'*Results show that there were significant autoregressive effects for both the happiness and sadness ratings (accounting for 4% of variance). We also observed cross-over effects, such that the happiness rating of an AM was predicted by the sadness rating of the previously reported AM (and vice versa) (***Results***). ... Together, these findings demonstrate that there is a substantive effect of emotional intensity on the output order with which AMs are reported—even when elicited by cue words. Based on the premise that the output order of AMs informs about the organisation of autobiographical memory, our results highlight the role of emotional associations among AMs in old age (***Conclusion***).*

Example 12 (*GPT 1*)

*'Preliminary findings indicated that output order effects significantly influenced the recall of autobiographical memories in older adults. Specifically, memories presented in categorical order were retrieved more efficiently and accurately compared to memories presented in random order. Moreover, memories with a higher emotional valence were recalled more quickly and were associated with greater subjective ratings of emotional intensity and vividness. These results provide further evidence for the emotional organization hypothesis in autobiographical memory among older adults, highlighting the importance of emotional valence in memory retrieval. The findings have implications for understanding the mechanisms underlying age-related changes in autobiographical memory and may inform interventions aimed at enhancing memory performance and emotional well-being in older individuals (***Results***). Overall, this study contributes to the existing literature by elucidating the role of output order effects and emotional organization in autobiographical memory in old age. The findings underscore the significance of emotional valence as a cognitive factor influencing memory retrieval and suggest potential avenues for future research in the field of aging and memory (***Conclusion***).'*

Regarding active and passive voice usage, a distinct difference was found between the abstracts that were written by ChatGPT and those written by humans. Overall, there was a clear preference for the active voice in both sources, indicating a general desire for a direct and straightforward communication style that is frequently preferred in academic writing and in which the subject's activities are explicitly attributed to them for clarity and agency. Candarh (2012) reported similar findings for English abstracts. However, another approach where abstracts are commonly conveyed with passive voice for purposes of the text sounding objective and research oriented is reported by Cross and Oppenheim (2006).

A closer look at table 9, nevertheless, indicated a subtle variation in the voice distribution between the ChatGPT and original abstracts. In the ChatGPT-generated abstracts, active clauses were found to be more common, indicating a default syntactic propensity in the language model to prefer active constructions for clarity and simplicity. ChatGPT-generated abstracts had significantly more use of active voice than original abstracts. Such finding was also reported by AlAfnan and MohdZuki (2023). However, this usage might not always follow the syntactic decisions made by human authors and does not indicate that passive clauses are very rare in ChatGPT abstract. Referring back to example 12 (*GPT 1)*, multiple examples of employing passive voice can be found in reporting the results (e.g., *were retrieved, were recalled, were associated*). However, there was an overall more usage of passive voice in abstracts that were authored by humans than those generated by ChatGPT. Various instances can be found such as, *was predicted,* in example 11 (*Original 1)*, and, *were observed,* in example 6 *(Original 25)*.

This style decision in human-authored content could point to a purposeful tactic writers take to move the emphasis from the action's doer to the action itself. Although the active voice is generally preferred in both sources, the variable distribution of active and passive constructions inside clauses indicates possible differences in grammatical preferences and the complex interaction between ChatGPT's default generation patterns and the intentional choices made by human authors for specific communicative goals. The more important variation is the use of personal pronouns as subjects in active clauses in original abstracts much more than in ChatGPT abstracts. For example, the author in (*Original 6*) states '*I provide unbiased estimates of the model's elasticities, using a regression discontinuity derived from Indian government policy'*, another example was found in (*Original 20*), *'Based on my findings, I illustrate that it is helpful to use a public–private dichotomy framework to understand different types of political* *interactions in Iceland, other small states, and more widely'*. Other examples can also be found of using plural personal pronoun 'we'. For instance, '*To organize this literature, we present an ecological view of collective memory.*' (Original 9).

4.4 Pattern Analysis

Another important goal of the study is to analyze the patterns of the moves as they are used in the collected abstracts. To fulfill this aim, I identified the patterns first by surveying the annotation of the abstracts, then the occurrence and percentage of each pattern was counted. Nine patterns were found in the collected corpus of abstracts. However, only one pattern, Introduction, Methodology, Results, and Conclusion (IMRC), was found in ChatGPT abstracts. On the other hand, all of the nine patterns were identified in the original abstracts. These patterns are a departure from the model suggested by Hyland (2000a), which was a five-move model (IPMPC).

Table 11

Pattern analysis of moves in original and ChatGPT abstracts

No.	Pattern	Original abstracts		ChatGPT abstracts		Example
A	I M R C	7	28%	25	100%	GPT 1 / Original 25
B	I M R	7	28%	0	0%	Original 1
C	I M I R	2	8%	0	0%	Original 18
D	I M C	1	4%	0	0%	Original 7
E	I M	1	4%	0	0%	Original 12
F	I	1	4%	0	0%	Original 5
G	I C	1	4%	0	0%	Original 19
H	I R	3	12%	0	0%	Original 10
I	I R C	2	8%	0	0%	Original 8
SUM		25	100%	25	100%	

Table 11 above shows the frequencies and percentages of each pattern in both ChatGPT and original abstracts respectively. As table 11 shows, all 25 collected abstracts from ChatGPT followed the same four-moves pattern. However, in the original abstracts, there was a variety of move-sequence patterns. As demonstrated in Table 11, the IMRC pattern and the IMR pattern were the most frequent in the collected abstracts. The IMR pattern is very similar to the Purpose, Methodology, Product (PMP) model reported by Bonsu and Adusei (2023), where Purpose is represented to be part of Introduction in the current study's model and Product is represented by the Results Move.

These findings suggest that ChatGPT, as an AI model, has a systematic unchanging pattern according to which the abstracts are based. This pattern (Introduction, Methodology, Results, Conclusion) mimics the whole structure of a research article, which is what abstracts are traditionally used for, to mirror the whole scientific work. The fact that this same pattern is the most repetitive one in original abstracts followed by the same pattern without the Conclusion move, can be used to explain the fixed pattern of ChatGPT. As the IA model is trained on real data from published scientific papers, the regulative pattern affected the ChatGPT behavior in this regard. However, human authors have no compulsion to abide by a certain pattern; instead, they use whatever pattern or style they find suitable. Human authors may not even be aware of the pattern concerns, and they might be writing unconsciously in this regard, solely based on their experience and choices. Several factors motivate them to adjust their writing style; for example, the field of research, the nature of the study, the rules of the journal in which the article is published, and the characteristics of the intended audience. This contextual awareness makes it possible to apply patterns in a more deliberate and flexible way. A solid argument for this suggestion is that four patterns were repeated only once, indicating a wide range of choices made by the authors of original abstracts. Furthermore, as the rhetorical moves pattern analysis above has shown, only two or three moves are deployed in many original abstracts, meaning that they neglect summarizing a certain aspect of the study. On the other hand, ChatGPT might opt for providing more details when generating content as the enquiry only asks for an abstract about the title, thus the AI application might be trying to cover as much of a wholesome aspect of the supposed study, covering as many aspects as possible.

5. Conclusion

5.1 Summary of Findings

In light of the previous background on genre analysis and AI language models, this study came as an attempt to examine and analyze the genre of some texts generated by the well-known AI language model (ChatGPT). As abstracts are a fundamental part of any publication, the researcher chose to direct the concern of the current study on analyzing abstracts generated by ChatGPT. Subsequently, comparing original human-written abstracts with ChatGPT abstracts was the purpose of this study. The rationale behind the study was the scarcity of studies on the genre and style of the academic texts generated by ChatGPT. Moreover, the steps or methods utilized by the chatbot have not received sufficient research. Therefore, this study attempted to help in filling these gaps by addressing the following questions:

1. How does the genre of abstracts differ between published scientific articles and ChatGPT-generated abstracts, in terms of the moves used?
2. What are the linguistic features used to achieve the moves in linguistic scientific article abstracts from ChatGPT-generated abstracts?
3. To what extent do the abstracts generated by ChatGPT mimic the genre and linguistic features of linguistic scientific article abstracts?

Based on the analysis of the corpus of 50 abstracts extracted from Q1 journals and generated by ChatGPT, the three main aims of the study were addressed. Similarly to the findings of Guo et al. (2023), it was found that ChatGPT abstracts were longer, more complex, and more lexically dense than original abstracts in terms of general text statistics.

Different rhetorical patterns were revealed by the genre analysis, which focuses on steps and moves. Four moves were found to be utilized, which are 'Introduction', 'Methodology', 'Results', and 'Conclusion' (IMRC). Those moves are employed using 21 various steps. It was revealed that the Results move was the most employed move in the ChatGPT-generated abstracts, and the least employed move was Methodology. On the other hand, the Introduction move was the most used by the authors in the original abstracts, while the Results move was the second most used. Different from ChatGPT-generated abstracts, original abstracts have less frequent Conclusion moves. Findings related to original abstracts echo the results of multiple previous studies (Al-Khasawneh, 2017; Bonsu & Adusie, 2023; Candarh, 2012; Hyland, 2000a).

It was found that more information was usually included in ChatGPT abstracts, especially in the introduction section. The findings also revealed that different introduction moves use different steps, with ChatGPT omitting the hypothesis step. The study highlights the importance of having a thorough understanding of ChatGPT language model's capabilities and constraints, taking into account how training data and user interaction dynamics may affect the produced texts. The results indicate that ChatGPT's abstract generation systematically deviate from some human-authored patterns, especially when it comes to how hypotheses are presented and some aspects of the Introduction and Conclusion. A thorough understanding of Large Language Model (LLM) outputs and the effect of their training datasets is emphasized in the discussion of the findings. Furthermore, the results of the study found that the Methodology move was used more in original abstracts than in ChatGPT abstracts. This is because the methodology descriptions frequently require a detailed breakdown of procedures, which may be more difficult for ChatGPT because of its architecture. If the model is meant to generate more simple material, such as results, it may have trouble representing complex procedural information. A more detailed explanation could be that ChatGPT abstracts are not based on an actual study with an actual data collection process, making ChatGPT briefly discuss such details.

In the Results move, the findings step was the most frequent step for both ChatGPT and original abstracts, and purpose was the least repetitive one for both. Moreover, only one instance of significance of the findings step was found in the original abstracts' corpus. It was also shown that there were no significant differences in the steps of the Results move. However, while more analysis was described in original abstracts, more implications were suggested by ChatGPT. These findings might imply that there is a widespread focus on clearly outlining the key findings of the study in the abstracts' Results move. Similar findings were reported by Buruk et al. (2023) for providing findings and analysis, and by Gupta (2022) for ChatGPT's ability to suggest implications.

Analyzing the move patterns as they appear in the gathered abstracts is a key component of the study. In order to achieve this goal, the patterns were first found, and then each pattern's frequency and percentage were determined. The gathered corpus of abstracts revealed nine patterns. Nevertheless, ChatGPT abstracts only contained one pattern: Introduction, Methodology, Results, and Conclusion (IMRC). However, the original abstracts contained identifications for each of the twelve patterns. The frequencies and percentages of each pattern in the original abstracts and ChatGPT were displayed in Table 11. Table 11 demonstrates that the 25 collected abstracts from ChatGPT all adhered to the same four-step sequence. However, in original abstracts there were some instances of a five-move pattern, in which a certain move is used and then repeats after another one. As demonstrated in Table 11, the IMRC pattern was the most frequent one, along with the Introduction, Methodology, Results (IMR) pattern.

As for the grammatical features, content words were found to be used much more significantly than function words. Nouns were the most used part of speech in both types of texts, which is unsurprising since nouns are the most commonly used word category in language. Among verbs and adjectives, no significant differences were found between original and ChatGPT abstracts. It can be argued that the contextual limitations of using verbs and adjectives, which are used to convey information or provide descriptions, limits the possible differences that can exist, especially since the context examined is research articles' abstracts where brevity is essential and there is no room for circumlocutions or excessive literary language.

Based on the findings of the study and considering the examination of ChatGPT-generated abstracts and their comparison to original abstracts, some important observations can be summarized about the understanding and credibility of ChatGPT representing AI language generated academic content. First, ChatGPT follows a fixed pattern/order/style when producing content as shown in the moves pattern. Second, it mostly offers lengthier content. Third, ChatGPT tends to offer generic findings and too many implications when compared to human-authored content. It may make use of its huge and diverse database to come up with such implications. Similarly, and likely for the same reasons, it also pays more attention to generating conclusions than offering details about data collection and analysis details, as it does not actually perform any data collection. Finally, it focuses more on providing general background and study aims in the introduction than on reviewing literature or offering the hypothesis of the study. All in all, these summarized points require more investigation from different angles to be able to confidently generalize them. However, many factors, the most important of which is constant updates, can affect the above-mentioned conclusions.

5.2 Limitations

Because the sample of abstracts in this study was relatively small (N=50), the findings cannot be generalized to all ChatGPT-generated abstracts. Also, a second limitation is that AI models are constantly and rapidly developing. This rapid evolution may make the results of the current study become outdated in a very short period or even by the time this work sees light. This shortcoming could have been made less crucial if GPT-4, the latest version of ChatGPT as of early 2024, was used instead of using the 3.5 version. The current study comes with another limitation that the abstracts generated by ChatGPT were not conditioned by any word limits, unlike the human-authored ones. The reason why the researcher did not use any word limits for ChatGPT is to leave it unrestricted so that it would produce its ideal content. However, this remains a limitation of the current study as the human-authored abstracts were put under various restrictions based on publishing journals standards.

5.3 Recommendations

Based on the results of this study, the researcher recommends future researchers to investigate the following research issues related to comparing other parts of research articles, such as introduction or methodology between original human-written ones and ChatGPT generated parts. Such studies would aid in reaching a wider understanding of the differences and similarities between AI and the human sources of academic writing. The current study can be the way for researchers to discuss those matters as it is based on comparing abstracts, which are mirrors of their whole studies. Moreover, another kind of comparison that can be recommended is to compare abstracts generated by ChatGPT with abstracts generated by other AI language models. The aim of a similar investigation would be to check how the language model training biases can affect AI responses. A further recommendation would be to take the current study's limitations, namely word limit of ChatGPT generated abstracts, into consideration.

Appendix A

Python code used for inter-rater reliability.

URL link to the Python code for checking inter-coder reliability

https://github.com/AbdQ96/Inter-rater-Reliability-Code/blob/main/inter_rater_reliability%202.py

Note: To be able to run the code, the following libraries must be installed:

pip install pandas==2.1.4

pip install scikit-learn==1.3.2

pip install openpyxl==3.1.2

Appendix B

Corpus of the study

Corpus of abstracts

https://drive.google.com/drive/folders/1LtJMiGEeInWCx0Na5SX0QCIoOO0MJmmR?usp=drive_link

List of Titles

No.	Article Title	Journal Name
.1	Output Order Effects in Autobiographical Memory in Old Age: Further Evidence for An Emotional Organization.	*Memory & Cognition*
2	The Role of Affective Interference and Mnemonic Load in The Dynamic Adjustment in Working Memory	
3	Depressive Symptoms are Positively Linked to Malevolent Creativity: A Novel Perspective on The Maladaptive Nature of Revenge Ideation	*The Journal of Creative Behavior*
4	New Deal, New Patriots: How 1930s Government Spending Boosted Patriotism During World War II	*The Quarterly Journal of Economics*
5	Regulation of Ribosomal RNA Gene Copy Number, Transcription and Nucleolus Organization in Eukaryotes	*Nature Reviews Molecular Cell Biology*
6	Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India	*Journal of Political Economy*
7	Synchrotron Radiation Techniques and Their Application to Actinide Materials.	*Reviews of Modern Physics*
8	Measuring Interfacial Dzyaloshinskii-Moriya Interaction in Ultrathin Magnetic Films	
9	Ecologies of Memories: Memory Work Within and between Organizations and Communities	*Academy of Management*
10	Monetary Stimulus amidst the Infrastructure Investment Spree: Evidence from China's Loan-Level Data	*Journal of Finance*

11	Feasibility Study of Operation Volume Design Concept for Integration of Aircraft Flying Under Visual Flight Rules Based on Actual Flights	*IEEE*
12	Convective Thermal Metamaterials: Exploring High-Efficiency, Directional, and Wave-Like Heat Transfer	*Advanced Materials*
13	Permian to Triassic Protolith Ages of Type Locality Eclogites in the Eastern Alps: Implications for the Opening of the Meliata Back-Arc Basin	*Geology*
14	Memory Based Hybrid Crow Search Algorithm for Solving Numerical and Constrained Global Optimization Problems	*Artificial Intelligence Review*
15	Effects of Word Definitions on Meaning Recall: A Multisite Intervention in Language-Diverse Second Language English Classrooms	*Language*
16	Greeting in English as a Foreign Language: A Problem for Speakers of Chinese	*Applied Linguistics*
17	Origin versus Substance: Competing Determinants of Disruption in Duplication Technologies	*Journal of Consumer Research*
18	Untangling the Team Social Capital–Team Innovation Link: The Role of Proportional Task Conflict as well as Group- and Differentiated Individual-Focused Transformational Leadership	*Human Relations*
19	Cuttlefish eye–inspired artificial vision for high-quality imaging under uneven illumination conditions	*Science Robotics*
20	Politicians and the General Public Communicating on Facebook and Messenger: Public and Private Interactions in a Two-Level Online Sphere	*Social Media and Society*
21	Why Do People Legitimize and Cooperate with the Police? Results of a Randomized Control Trial on the Effects of Procedural Justice in Quito, Ecuador	*Crime Science*
22	Racisms of Commission and Omission in Educational Psychology: A Historical Analysis and Systematic Review	*Educational Psychologist*

23	The Divergent Adolescent and Adult Transitions of Latin American Undocumented Minors	*Child Development Perspective*
24	Effectiveness of Physical Activity Interventions for Improving Depression, Anxiety and Distress: An Overview of Systematic Reviews	*Sport Medicine*
25	The Effect of upland crop planting on field N2O emission from rice-growing seasons: A case study comparing rice-wheat and rice-rapeseed rotations	*Agriculture Ecosystem and Environment*

References

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology*, *3*(3), 85-94. https://doi.org/10.37965/jait.2023.0267

Al-Ali, M. N. (2004). How to get yourself on the door of a job: A cross-cultural contrastive study of Arabic and English job application letters. *Journal of Multilingual and Multicultural Development*, *25*(1), 1-23. http://dx.doi.org/10.1080/01434630408666517

Alhadlaq, H., Chen, W., & Huang, Y. (2021). Application of artificial intelligence in education: A review. *IEEE Access*, 9, 15129-15140. https://doi.org/10.1109/ACCESS.2020.2988510

Al-Khasawneh, F. M. (2017). A genre analysis of research article abstracts written by native and non-native speakers of English. *Journal of Applied Linguistics and Language Research*, *4*(1), 1-13.

Allam, H., Dempere, J., Akre, V., Parakash, D., Mazher, N., & Ahamed, J. (2023, May). Artificial intelligence in education: an argument of Chat-GPT use in education. In *2023 9th International Conference on Information Technology Trends (ITT)* (pp. 151-156). IEEE. https://doi.org/10.1109/ITT59889.2023.10184267

M Alshater, M. (2022). Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT. *Available at SSRN 4312358*. http://dx.doi.org/10.2139/ssrn.4312358

Anderson, N., Belavy, D. L., Perle, S. M., Hendricks, S., Hespanhol, L., Verhagen, E., & Memon, A. R. (2023). AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in Sports & Exercise Medicine manuscript generation. *BMJ Open Sport—Exercise Medicine*, *9*(1). https://doi.org/10.1136/bmjsem-2023-001568

Askehave, I., & Nielsen, A. E. (2005). Digital genres: a challenge to traditional genre theory. *Information Technology & People*, *18*(2), 120-141. https://doi.org/10.1108/09593840510601504

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. *Genre and the NewRhetoric*, 79101.

Behnam, B., & Golpour, F. (2014). A genre analysis of English and Iranian research articles abstracts in applied linguistics and mathematics. *International Journal of Applied Linguistics and English Literature*, *3*(5), 173-179. https://doi.org/10.7575/aiac.ijalel.v.3n.5p.173

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics, 6,* 587-604. https://doi.org/10.1162/tacl_a_00041

Bhatia, V. K. (1993). *Analysing genre: Language use in practical settings*. Routledge.

Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. A&C Black.

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., & Urzua, A. (2004). *Representing language use in the university: Analysis of the TOEFFL 2000 spoken and written academic language corpus*. ETS TOEFL *monograph series*, *MS-25*. Princeton, NJ: Educational Testing Service..

Biber, D. and Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Bonsu, E. M., & Adusei, D. K. (2023). Trending Technologies: A Corpus-based Genre Analysis of Abstracts on ChatGPT. *Linguistics Initiative, 3*(2), 107-124. https://doi.org/10.53696/27753719.3283

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. *Neural Information Processing Systems, 33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Burkette, A., & Kretzschmar Jr, W. A. (2018). *Exploring linguistic science: Language use, complexity, and interaction*. Cambridge University Press.

Buruk, O. O. (2023, October). Academic writing with GPT-3.5 (ChatGPT): reflections on practices, efficacy and transparency. In *Proceedings of the 26th International Academic Mindtrek Conference* (pp. 144-153). https://doi.org/10.1145/3616961.3616992

Cai, Z. G., Haslett, D. A., Duan, X., Wang, S., & Pickering, M. J. (2023). Does ChatGPT resemble humans in language use? *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2303.08014

Candarh, D. (2012). A cross-cultural investigation of English and Turkish research article abstracts in Educational Sciences. *Studies about Language*, *2*, 12-17. https://doi.org/10.5755/j01.sal.0.20.1770

Casal, J. E., & Kessler, M. (2024). Rhetorical move-step analysis. In M. Kessler & C. Polio (Eds.), *Conducting genre-based research in applied linguistics: A methodological guide* (pp. 82–

104). Routledge.

Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in language teaching: A guide for language teachers*. Cambridge University Press.

Chalak, A., & Norouzi, Z. (2013). Rhetorical moves and verb tense in abstracts: A comparative analysis of American and Iranian academic writing. *International Journal of Language Studies, 7*(4).

Chu, S. K. W. (2021). An AI-based Chatbot for Student Support in Higher Education. *Educational Technology Research and Development*, *69*(3), 1297-1316.

Cohen, N., & Ellis, C. (2021). AI Writing Feedback: A Review and Future Directions. *Journal of Writing Research*, *13*(1), 43-71.

Connor, U., Davis, K. W., & De Rycker, T. (1995). Correctness and clarity in applying for overseas jobs: A cross-cultural analysis of US and Flemish applications. *Text & Talk*, *15*(4), 457-476. https://doi.org/10.1515/text.1.1995.15.4.457

Copeland, B. (2023). artificial intelligence. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Cross, C., & Oppenheim, C. (2006). A genre analysis of scientific abstracts. *Journal of documentation*, 62(4), 428-446. https://doi.org/10.1108/00220410610700953

Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of sport, 40*(2), 615-622. https://doi.org/10.5114/biolsport.2023.125623

Drury, A., Pape, E., Dowling, M., Miguel, S., Fernández-Ortega, P., Papadopoulou, C., & Kotronoulas, G. (2023, April). How to write a comprehensive and informative research abstract. *In Seminars in Oncology Nursing* (Vol. 39, No. 2, p. 151395). WB Saunders. https://doi.org/10.1016/j.soncn.2023.151395

Dudley-Evans, A. (1986). Genre analysis: An investigation of the introduction and discussion sections of MSc. dissertations. In M. Coulthard (Ed.), *Talking about text* (Discourse Monograph No. 13). English Language Research, University of Birmingham, 128- 145.

Dudley-Evans, T. (1989). An outline of the value of genre analysis in LSP work. *Special language: From humans thinking to thinking machines*, 72-79.

Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. London: Printer Publishers, Ltd.

El-Dakhs, D. A. S. (2018). Comparative genre analysis of research article abstracts in more and less prestigious journals: Linguistics journals in focus. *Research in Language (RiL)*, *16*(1), 47-63. http://dx.doi.org/10.2478/rela-2018-0002

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching, 12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

Frow, J. (2005). *Genre*. London: Routledge.

Graetz, N. (1985). Teaching EFL Students to Extract Structural Information from Abstracts. In *Reading for Professional Purposes: Methods and Materials in Teaching Languages* (pp. 123-135). ACCO.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2301.07597

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language* *and Meaning*. London: Edward Arnold.

Harris, Z. S. H. (1952). *Discourse Analysis*. Linguistic Society of America.

Haroon, J., & Azam, M. K. S. (2023). AI generated language: Prospects and challenges. *Erevna: Journal of Linguistics and Literature*, *7*(1), 47-60.

Helms-Park, R., & Stapleton, P. (2003). Questioning the importance of individualized voice in undergraduate L2 argumentative writing: An empirical study with pedagogical implications. *Journal of Second Language Writing, 12*(3), 245–265. https://doi.org/10.1016/j.jslw.2003.08.001

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports, 13*(1), 18617. https://doi.org/10.1038/s41598-023-45644-9

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research, 8*(1), 5-29. https://doi.org/10.1191/1362168804lr132oa

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

Hyland, K. (2000b). Hedges, boosters and lexical invisibility: Noticing modifiers in academic texts. *Language Awareness*, *9*(4), 179-197. https://doi.org/10.1080/09658410008667145

Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.

Jenkins, S., & Hinds, J. (1987). Business letter writing: English, French, and Japanese. *TESOL Quarterly*, *21*(2), 327-349. https://doi.org/10.2307/3586738

Jiang, F.K., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes, 46*, 1-14.

https://doi.org/10.1016/j.esp.2016.11.001

Jiang, F., & Hyland, K. (2023). Changes in research abstracts: Past tense, third person, passive, and negatives. *Written Communication, 40*(1), 210-237. https://doi.org/10.1177/07410883221128876

Jogthong, C. (2001). *Research article introductions in Thai: Genre analysis of academic writing*. West Virginia University.

Johansson, I.-R. (2023). A Tale of Two Texts, a Robot, and Authorship : A Comparison Between a Human-Written and a ChatGPT-Generated Text (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-60674

Kalyan, K. S. (2023). A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal*, 100048. https://doi.org/10.1016/j.nlp.2023.100048

Kaplan, R. B., Cantor, S., Hagstrom, C., Kamhi-Stein, L. D., Shiotani, Y., & Zimmerman, C. B. (1994). On abstract writing. *Text-interdisciplinary Journal for the Study of Discourse, 14*(3), 401-426. https://doi.org/10.1515/text.1.1994.14.3.401

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Kretzenbacher, H. L. (1990). *Rekapitulation: Textstrategien der Zusammenfassung von wissenschaftlichen Fachtexten* (Vol. 11). Gunter Narr Verlag.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174. https://doi.org/10.2307/2529310

Liu, D. (2015). Moves and wrap-up sentences in Chinese students' essay conclusions. *Sage*

Open, *5*(2), 2158244015592681. https://doi.org/10.1177/2158244015592681

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. *Library Hi Tech News*, *40*(3), 26-29. https://dx.doi.org/10.2139/ssrn.4333415

Maier, P. (1992). Politeness strategies in business letters by native and non-native English speakers. *English for Specific Purposes*, *11*(3), 189-205. https://doi.org/10.1016/S0889-4906(05)80009-2

Malyuga, E., & McCarthy, M. (2018). English and Russian vague category markers in business discourse: Linguistic identity aspects. *Journal of pragmatics*, *135*, 39-52. https://doi.org/10.1016/J.PRAGMA.2018.07.011

Makarova, V., Li, Z., & Wang, Z. (2024). Can ChatGPT Grade Non-Native Academic English Writing?. In O. Eybers & A. Muller (Eds.), *AI Approaches to Literacy in Higher Education* (pp. 97-116). IGI Global. https://doi.org/10.4018/979-8-3693-1054-0.ch005

Martin, J. R. (1992). Genre and literacy-modeling context in educational linguistics. *Annual Review of Applied Linguistics*, *13*, 141-172. https://doi.org/10.1017/S0267190500002440

Martin, P. A. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes, 22*, 25-43. https://doi.org/10.1016/S0889-4906(01)00033-3

Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, *2*, 1-11. https://doi.org/10.1016/j.amper.2014.12.002

Mehrpour, S., & Mehrzad, M. (2013). A Comparative Genre Analysis of English Business Emails Written by Iranians and Native English Speakers. *Theory & Practice in Language Studies*, *3*(12). https://doi.org/10.4304/TPLS.3.12.2250-2261

Miller, C. R. (1994). Rhetorical community: The cultural basis of genre. *Genre and the New Rhetoric*, 6778.

Mirador, J. F. (2000). A move analysis of written feedback in higher education. *RELC Journal*, *31*(1), 45-60. http://dx.doi.org/10.1177/003368820003100103

Moshtaghi, P. (2010). The comparative genre analysis of psychology journal articles and popularized psychology texts in e-magazines and e-journals. *Procedia-Social and Behavioral Sciences*, *5*, 2067-2071. https://doi.org/10.1016/j.sbspro.2010.07.414

Nundy, S., Kakar, A., Bhutta, Z. A., Nundy, S., Kakar, A., & Bhutta, Z. A. (2022). How to Write an Abstract?. *How to Practice Academic Medicine and Publish from Developing Countries? A Practical Guide*, 179-184. https://doi.org/10.1007/978-981-16-5248-6_15

O'Donnell, M. (2008). UAM CorpusTool. *Universidad Autónoma de Madrid. Retrieved from www.wagsoft.com/CorpusTool.*

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, *36*, 1-14. https://doi.org/10.1016/j.jeap.2018.08.002

PLOS Medicine. (2024, May 8). *Submission Guidelines.*
https://journals.plos.org/plosmedicine/s/submission-guidelines

Porter, J. E. (1986). Intertextuality and the discourse community. *Rhetoric Review*, *5*(1), 34-47. https://doi.org/10.1080/07350198609359131

Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes*, *18*(2), 139-160. https://doi.org/10.1016/S0889-4906(98)00001-5

Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*, *15*(6), 192. https://doi.org/10.3390/fi15060192

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes*, *22*(4), 365-385. https://doi.org/10.1016/S0889-4906(02)00026-1

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Routledge.

Semrl, N., Feigl, S., Taumberger, N., Bracic, T., Fluhr, H., Blockeel, C., & Kollmann, M. (2023). AI language models in human reproduction research: exploring ChatGPT's potential to assist academic writing. *Human Reproduction, 38*(12), 2281-2288.

Soler, V. (2011). Comparative and contrastive observations on scientific titles written in English and Spanish. *English for specific purposes*, *30*(2), 124-137.

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

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Swales, J. M., & Najjar, H. (1981). The Writing of Research Article Introductions. *Written Communication*, *4*(2), 175-191. http://dx.doi.org/10.1177/0741088387004002004

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature, 614*(7947), 224-226. https://doi.org/10.1038/d41586-023-00288-7

van Woudenberg, R., Ranalli, C., & Bracker, D. (2024). Authorship and ChatGPT: a Conservative View. *Philosophy & Technology, 37*(1), 34.

https://doi.org/10.1007/s13347-024-00715-1

Widdowson, H. G. (2007). *Discourse analysis*. Oxford: Oxford University Press.

Yates, J., & Orlikowski, W. J. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review*, *17*(2), 299-32. https://doi.org/10.2307/258774

Zhang, P. (2023). Review of Sydney School Genre Studies. *Linguistics and Literature Studies*, *11*(4), 57–63. https://doi.org/10.13189/lls.2023.110401