A Comparison of Rhetorical Move Analysis by GPT-4 and Humans in Abstracts of Scopus-Indexed Tourism Research Articles

Article *in* International Linguistics Research · June 2024

DOI: 10.30560/ilr.v7n2p1

CITATION
1

READS
380

4 authors:

Hui Geng
Guilin University of Technology
40 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Vahid Nimehchisalem
University of Economics and Human Sciences in Warsaw
179 PUBLICATIONS **1,442** CITATIONS
SEE PROFILE

Mohsen Zargar
2 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Jayakaran Mukundan
Taylor's University
138 PUBLICATIONS **1,848** CITATIONS
SEE PROFILE

International Linguistics Research; Vol. 7, No. 2; 2024
ISSN 2576-2974 E-ISSN 2576-2982
https://doi.org/10.30560/ilr.v7n2p1

A Comparison of Rhetorical Move Analysis by GPT-4 and Humans in Abstracts of Scopus-Indexed Tourism Research Articles

Hui Geng[1], Vahid Nimehchisalem[1], Mohsen Zargar[1] & Jayakaran Mukundan[2]

[1] English Department, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Malaysia

[2] School of Education, Faculty of Social Sciences & Leisure Management, Taylor's University, Malaysia

Correspondence: Vahid Nimehchisalem, English Department, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Malaysia. Tel: 60-176-678-715. E-mail: vahid@upm.edu.my

Received: May 24, 2024 Accepted: June 3, 2024 Online Published: June 5, 2024

Abstract

AI advancements have made ChatGPT a remarkable and versatile tool in education and linguistics, showcasing its potential to mimic human conversation and comprehend language. Scholars are intrigued by ChatGPT's text data handling, yet its application in rhetorical move analysis remains largely unexplored. Therefore, the objective of this study is to investigate the ability of GPT-4 in the identification of rhetorical moves employed in the abstracts of tourism research articles indexed in Scopus. The essentiality of moves was also reported. Additionally, this research seeks to compare the accuracy of GPT-4's analysis with that of humans. Adopting Hyland's (2000) five-move model, the results indicated that GPT-4 analyzes moves more quickly but less accurately than human experts, and the four principal types of errors committed by GPT-4 include redundancy/over-count, unmatched categorization, incorrect sequence, and vague identification. The findings also revealed that Move 2 (Purpose) and Move 4 (Findings) are obligatory with a 100% essentiality rate through both GPT-4 and human analysis. Differences arise in certain steps of Move 1 (Introduction), Move 3 (Methods), and Move 5 (Conclusion), where GPT-4 often sees higher essentiality rates. This study shed light on the testament to AI's current capabilities in move analysis in academic discourse.

Keywords: ChatGPT, rhetorical moves, abstracts, tourism research articles, Scopus-indexed journal

1. Introduction

In recent years, there has been wide interest in genre analysis of abstract writing. An abstract serves as a representative of the research by providing a concise yet precise synopsis of the article. The abstract functions as a guide to the article, stimulating readers to proceed with the reading. In the case of journals, the acceptance or rejection of a paper is typically determined by the quality of the abstract. For research articles to be published in prestigious journals indexed by Scopus, a concise and informative abstract is essential (Geng & Wei, 2023). Previous studies have investigated the rhetorical structure of research article abstracts across multiple soft science disciplines, including education (Sidek, 2017), literature (Tankó, 2017), applied linguistics (Fauzan et al., 2020), psychology (Gusmana, 2023), history (Saidi & Karami, 2024), geography (Alyousef, 2023), economics (Fitria, 2022), among others. However, within the tourism discipline, there have been few studies on the rhetorical move structures of research article abstracts, particularly those indexed by Scopus (Kurniawan & Sabila, 2021).

Furthermore, in most move analysis studies, researchers commonly conducted move identification manually, a process that consumes a significant amount of time. With the advent and the evolution of artificial intelligence, ChatGPT which employs natural language processing has shown potential as an effective aid in the academic community. At present, several studies are looking into the suitability and effectiveness of ChatGPT in educational settings. An illustration of this can be seen in the research conducted by Adeshola and Adepoju (2023), which utilized sentiment analysis to conclude that the majority of individuals express positive opinions regarding ChatGPT. Another example is that Livberber (2023) investigated the utility of ChatGPT (GPT-3.5 and GPT-4) as a writing tool and found that it has the potential to logically and appropriately specify the steps to follow in the design of an academic article. Nevertheless, scant attention has been paid to ChatGPT's capability to conduct textual analysis, which is a research method used by scholars to interpret, analyze, and understand the content, structure, and various elements of the text. So far only one study has shown that GPT-4 could conduct conceptual metaphors and thematic analysis of short stories (Geng & Nimehchisalem, 2023).

Thus, is there a possibility of employing AI to conduct other various types of textual analysis including discourse, sentiment, and stylistic analysis, as well as for linguistic analysis encompassing lexical, semantic, pragmatic, genre, frame, and intertextuality analysis? Given that rhetorical move analysis can be considered a part of genre analysis, which focuses on understanding how texts achieve specific communicative objectives by examining their structural and strategic components, it is worthwhile to leverage the AI-powered tool ChatGPT for experiments to assess its feasibility to accurately identify these rhetorical moves. Based on this, the current study aims to fill the gap by investigating the ability of GPT-4 to identify rhetorical moves in abstracts of tourism research articles from a Scopus-indexed journal. The findings of this research not only underscore the remarkable prospects of exploiting artificial intelligence, especially ChatGPT, for improving academic writing as well as research methods but also open new possibilities for genre analysis.

2. Literature Review

In academic writing, the abstract is recognized by authors as a concise and accurate summary of the content of a research article, marking its unique place as a specific genre within scholarly prose. According to Swales & Feak (2004), it is a crucial navigational tool, allowing readers to quickly ascertain the relevance and applicability of the research to their work. In the digital age, the abstract's role has expanded beyond the article itself to include its visibility in electronic databases and search engines. Biber et al. (2007) noted that well-crafted abstracts improve the article's discoverability, significantly affecting its readership and citation rates. Keywords and phrases within the abstract enhance the article's search engine optimization (SEO), making it more accessible to a global audience. Given the pivotal role of abstracts in deciding the fate of manuscripts, several frameworks have been developed to analyze their structural composition (Bhatia, 1993; Santos, 1996; Swales, 1990), among which Hyland's (2000) five-move model is notably comprehensive, covering introduction, purpose, method, results, and conclusion sections (Ghasempour & Farnia, 2017).

These models suggest universal guidelines for organizing abstracts across various fields. However, the specific rhetorical structures of tourism abstracts in Scopus-indexed journals have not been extensively explored. Founded in November 2004, Scopus is the largest abstract and citation database, exceeding the Web of Science Core Collection in journal coverage. It provides advanced tools for research tracking, analysis, and visualization, establishing it as a highly reputable bibliometric database (Joshi, 2016). As mentioned by Geng (2024), Scopus-indexed journals have gained much recognition on various occasions. For example, many universities in Malaysia as well as in the world hope their academicians publish in Scopus-indexed, ISI (now Web of Science), and other high-impact factor journals. As a result, a detailed analysis of how the rhetorical moves were used in Scopus-indexed tourism research articles is worth further exploring.

Rhetorical moves, as identified by Ding (2007), are schematic units that denote textual regularities, socially recognized for their formal structure, which fulfills a coherent communicative function in written discourse (Swales, 2004). A move is seen as a chunk of text, encompassing at least one complete sentence with a specific communicative function, indicating that any sentence or group of sentences meeting the criteria of a move or step in the coding system was considered an instance of that move or step, irrespective of length (Geng et al., 2023; Zhang & Wannaruk, 2016).

In the coding phase, moves or steps are counted as single instances if uninterrupted, but as separate instances if interrupted by different moves or steps (Geng et al., 2023). Various assisted annotation software tools for coding rhetorical moves have been examined, such as Mover, Research Writing Tutor, Academic Writing Analytics (AWA), and AcaWriter tools (Knight et al., 2020). However, to the researchers' knowledge, there has been no study to date that utilizes ChatGPT (GPT-4) for the coding of rhetorical moves within academic texts. As a type of Large Language Model (LLM) and a prominent framework for generative artificial intelligence (AI), The Generative Pre-trained Transformer (GPT), commonly referred to as ChatGPT, has attracted significant public interest after it reached version 3.5 (Reiss, 2023). GPT-4, unveiled by OpenAI on March 14, 2023, is a recent version of OpenAI's GPT series. It is an enhanced version of GPT-3.5 with the superior capacity to identify nuances and deliver more accurate and logical responses (Bubeck et al., 2023).

As tested by researchers, Savelka et al. (2023) claimed that GPT-4 can analyze textual data in the context of a task involving the interpretation of legal concepts at a level comparable to that of law student annotators who have received adequate training. Another study reported that GPT-4 can assist in qualitative data analysis, and careful prompt structuring can align AI responses closely with human interpretations, particularly in deductive analysis (Siiman, 2023). The previous findings suggested that there is a need to explore more about GPT-4's ability to analyze different types of textual data. Hence, the current study aims to fill this gap in how GPT-4 can assist humans in genre analysis, particularly in move analysis.

2.1 Research Objectives

1. To investigate the ability of GPT-4 to identify rhetorical moves in abstracts of tourism research articles from a Scopus-indexed journal.
2. To explore the essentiality of the moves and steps in abstracts of tourism research articles from a Scopus-indexed journal through analysis by GPT-4 and humans.
3. To compare the accuracy of GPT-4's identification and analysis of rhetorical moves with that of human coders in abstracts of tourism research articles from a Scopus-indexed journal.

2.2 Research Questions

1. Can GPT-4 identify rhetorical moves in abstracts of tourism research articles from a Scopus-indexed journal?
2. What is the essentiality of the moves and steps in abstracts of tourism research articles from a Scopus-indexed journal through analysis by GPT-4 and humans?
3. To what extent does the accuracy of GPT-4's identification and analysis of rhetorical moves compare with that of human coders in abstracts of tourism research articles from a Scopus-indexed journal?

3. Methods

3.1 Corpus Construction

In this study, 20 research article abstracts were chosen randomly from one esteemed tourism journal named the *European Journal of Tourism Research*. These abstracts were in English, and they were extracted from empirical research articles following Swales' IMRD (Introduction, Method, Results, Discussion) framework. The articles all underwent peer review processes before being published between 2022 and 2023. The corpus comprised only standardized academic research papers with excellent citation metrics. To enable comprehensive examination by GPT-4 and language specialists, the accessible abstracts were collated into a single Word file. Each abstract encompassed between 94 to 210 words in length. Collectively, the 20 research article abstracts amounted to 2,929 words, with an average of approximately 146 words per abstract.

3.2 Analytical Framework

The current research analyzed the use of rhetorical moves in abstracts of 20 research articles by employing Hyland's (2000) well-known five-move model (See Table 1). The selection of this innovative framework, which has been developed through a thorough review of 800 abstracts, is based on the fact that it can offer a more detailed analysis. It is greatly known for the interpretation of abstracts with high accuracy.

Table 1. Hyland's (2000) Five-Move Model for the Analysis of Abstracts

Moves	Steps	Labels
Move 1 Introduction (I)	Step 1	Arguing for topic significance
	Step 2	Making topic generalization
	Step 3	Defining key term(s)
	Step 4	Identifying gap
Move 2 Purpose (P)		Stating the research purpose
Move 3 Methods (M)	Step 1	Describing participant/data sources
	Step 2	Describing instrument (s)
	Step 3	Describing procedure and context
Move 4 Findings (F)		Describing the main results
Move 5 Conclusion (C)	Step 1	Deducing conclusion
	Step 2	Evaluating the significance of the research
	Step 3	Stating limitation
	Step 4	Presenting recommendation and implication

Table 1 shows that this model has five moves, each of which is aimed at enhancing coherence and clarity in the abstract. The initial one is the Introduction (I) move. It sets the article's context and themes, outlining the research, background, and rationale to establish a foundation and attract interest, enabling subsequent moves to be understood. The second one is the Purpose (P) move, which provides an accurate description of what was meant to be achieved from this study. It reveals both the writer's intentions and general goals behind their work, giving readers a glimpse of what they want to accomplish or prove with their research. The third move Methods (M) presents research design and procedural elements. It describes the approaches used along with tools and techniques to understand how processes work, as well as possibly replicate a study. As a further move, the abstract uses Findings (F) to present its main findings or results of the article. It defines the empirical or analytical outcomes that are obtained from using certain methods. As the last move, the Conclusion summarizes findings, interprets them within a broader context, explores implications, synthesizes the paper's key aspects, enhances the subject understanding, and suggests future research directions for broader impact. This five-move framework allows for the research to be presented in a well-organized and systematic manner. There are multiple steps in Move 1 (I), Move 3 (M), and Move 5 (C), but no steps in Move 2 (P) and Move 4 (F).

3.3 Data Analysis

This study employed an AI-assisted approach to conduct a qualitative analysis of rhetorical moves within academic writing. To address the first research question, a recent version of ChatGPT (GPT-4) was used to identify the moves and steps in 20 abstracts sourced from Scopus-indexed research articles in the field of tourism. The interaction with GPT-4 commenced with the researchers initiating a "new chat" and inputting specific prompts and cues.

Firstly, Hyland's (2000) comprehensive five-move model for abstract analysis was inserted into the chat interface in detail, serving as the instruction set for the AI program. Subsequently, GPT-4 was tasked with performing the analysis based on the provided requirements. It is crucial to highlight that the precision and clarity of the instructions and questions were directly proportional to the accuracy of the results rendered by GPT-4. The more lucid and explicit the directives, the more accurate and reliable the outcomes would be delivered by the AI system. Hence, the researchers tried to add some instructions on how to code chunks of text and underline the typical parts that realize certain moves and steps.

The prompt entered to GPT-4 was like:

- *The contents of Hyland's (2000) five-move model.*
- *A rhetorical move is characterized as a chunk of text that contains at least one complete sentence and serves a specific communicative function. This means that whenever a sentence or combination of sentences fits the description of any move or step in the coding system, it is regarded to be an instance of a move or step regardless of its length.*
- *By adopting Hyland's (2000) five-move model, could you please tag chunks of text as No. 1, 2, 3, underline the typical parts that realize each move and step, and make a table by comprehensively analyzing the moves and steps in the following abstract?*

With clicking "send message" to deliver the prompt, GPT-4 would answer like:

- *Certainly! We can break down the provided abstract according to Hyland's Five-Move Model. Here's a coded version of the abstract, with each move underlined and labeled.*
- *[A coded version of the full abstract with some bold (underlined) parts]*
- *This is a quick analysis, and it's important to note that this abstract does not contain all steps within every move as per Hyland's model. Here is a table that categorizes the moves and steps in the abstract:*
- *[Table]*

After getting the results for the identification of moves and steps in 20 abstracts, the accuracy of GPT-4's analysis was examined by human coders, and the comparison was further made by showing the primary differences.

Finally, the essentiality of moves and steps was analyzed manually based on an arbitrary cut-off point (Can et al., 2016). Essentiality refers to the percentage of texts that include the move or step at least once (Lu et al., 2021). Kanoksilapatham (2005) proposed a 60% threshold for determining the essentiality of moves, while Rasmeenin (2006) enhanced this approach by implementing a spectrum of percentages to provide a more precise evaluation of each move's stability. This study used Rasmeenin's (2006) rationale by defining a move as obligatory if it appeared in 100% of the analyzed corpus, conventional if it occurred in 66% to 99% of the corpus, and optional if it was present in less than 66%. The constituent steps within each move were evaluated using the same criteria.

4. Results and Discussion

4.1 GPT-4's Ability of the Identification of Moves and Steps

Based on the results, the researcher found that ChatGPT (GPT-4) could identify rhetorical moves in a fast and evident manner. Table 2 presents an example of GPT-4's identification of moves and steps in tourism research article abstracts.

Table 2. An Example of GPT-4's Identification of Moves and Steps in Tourism Research Article Abstracts

Move Numbers (Labels)	**Step Numbers (Labels)**	**Extracted Underlined Texts**
1 (Introduction)	**2 (Making topic generalization)**	little is known about how location-related factors influence…
1 (Introduction)	**4 (Identifying gap)**	it largely ignores on-site cognitive assessments…
2 (Purpose)	**- (Stating the research purpose)**	The present study intended to address these gaps…
3 (Methods)	**1 (Describing participant/data sources)**	…fear-arousing walking interviews with foreign tourists (n = 24) in the city center of Munich, Germany.
4 (Findings)	**- (Describing the main results)**	Many participants self-reported the absence of fear of terrorism…12 identified location-related factors...
5 (Conclusion)	**4 (Presenting recommendation and implication)**	Theoretical and managerial implications are discussed.

From the example, it could be seen that all five moves were detected by GPT-4 and their sequence was M1-M2-M3-M4-M5. For the accuracy of labels of moves, GPT-4 correctly recognized them. However, regarding the labels of the steps or sub-moves, GPT-4 miss-coded one step. For the chunks of text "little is known about…", the human coders put it into the category of Move 1 Step 4 (Identifying gap) instead of Move 1 Step 2 (Making topic generalization) because it was more like developing understudied area of the literature rather than explicitly providing statements about the current state of knowledge, consensus, practice or description of phenomena (Swales, 1990).

This suggested that GPT-4 could identify moves and steps, and it went through multiple phases to analyze them before generating results. Phase 1 is input processing. Initially, the GPT model processed the input by understanding the user's quest by parsing the language, identifying key concepts, and understanding the context and specific task. In Phase 2, preprocessing, GPT-4 tokenized and encoded the text, breaking it into tokens—words, parts of words, or punctuation. These tokens were then converted to a numeric format using a pre-trained vocabulary, translating language into a machine-readable format. In Phase 3, model computation, GPT-4 utilized its transformer architecture and attention mechanism to analyze and understand the context and significance of each token, focusing on the most important parts of the text for the task. Phase 4 is the application of specific models. For tasks like this using Hyland's model, GPT-4 relied on its training which contained knowledge of various analytical frameworks. The model used the formed knowledge to recognize and label parts of the text following the given framework. The model identified patterns consistent with the steps and moves of the particular model (such as Introductions, Purposes, Methods, Findings, and Conclusion in Hyland's model). In Phases 5 and 6, GPT-4 generates and delivers responses. After analyzing and synthesizing relevant information, it creates a coherent message in formats like summaries or tables. The final output is presented in user-understandable forms, fulfilling the specified request.

4.2 Human Coders' Identification of Moves and Steps

After the identification of moves and steps by GPT-4, human coders verified the results and some inconsistencies were found. Table 3 displays the types of errors and their frequencies identified by GPT-4 in move analysis, along with corresponding examples.

Table 3. Types of Errors through GPT-4 Move Analysis

Types of Errors	Frequencies	Examples		Number of Abstracts
		By GPT-4	By Human	
1.Unmatched Categorization	**5**	M1S1	M1S2	3, 4, 8, 9, 13
	2	M1S2	M1S4	1, 4
	1	M1S2	M2	20
	1	M1S4	M1S2	3
	1	M3S3	M3S2	5
	6	M3S2	M3S3	5, 6, 7, 8, 10, 11, 16
2. Redundancy/ Over-Count	**8**	M1S1	No M1S1	5, 7, 10, 11, 14, 15, 17, 18
	5	M1S2	No M1S2	6, 7, 11, 17, 18
	4	M1S3	No M1S3	9, 10, 11, 17
	2	M1S4	No M1S4	6, 17
	1	M2	No M2	17
	1	M3S2	No M3S2	9
	1	M5S1	No M5S1	16
	1	M5S2	No M5S2	19
	1	M5S4	No M5S4	14
3. Vague Identification	**1**	M5S1-3	M5S2-S1-S3	5
4. Incorrect Sequence	**1**	M2-M3S1-M3S3	M3S1-M3S3-M2	3
	1	M2-M3S1-M3S2	M3S2-M2-M3S1	4
	1	M1S4-M2-M3S1-M3S2	M3S2-M3S1- M2-M1S4	18

Table 3 lists the four types of errors identified in the analysis of abstracts by GPT-4, categorized as Unmatched Categorization, Redundancy/Over-Count, Vague Identification, and Incorrect Sequence. It also includes the frequency of each error type, examples of the specific errors, and the abstract numbers where these errors were found.

From the data presented, the most frequent type of error was identified through GPT-4 analysis, starting from the highest frequency Redundancy/Over-Count. This error type is the most frequent, with instances such as labeling a move or step as present by GPT-4 when it is not identified in the text by human coders. The highest frequency of errors in this category involves M1S1 (8 occurrences), followed by M1S2 (5 occurrences), and M1S3 (4 occurrences).

Taking M1S1 (arguing for topic significance) as an example, the initial sentence in Abstract 5 was redundantly coded by GPT-4. Upon the human coder's closer examination, this sentence more aptly fulfilled the criteria for M2 (stating the research purpose), as it explicitly outlined the objectives of the study. However, following the coding of Move 1, when GPT-4 subsequently identified any sentences in alignment with M2, it coded this same sentence, as recorded in Table 4.

Table 4. An Example of Redundancy/Over-Count of Moves or Steps by GPT-4

Move	Step	Extracted Text in Abstract 5
1	**1**	The present study investigates the structural associations between workplace bullying, employee psychological distress, and work engagement...
2	**-**	The present study investigates the structural associations between workplace bullying, employee psychological distress, and work engagement... (blended with Move 1, Introduction)

Based on Table 4, this result suggests that the move analysis by GPT-4 is mechanically applied based on the sequence in which each sentence appears within a paragraph (here, the paragraph refers to the abstract). Even if the first sentence of the paragraph does not employ M1S1, it will still forcefully fit this sentence into M1S1. This shows that one of the shortcomings of GPT-4 in move analysis is sequential bias. GPT-4 might be programmed or

trained to prioritize the sequence of sentences as a primary heuristic for categorization. This could lead to an overreliance on sentence order rather than a nuanced understanding of content and function within the text.

The second frequent error type is Unmatched Categorization in GPT-4 move analysis. This error occurs when GPT-4 assigns a sentence to a different move or step than a human expert would. Within this category, the highest frequency of error (6 times) was observed when GPT-4 incorrectly classified sentences as belonging to Step 2 (describing instruments) of Move 3 (methods), when they should have been categorized under Step 3 (describing procedure and context) of the same move, affecting abstracts numbered 5, 6, 7, 8, 10, and 11. Other frequent errors in this category involve misclassifications between steps within the Introduction move and between Introduction and Purpose moves.

Taking M3S2 (describing instruments) as an example, the extracted text in Abstract 6 was incorrectly identified by GPT-4, and it was supposed to be M3S3 (describing procedure and context) by human coders.

Table 5. An Example of Unmatched Categorization of Moves or Steps by GPT-4

Move	Step	Extracted Text in Abstract 6
3	**2**	To test the hypotheses of the proposed theoretical model, the PLS-SEM technique was used...
3	**1**	...based on 611 responses from consumers staying in hotels (254 in Ukraine and 357 in Spain).

In academic research, instruments are tools designed to gather data from participants in a structured and systematic way, such as observations, interviews, questionnaires, surveys, checklists, focus groups, and secondary data. As can be seen from the attached sentence in Table 5, due to the mention of the testament of the hypothesis of the theoretical model by PLS-SEM, it is more like describing procedure rather than depicting instruments because Partial Least Squares Structural Equation Modeling (PLS-SEM) is not an instrument in the traditional sense of tools. Instead, it is a statistical analysis method used to model complex relationships between observed variables and latent constructs. This result indicates that GPT-4 lacks domain-specific knowledge and has a misinterpretation of terminology. GPT-4 possesses a broad knowledge of "methods" but it fails to understand the distinctions between research instruments and analytical methods.

The next frequent error type is the Incorrect Sequence. Such an error is made when the order of moves in abstracts is misinterpreted by GPT-4. For example, in Abstract 18, the human coding of the moves and steps of the source text is:

This study surveyed [M3S2] 1562 wellness tourists [M3S1] in Finland, the St. Petersburg area, and Lithuania to explore what were their motivations for taking a wellness trip, to investigate what differences there were between different nationalities and determine how wellness tourists could be segmented based on their motivation [M2]. As most previous studies analyzed wellness tourism motivations in one chosen destination, this study fills the gap in cross-cultural analysis of motivations [M1S4].

Based on the text, human coders' order of moves and steps is M3S2 (Describing instruments) → M3S1 (Describing participant/data sources) → M2 (Stating the research purpose) → M1S4 (Identifying gap) while GPT-4's order of moves and steps is: M1S4 (Identifying gap) → M2 (Stating the research purpose) → M3S1 (Describing participant/data sources) → M3S2 (Describing instruments).

Table 6. An Example of Incorrect Sequence of Moves or Steps by GPT-4

Move	Step	Extracted Text in Abstract 18
1	**4**	As most previous studies analyzed wellness tourism motivations in one chosen destination, this study fills the gap in cross-cultural analysis of motivations.
2	**-**	This study aimed to explore what were the motivations for taking a wellness trip, to investigate what differences there were between different nationalities and determine how wellness tourists could be segmented based on their motivation.
3	**1**	This study surveyed 1562 wellness tourists in Finland, the St. Petersburg area, and Lithuania…
3	**2**	Exploratory factor analysis and Seemingly unrelated regression analysis were used…
3	**3**	…to explore what were their motivations for taking a wellness trip…

As can be seen in Table 6, GPT-4 even changes the sentence sequence of the original text from the Abstract. It disrupts the logical flow that is typically expected in academic writing. This result suggests that the effectiveness of AI models in performing specific tasks is heavily dependent on the quality and variety of their training data. If the training data lacked a sufficient range of examples illustrating the proper sequence of moves and steps in academic texts, the model might not learn to recognize and replicate the expected order accurately. Moreover, in trying to infer the move structure of an abstract, GPT-4 infers more than relies more. It might rely too heavily on certain keywords or phrases that typically signal specific moves or steps, leading to incorrect assumptions about the order in which these elements should appear.

The last frequent error type is Vague Identification. In one case (See Table 7), GPT-4 only analyses the range of steps in a move, rather than listing them one by one as M5S1, M5S2, and M5S3 based on the corresponding chunks of texts that realize each step. Though this error is not very common, it shows that when faced with the task of analyzing specific text segments, the AI model might prioritize broader patterns it has learned over the detailed, fine-grained analysis needed to distinguish between closely related steps within a move.

Table 7. An Example of Vague Identification of Moves or Steps by GPT-4

Move	Step	Extracted Text in Abstract 5
5	**1-3**	The study findings add solid and valuable contributions...by illustrating how bullying behavior could impact employees' psychological distress and their work engagement, considering the interaction role of self-compassion... Limitations and future research are further discussed.

4.3 Gpt-4's Analysis of the Essentiality of Moves and Steps

In the context of Scopus-indexed tourism research articles, the essentiality of various moves and steps, as framed by Hyland's (2000) Five-Move Model, offers a fascinating insight into the construction of academic writing. The term "essentiality rate" denotes the proportion of tourism research article abstracts (N=20) that incorporate each move or step. This metric is alternatively known as the inclusion rate, which reflects the extent to which the texts encompass each move or step. By examining the essentiality rate, we can ascertain the level of essentiality, thereby gauging the necessity of each move or step in the abstracts of Scopus-indexed tourism research articles.

Table 8 shows the number of abstracts that contain each specific move and step, their corresponding essentiality rates, and their categorization based on essentiality through GPT-4 analysis. However, due to GPT-4's identification errors in the analysis of moves and steps, the essentiality rates and essentiality categories were also inaccurate. Hence, the results regarding the essentiality of moves and steps were checked by humans, as reported in Table 9.

Table 8. Essentiality of Moves and Steps in Abstracts of Scopus-Indexed Tourism Research Articles through GPT-4 Analysis

Tourism Abstracts (N=20)	Moves (M) & Steps (S)	Number of Abstracts	Essentiality Rates	Essentiality
	M1	**20**	**100%**	**Obligatory**
	M1 S1	19	95%	Conventional
	M1 S2	13	65%	Optional
	M1 S3	5	25%	Optional
	M1 S4	9	45%	Optional
	M2	**20**	**100%**	**Obligatory**
	M3	**19**	**95%**	**Conventional**
	M3 S1	19	95%	Conventional
	M3 S2	15	75%	Conventional
	M3 S3	15	75%	Conventional
	M4	**20**	**100%**	**Obligatory**
	M5	**15**	**75%**	**Conventional**
	M5 S1	11	55%	Optional
	M5 S2	9	45%	Optional
	M5 S3	3	15%	Optional
	M5 S4	10	50%	Optional

Table 9. Essentiality of Moves and Steps in Abstracts of Scopus-Indexed Tourism Research Articles through Human Analysis

Tourism Abstracts (N=20)	Moves (M) & Steps (S)	Number of Abstracts	Essentiality Rates	Essentiality
	M1	**14**	**70%**	**Conventional**
	M1 S1	6	30%	Optional
	M1 S2	8	40%	Optional
	M1 S3	1	5%	Optional
	M1 S4	7	35%	Optional
	M2	**20**	**100%**	**Obligatory**
	M3	**19**	**95%**	**Conventional**
	M3 S1	19	95%	Conventional
	M3 S2	9	45%	Optional
	M3 S3	15	75%	Conventional
	M4	**20**	**100%**	**Obligatory**
	M5	**14**	**70%**	**Conventional**
	M5 S1	10	50%	Optional
	M5 S2	7	35%	Optional
	M5 S3	3	15%	Optional
	M5 S4	8	40%	Optional

Based on Table 8 and Table 9, there are some similarities and differences based on GPT-4 and human results.

The presence of a 100% essentiality in both Move 2 (stating the research purpose) and Move 4 (describing the main results) assures that they are invariable elements shared among all the 20 abstracts in tourism research articles. Just like laying the stone of the foundation of a building, the research purpose is the vital act of setting the intentions and goals statement, and this is a clear directive for the whole structure that follows. As the actual results are often the strongest part of the research articles, they are like the keystones of the arch, the central pieces that hold the structure together, unifying the different parts into a single whole.

Regarding Move 1 (Introduction), the findings by GPT-4 analysis showed that this move is obligatory, which means 100% of the tourism abstracts include it. However, through human analysis, it is found conventional, and 70% percent of the same abstracts contain it. As for the four steps in this move, GPT-4 analysis suggests that arguing for the topic's significance (Step 1) is almost universally necessary, with a 95% essentiality rate, indicating it is a conventional practice. However, making topic generalizations (Step 2), defining key terms (Step 3), and identifying a gap (Step 4) are viewed more optionally, with essentiality rates of 65%, 25%, and 45% respectively. In contrast, human analysis rates all these steps as more optional: Step 1 at 30%, Step 2 at 40%, Step 3 at a mere 5%, and Step 4 at 35%.

For Move 3 (Methods), both GPT-4 and human analyses find common ground—a 95% essentiality rate underscores its high conventional status and reaffirms the importance of presenting the research methodology as the foundation for understanding and evaluating the study's validity. Step 1, describing participant/data sources, is unanimously seen as conventional by both analyses, emphasizing the necessity of transparency in research design. However, a notable divergence appears in the perception of Step 2, describing instruments. GPT-4 views this step as conventionally essential (75%) whereas humans deem it more optional (45%). Step 3, describing the procedure and context, is regarded as conventional by both, but with a higher consensus from GPT-4, reflecting a shared understanding of the importance of this step in replicating and understanding the research framework.

As for Move 5 (Conclusion), both GPT-4 and human analysis tend to stick to the conventional category, with GPT-4 reporting the essentiality rate at 75% and the human counterparts at only 70%. Concerning the steps, both analyses deemed all steps as optional, but with varying degrees of essentiality. Step 1 (deducing conclusions), Step 2 (evaluating the significance of research), and Step 4 (presenting recommendations and implications) were seen with slightly higher essentiality rates by GPT-4 compared to human analysis. However, both analyses agreed that Step 3 (stating limitations) is least emphasized, marked by a 15% essentiality rate, highlighting a shared view on the variability of including limitations within abstract conclusions.

5. Conclusion

Adopting Hyland's (2000) classical five-move model, this study investigated the ability of GPT-4 in text analysis, specifically its role in the analysis of rhetorical moves within tourism research article abstracts indexed in Scopus. It also explored the essentiality of the moves and steps by GPT-4 and compared the accuracy of GPT-4's identification and analysis of rhetorical moves with that of human coders. The results highlighted ChatGPT's significant advancements in linguistics, marking its utility as a versatile tool capable of emulating human-like conversations and showcasing impressive language comprehension skills.

GPT-4 can identify rhetorical moves in a fast and evident manner through six phases, which are input processing, preprocessing, model computation, application of specific models, response generation, and output delivery. However, the accuracy of move identification is not as clear as that of human coders. The errors mainly result from redundancy/over-count, unmatched categorization, incorrect sequence, and vague identification of moves or steps. The reasons behind these types of errors include overreliance on sentence order rather than a nuanced understanding of content, a lack of domain-specific knowledge and a misinterpretation of terminology, overreliance on keyword signaling instead of the appearance of the original texts, and broad-brush analysis over granularity.

Concerning the essentiality of moves and steps, both analyses agree on the obligatory nature of Move 2 (stating the research purpose) and Move 4 (describing the main results), recognizing these two moves as essential across all examined abstracts. Move 3 (Methods), with its high essentiality rate of 95%, stands as a point of consensus between GPT-4 and human analyses, emphasizing the universal acknowledgment of the importance of thoroughly describing research methodology to guarantee the study's validity and replicability. This agreement extends specifically to Move 3 Step 1 (Describing Participant/Data Sources) and Move 3 Step 3 (Describing Procedure and Context). Both steps are recognized equally by GPT-4 and humans as critical, with essentiality rates of 95% and 75% respectively, highlighting a shared understanding of the need for transparency in research design and the detailed explanation of procedural context. As for Move 5 (Conclusion), both analyses classify it as conventional (with a 75% essentiality rate through GPT-4 analysis and 70% through human analysis), indicating a shared recognition of the importance of concluding remarks, although there is some leeway regarding the incorporation of particular steps within this move. Concerning steps, Move 5 Step 3 had the same essentiality rate of 15% recognized by both GPT-4 and humans.

However, disparities emerge in specific steps of Move 1 (Introduction), Move 3 (Methods), and Move 5 (Conclusion), with GPT-4 frequently achieving higher rates of essentiality, specifically in Move 1 Step 1 (Arguing for the Significance of the Topic), Move 3 Step 2 (Describing the Instruments), Move 1 Step 2 (Making Topic Generalization), Move 1 Step 3 (Defining key terms), Move 5 Step 2 (Evaluating the Significance of the Research), Move 5 Step 4 (Presenting Recommendation and Implication), and Move 5 Step 1 (Deducing Conclusion). These inconsistencies are mainly due to the errors made by GPT-4 in the identification process of moves and steps. Tackling these obstacles would entail the continuous improvement of AI models to achieve an effective feedback loop that incorporates expert human analysis. The process can create a more accurate AI in mitigating academic rhetoric, which is mostly determined by practical knowledge. This AI then can be useful as a tool for text analysis.

The significance of this research is underscored by the need to understand the advancements and shortcomings of AI in academic discourse analysis. This study also paves the way for future research to explore the potential enhancements of AI capabilities by enlarging the corpus size.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 1-14. https://doi.org/10.1080/10494820.2023.2253858

Alyousef, H. S. (2023). A multidimensional analysis of linguistic realizations and rhetorical move structure in geography research article abstracts: A corpus-based study. *Atlantis: Revista de la Asociación Española de Estudios Anglo-Norteamericanos*, *45*(1), 190-212. https://doi.org/10.28914/Atlantis-2023-45.1.11

Bhatia, V. K. (1993). *Analyzing Genre: Language Use in Professional Settings*. Longman Publishing.

Biber, D. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure* (Vol. 28). John Benjamins Publishing. https://doi.org/10.1075/scl.28

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *Preprints*. https://doi.org/10.48550/arXiv.2303.12712

Can, S., Karabacak, E., & Qin, J. (2016). Structure of moves in research article abstracts in applied linguistics. *Publications, 4*(3), 1-16. https://doi.org/10.3390/publications4030023

Ding, H. (2007). Genre analysis of personal statements: Analysis of moves in application essays to medical and dental schools. *English for specific purposes, 26*(3), 368-392. https://doi.org/10.1016/j.esp.2006.09.004

Fauzan, U., Lubis, A., & Kurniawan, E. (2020). Rhetorical moves and linguistic complexity of research article abstracts in international applied linguistics journals. *The Asian E.S.P. Journal, 16*(5.2), 219-247.

Fitria, T. N. (2022). Structure Analysis of English Abstracts in International Journal Published in 2020. *Journal of English Language and Culture*, *12*(1). https://doi.org/10.30813/jelc.v12i1.2815

Geng, H. (2024). *Lexical Bundles in Rhetorical Moves of Introduction in English Linguistics Research Articles from Non-Scopus and Scopus Journals* (Doctoral thesis, Universiti Putra Malaysia, Serdang, Malaysia).

Geng, H., & Nimehchisalem, V. (2023). Can ChatGPT Analyse Textual Data? The Sub-Themes Reflected by Typical Conceptual Metaphors in Short Stories of Language Assessment. *ASEAN Journal of Applied Linguistics*, *2*, 16-31.

Geng, H., & Wei, H. (2023). Metadiscourse Markers in Abstracts of Linguistics and Literature Research Articles from Scopus-Indexed Journals. *Journal of Modern Languages*, *33*(1), 29-49. https://doi.org/10.22452/jml.vol33no1.2

Geng, H., Lee, G. I., Jalaluddin, I., & Tan, H. (2023). Occurrence Frequency of Rhetorical Moves in Introductions of Linguistics Research Articles From Non-Scopus and Scopus Journals. *Journal of Language Teaching and Research*, *14*(5), 1279-1289. https://doi.org/10.17507/jltr.1405.16

Geng, H., Lee, G. I., Jalaluddin, I., & Tan, H. (2023). Rhetorical Moves of Introduction Sections in English Linguistics Research Articles From Two Non-Scopus and Two Scopus Journals. *Theory and Practice in Language Studies*, *13*(8), 2087-2096. https://doi.org/10.17507/tpls.1308.25

Ghasempour, B., & Farnia, M. (2017). Contrastive move analysis: Persian and English research article abstracts in law. *The Journal of Teaching English for Specific and Academic Purposes, 5*(4), 739-753. https://doi.org/10.22190/JTESAP1704739G

Gusmana, A. L. (2023). *An Analysis Study of Rhetorical Moves in Sinta-Indexed Psychological Journal Abstracts* [Doctoral dissertation] Universitas Pendidikan Indonesia.

Hyland, K. (2000). *Disciplinary discourses: Social interaction in academic writing.* Longman Pearson Education.

Joshi, A. (2016). Comparison between Scopus and ISI web of science. *Journal Global Values*, *7*(1), 1-11.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for specific purposes, 24*(3), 269 292. https://doi.org/10.1016/j.esp.2004.08.003

Knight, S., Abel, S., Shibani, A., Goh, Y. K., Conijn, R., Gibson, A., ... & Shum, S. B. (2020). Are you being rhetorical? a description of rhetorical move annotation tools and open corpus of sample machine-annotated rhetorical moves. *Journal of Learning Analytics*, *7*(3), 138-154. https://doi.org/10.18608/jla.2020.73.10

Kurniawan, E., & Sabila, N. A. A. (2021). Another look at the rhetorical moves and linguistic realizations in international and Indonesian journal articles: A case of tourism research article abstracts. *Indonesian Journal of Applied Linguistics*, *11*(2), 318-329. https://doi.org/10.17509/ijal.v11i2.32055

Livberber, T. (2023). Toward non-human-centered design: Designing an academic article with ChatGPT. *Profesional de la información/Information Professional*, *32*(5). https://doi.org/10.3145/epi.2023.sep.12

Lu, X., Casal, J. E., Liu, Y., Kisselev, O., & Yoon, J. (2021). The relationship between syntactic complexity and rhetorical move-steps in research article introductions: Variation among four social science and engineering disciplines. *Journal of English for Academic Purposes, 54*, 1-13. https://doi.org/10.1016/j.jeap.2021.101006

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

Reiss, M. V. (2023). Testing the reliability of chatgpt for text annotation and classification: A cautionary remark. Preprints. https://doi.org/10.48550/arXiv.2304.11085

Saidi, M., & Karami, N. (2024). A Cross-Move Analysis of Interactional Metadiscourse Markers in Abstracts of Local and International Journals of History. *Journal of Language Horizons*, *7*(4). https://doi.org/10.22051/lghor.2023.41767.1738

Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics. *Text-Interdisciplinary Journal for the Study of Discourse, 16*(4), 481-499. https://doi.org/10.1515/text.1.1996.16.4.481

Savelka, J., Ashley, K. D., Gray, M. A., Westermann, H., & Xu, H. (2023). Can GPT-4 Support Analysis of Textual Data in Tasks Requiring Highly Specialized Domain Expertise? Preprints. https://doi.org/10.1145/3587102.3588792

Sidek, H. M. (2017). Entrepreneurial education conference proceedings: a rhetorical moves analysis of abstracts. *IJASOS-International E-journal of Advances in Social Sciences*, *3*(9), 1112-1119. https://doi.org/10.18769/ijasos.401178

Siiman, L. A., Rannastu-Avalos, M., Pöysä-Tarhonen, J., Häkkinen, P., & Pedaste, M. (2023, August). Opportunities and Challenges for AI-Assisted Qualitative Data Analysis: An Example from Collaborative Problem-Solving Discourse Data. In *International Conference on Innovative Technologies and Learning* (pp. 87-96). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-40113-8_9

Swales, J. (1990). *Genre analysis: English in academic and research settings*. CUP.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (Vol. 1). University of Michigan Press.

Tankó, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes*, *27*, 42-55. https://doi.org/10.1016/j.jeap.2017.04.003

Zhang, B., & Wannaruk, A. (2016). Rhetorical Structure of Education Research Article Methods Sections. *PASAA: Journal of Language Teaching and Learning in Thailand, 51*, 155-184. https://doi.org/10.58837/CHULA.PASAA.51.1.6

Copyrights