



<https://doi.org/10.22363/2687-0088-45774>

EDN: NNGCIG

Research article / Научная статья

Using ChatGPT in identifying rhetorical moves in grant proposal abstracts

Olga A. BOGINSKAYA  

Irkutsk National Research Technical University, Irkutsk, Russian Federation

 olga_boginskaya@mail.ru

Abstract

Rhetorical move analysis has gained popularity in research aimed at describing the rhetorical structure of academic genres, but traditional methods involving humans are sometimes subjective, time-intensive, and inconsistent. This study explores the viability of using ChatGPT for move analysis of grant proposal abstracts (GPAs). It aims to evaluate the accuracy and reliability of ChatGPT in identifying rhetorical moves in grant proposal abstracts compared to human annotation. To achieve this purpose, the research seeks answers to the questions how ChatGPT can be instructed to assist in identifying and annotating rhetorical moves in grant proposal abstracts, what types of rhetorical moves are more accurately identified by ChatGPT, and which types present challenges for it, and to what extent ChatGPT's annotation differs from that of a human annotator. Using a corpus of Russian-language GPAs submitted to the Russian Scientific Fund, the study developed a prompt intended to guide ChatGPT in identifying rhetorical moves typical of this genre. The accuracy of AI identification was assessed in comparison with the annotation performed by a human. The results show that a carefully created prompt with detailed examples and clear definitions yields high levels of agreement with a human annotator. However, the study showed that while ChatGPT demonstrated good performance in identifying rhetorical moves, its annotations differed from the annotations performed by a human. The human annotator frequently outperformed ChatGPT, particularly in recall (the completeness of detecting all instances of a given move) for certain categories. Nevertheless, the study showed that ChatGPT can be a useful tool not just for applied linguistics researchers, but also for improving pedagogical approaches to grant writing teaching. The AI-based identification of rhetorical moves can be used to provide EAP students with examples of efficient rhetorical methods in successful grant applications, allowing them to get a better understanding of genre conventions and improve their grant writing skills.

Keywords: *ChatGPT, prompt, move, grant proposal abstracts, automated move analysis*



For citation:

Boginskaya, Olga A. 2026. Using ChatGPT in identifying rhetorical moves in grant proposal abstracts. *Russian Journal of Linguistics* 30 (3). 585–609. <https://doi.org/10.22363/2687-0088-45774> EDN: NNGCIG

Использование ChatGPT для анализа риторической структуры аннотаций заявок на грант

Ольга Александровна БОГИНСКАЯ  

*Иркутский национальный исследовательский технический университет, Иркутск,
Российская Федерация*
 olgae_boginskaya@mail.ru

Аннотация

Анализ риторических ходов является актуальным направлением исследований риторической структуры академических жанров, однако традиционные методы анализа с участием человека субъективны, трудоемки и не всегда дают точные результаты. В данном исследовании изучается возможность применения модели ChatGPT для анализа риторических ходов в аннотациях заявок на грант. Цель исследования — определить степень точности аннотирования риторических ходов, выполненного с помощью ChatGPT, в сопоставлении с экспертной разметкой. Для достижения поставленной цели осуществляется разработка инструкции (промпта) для наиболее точной идентификации риторических ходов; выявляются типы риторических ходов, которые вызывают наибольшие трудности у ChatGPT; сопоставляются результаты риторического анализа, полученные ChatGPT и экспертом в области академического письма. В исследовании использовался корпус аннотаций заявок на грант, поданных в Российский научный фонд. В процессе исследования был разработан промпт, предназначенный для помощи ChatGPT в процессе идентификации типичных для данного академического жанра риторических ходов. Точность аннотации, выполненной ИИ, оценивалась в сравнении с разметкой, сделанной экспертом. Выявлено, что промпт, содержащий примеры и определения ходов, позволяет достичь высокой степени соответствия между автоматической аннотацией и экспертной разметкой. Наряду с этим были обнаружены и некоторые несовпадения: экспертная аннотация показала более высокую точность по критерию полноты. Результаты исследования свидетельствуют о том, что ChatGPT можно использовать как в лингвистических исследованиях, так и при написании грантовых заявок: автоматизированная идентификация риторических ходов даёт возможность анализировать успешные риторические приёмы, углубляя понимание жанровых особенностей текста и повышая навыки академического письма.

Ключевые слова: *ChatGPT, промпт, риторический шаг, аннотация, заявка на грант, автоматизированный анализ риторических ходов*

Для цитирования:

Boginskaya O.A. Using ChatGPT in identifying rhetorical moves in grant proposal abstracts. *Russian Journal of Linguistics*. 2026. Vol. 30. № 3. P. 585–609. <https://doi.org/10.22363/2687-0088-45774> EDN: NNGCIG

1. Introduction

The rapid development of Artificial Intelligence (AI), particularly Large Language Models (LLMs), is transforming research practices across disciplines. In applied linguistics and discourse studies, LLMs are employed to automate the analysis of written texts, including the identification of rhetorical structures. However, despite their growing potential, AI tools remain imperfect, producing inconsistent and inaccurate annotations and struggling with context-dependent rhetorical functions. As Ozyumenko and Larina (2025) argue in their analysis of AI in translation, while AI offers significant advantages in processing linguistic data, it also exhibits clear limitations related to contextual understanding, cultural nuance, and the interpretation of implicit meaning. This raises a question: to what extent can we trust AI to perform complex rhetorical tasks previously carried out by human experts?

One area where this issue is crucial is move analysis, which involves identifying rhetorical moves that fulfil specific communicative purposes. However, traditional manual move analysis is time-consuming, labour-intensive, and can be subjective. These limitations have encouraged researchers to explore the potential of LLMs such as ChatGPT or DeepSeek for automating move identification.

Recent studies have demonstrated that LLMs can accurately annotate rhetorical moves in certain academic genres. For instance, Yu et al. (2024) used ChatGPT to identify moves in research article abstracts. Kim and Lu (2024) showed that fine-tuning ChatGPT significantly improves its performance in analysing research article introductions. These findings suggest that LLMs may offer an alternative to manual annotation, at least for genres with established move schemes.

However, one high-stakes genre has remained unexplored in this body of research: the grant proposal abstract. It is crucial for securing funding and advancing research careers, yet its rhetorical structure is less standardised than that of research articles. It remains unknown whether ChatGPT can accurately identify rhetorical moves in this complex genre, and which types of moves pose particular challenges for it.

To address this gap, the present study aims to evaluate the accuracy of the LLM GPT-4o developed by OpenAI (accessed via its web interface, 2024-05-13) in identifying rhetorical moves in grant proposal abstracts compared to human annotation. To achieve this purpose, the research seeks answers to the following questions:

- 1) How can ChatGPT be instructed to assist in identifying and annotating rhetorical moves in grant proposal abstracts?
- 2) What types of rhetorical moves are more accurately identified by ChatGPT?
- 3) Which types of moves present challenges for ChatGPT, and to what extent its annotation differs from that of a human annotator?

The results of investigating the rhetorical structure of successful grant proposal abstracts can be useful for novice academic writers, taking into account the role of this genre in securing funding and advancing research initiatives.

2. Theoretical framework

2.1. Move analysis

Move analysis, first developed by Swales (1981, 1990) and regarded as a crucial component of genre analysis, has been applied not only to academic genres, such as argumentative essays (Liu et al. 2024), research articles (Boginskaya 2025a, Alyousef 2021, Basturkmen 2012, Cotos et al. 2017), book reviews (Guangyuan & Zhaoxia 2025), grant proposals (Boginskaya 2025), and three-minute thesis presentations (Hu & Liu 2018) but also to the examination of grant recommendation letters (Afful et al. 2022), business emails (van Herck et al. 2022), legal cases (Bhatia 1993), sales promotion letters and advertisements (Khedri et al. 2022, Vergaro 2004), annual reports (de Groot 2008), corporate social responsibility reports (Yu & Bondi 2017, Yu 2025), and CEO statements (Wu 2020).

Swales (2004) defined ‘move’ as a discursual unit that performs a communicative function, shaping a writer’s or speaker’s rhetorical choices to align with the expectations of a discourse community. Moves thus represent strategic rhetorical actions in achieving specific communicative goals typical of a specific discourse community and serve both individual purposes such as introducing a topic, presenting evidence, and making a claim and the overall communicative purpose (Yu et al. 2024). They represent collective knowledge shared by members of a discourse community, which is critical for newcomers to integrate successfully into that community (Swales 1990).

By mastering the genre conventions, including move structures, they demonstrate competence and shape their professional identities, establishing themselves as legitimate members (Hyland 2015). As a result, examining move structures gives useful information for new members of discourse communities, allowing them to contribute to the discourse in their field.

Move analysis has gained considerable attention of researchers applying it as a method for examining linguistic choices that define academic or professional genres (Feng 2019, Hyland & Tse 2012, Jiang & Hyland 2017, Liu et al. 2024, Yu 2022, Yu 2025, Yu & Bondi 2017, van Herck et al. 2022). This approach allows for identifying recurring patterns, provides valuable insights into various aspects of genre construction, helps identify key linguistic features, calculate move frequency and length, and delineate typical move sequencing (Biber et al. 2007). It also contributes to the development of genre prototypes, which can then be effectively used in developing teaching materials and strategies to facilitate genre-specific writing proficiency of students.

Rhetorical move analysis includes two stages: move identification and move annotation (Yu 2022). The first stage involves a careful analysis of textual segments with the aim of identifying their communicative functions. Following this identification process, a clearly defined move scheme is applied to annotate the corpus analyzed both quantitatively and qualitatively, which allows researchers to evaluate the frequency of moves and to identify their linguistic markers (Yu 2022).

Despite being a valuable approach to examining genres, move analysis has some limitations, including potential simplification, subjectivity, and labor intensity. Human analysts sometimes simplify the functions of textual units and may ignore the complexity of authorial intentions (Hyland 2004), which can obscure the ways the writers construct their arguments. Crookes (1986) highlighted the need for validating these analyses, ensuring they are grounded in the text rather than merely reflecting the analyst's interpretations. Because move analysis relies on functional rather than formal units, it is not always possible to correctly connect linguistic features with particular rhetorical functions. While transitions between moves are influenced by writers' objectives, identifying the explicit markers of these moves within the text remains a challenge (Gray et al. 2020). Another challenge is subjectivity and contextual dependency of move analysis, which can contribute to its complexity and time-intensiveness (Yu et al. 2024).

The cultural specificity of academic writing conventions is an additional problem. Most move analysis frameworks have been developed on the basis of English-language academic texts. However, rhetorical conventions vary across cultural contexts. Ignoring cross-cultural differences risks imposing a Western-centric model of rhetorical organization onto genres produced in other cultural contexts, potentially leading to misidentification of moves. Nevertheless, it is important to note that despite some cultural differences, contemporary Russian academic discourse has shown a clear tendency to align with Anglophone academic writing norms, which is partly driven by the increasing pressure to publish in international peer-reviewed journals, the global dominance of English as the language of science, and the adoption of Western-style formats by leading Russian journals. As a result, the move structures found in Russian-language academic texts often closely mirror those described in English-medium genre analyses.

The labor-intensive nature of manual move analysis also creates obstacles to studies of discursive phenomena. This process requires training and time. Furthermore, the possibility of human errors might undermine the reliability of the manual analysis results, thus limiting the use of this approach. Given these challenges, researchers have begun to apply Large Language Models (LLMs) to address these challenges.

The rapid advancement of LLMs is reshaping the capacity to conduct rhetorical analysis, where the manual annotation of linguistic features is time-consuming and subjective. The studies have demonstrated the potential of LLMs in annotating linguistic phenomena in academic genres (Gilardi et al. 2023, Kuzman et al. 2023, Yu et al. 2024, Kim & Lu 2024). Yu et al.'s (2024) research, for example, investigated the effectiveness of LLMs in annotating rhetorical moves in research article abstracts. Employing the move scheme, including Background, Purpose, Method, Result, and Conclusion, their findings yielded accurate annotations, indicating the benefits of the LLM in analyzing the patterns of academic discourse. Using a manually annotated corpus of 100 applied linguistics research article introductions, Kim and Lu (2024), who assessed the feasibility of rhetorical move

analysis assisted by ChatGPT, demonstrated that while few-shot learning and prompt refinement offer incremental gains, fine-tuning contributes to significantly higher accuracy rates. However, the question of AI imperfection remains central. Baltezarević, Stošić, and Mikhailova (2026) emphasize that AI-assisted academic writing tools require not only technical refinement but also legal and ethical oversight, particularly when they are used for evaluative purposes.

3.2. Grant proposals as an academic genre

Grant proposals represent a genre of persuasive communication, aimed at securing funding from reviewers and agency officials for proposed research projects (Connor 2000). In grant proposals, as Myers (1990: 42) states, “one must persuade without seeming to persuade”.

Grant proposal abstracts hold a uniquely precarious position compared to their research article counterparts. While an article abstract is primarily a summary of the study, a grant proposal abstract aims, first of all, to secure funding (Locke et al. 2014), serving as the initial rhetorical test for persuading reviewers of the need to support the project. The initial impression is particularly crucial given the highly competitive funding landscape, where reviewers need to assess a large number of proposals (Boginskaya 2025b).

Structured abstracts have gained popularity across disciplines in response to the growing demand for efficient information retrieval. They use predetermined headers that replicate the IMRD framework to improve clarity and accessibility. Swales (1981) was the first to propose a four-move model comprising moves to create the field, report previous research, prepare for the current research, and present it. This was subsequently revised in Swales’ (1990) ‘Create a Research Space (CARS)’ model, which proposed a three-move approach, including Establishing a territory, Establishing a niche, and Occupying the niche.

Based on the Swalesian model, Connor and Mauranen (1999) developed a ten-move framework including components such as Territory, Gap, Goal, Means, Reporting previous research, Achievements, Benefits, Competence Claim, Importance Claim, and Compliance Claim. This framework was used for the analysis of both full-length research articles and grant proposals. All the subsequent studies of rhetorical moves in grant proposals rely on this model. Feng and Shi (2004), for instance, adapted Connor and Mauranen’s framework and identified a three-move structure for abstracts and a ten-move structure for full grant proposals. Feng (2006) refined this approach, developing a six-move framework, including Territory, Niche, Objectives, Means, Explanation and Justification, and Contributions. Tardy (2011) identified six moves, aimed at announcing the project, describing context, objectives, and methods, identifying project outcomes and impacts. Flowerdew (2016), also drawing on Connor and Mauranen’s model, developed a seven-move framework that includes Territory, Gap, Goal, Means, Achievements, Benefits, and Future recommendations. Matzler’s study (2021) reduced Flowerdew’s framework to five core moves: Territory, Niche, Goal,

Means, and Benefits, while Wang (2025) in her turn modified Matzler’s model by incorporating an additional move for Expected Outcome. Table 1 provides a summary of the frameworks used for analyzing move structures in grant proposal abstracts.

Table 1

Moves in Feng & Shi’s, Feng, Tardy’s, Flowerdew’s, Matzler’s, and Wang’s frameworks

Feng & Shi (2004)	Feng (2006)	Tardy (2011)	Flowerdew (2016)	Matzler (2021)	Wang (2025)
Need	Territory	Project	Territory Gap / niche	Territory	Territory
Means	Niche	announcing	Goal	Niche	Niche
Contributions	Objectives	Context	Means Achievements	Goal	Goal
	Means	Objectives	Benefits	Means	Means
	Explanation and justification	Methods Outcomes	Recommendations	Benefits	Benefits
	Potential contribution	Impacts			Expected outcome

Despite variations in terminology, these six frameworks have several features in common. All of them incorporate the five moves identified by Matzler’s (2021) study, indicating their significance for the genre. A key difference lies in Tardy’s (2011) framework, which is the only one that identifies ‘Project Announcing’ as a distinct move. Feng’s (2006) framework expands upon other models by including an ‘Explanation and Justification’ move, reflecting its observed use in her corpus to clarify the rationale and feasibility of the proposed research. Similarly, Wang (2025) builds upon the ‘Benefits’ move by adding ‘Expected Outcomes,’ a category that mirrors the ‘Results’ move commonly found in research article abstracts.

It is important to emphasize that existing schemes may contain moves that are too vaguely defined or overlapping, which can hinder AI and human annotators to accurately identify them. For instance, Flowerdew’s (2016) and Tardy’s (2011) models combine a larger number of sometimes overlapping moves (e.g., Achievements and Benefits or Outcomes and Impacts), causing confusion for both LLM and human annotators. Therefore, a more rigorously defined move scheme that minimizes ambiguity and maximizes replicability is required. Matzler’s (2021) framework, with its distinct categories for Territory, Niche, Goal, Means, and Benefits, appears to provide a more precise approach to analyzing grant proposal abstract elements, making it exceptionally suitable for this study as simpler taxonomies allow for more accurate and consistent annotation, while taxonomies with more abstract categories (‘Explanation and Justification,’ ‘Potential Contribution,’ or ‘Project announcing’) create difficulties for ChatGPT to accurately identify them. Matzler’s framework’s efficiency is also confirmed by the fact that its key moves appear in other recent genre studies (Charles & Whiteside 2024, Wang 2025).

3. Data and methods

3.1. Overall research design

To ensure transparency regarding the methodological sequence and the distinct roles of the corpora and annotators, the study was conducted in three phases. Table 2 provides an overview of these phases, specifying the corpus used, the executor(s) involved, and the specific objective of each phase.

Table 2

Overview of analytical phases, corpora, executors, and objectives

Phase	Corpus	Executor(s)	Objective
Phase 1: Prompt development and error analysis	C1 (30 abstracts)	Author (provided reference annotation) + ChatGPT (iterative outputs)	Design and refine the prompt; identify errors; develop the final prompt.
Phase 2: Internal validation	C2 (30 abstracts)	Author (provided reference annotation) + ChatGPT (final prompt)	Test the finalized prompt on unseen data to verify generalizability and prevent overfitting to C1; confirm acceptable performance.
Phase 3: Formal evaluation	C3 (40 abstracts)	Independent human annotator + ChatGPT (final prompt)	Compare ChatGPT’s annotations against a human expert on the main test set; calculate precision, recall, and F1 for each move category; conduct qualitative error analysis.

3.2. Corpus composition and selection criteria

In Russia, a major source of funding for research across various disciplines is the Russian Science Foundation (RSF). This fund operates on a highly competitive basis. Both early-career and experienced researchers can apply for financial support through a system of grants. The application process follows a one-stage system: researchers submit a grant proposal, including an abstract both in Russian and English. The abstract describes scientific objectives and methodology, expected academic and practical contributions and is followed by a description of potential outcomes and benefits, a professional profile of the principal investigator, and a detailed budget justification.

All abstracts included in the corpus were sourced from the official website of the Russian Science Foundation (RSF) (www.rsf.ru), where information about successfully funded projects is made publicly available through the Foundation’s Project Finder service. As the data consist of published, non-confidential project summaries, no special ethical approval or data transfer agreement was required for their use in this study. The abstracts were accessed and processed only for the purpose of linguistic analysis. No personally identifiable information was extracted or analysed.

The RSF website search interface provides only a single combined category for ‘Humanities and Social Sciences’. Therefore, to compile the corpus, all projects

listed in this combined category for the years 2021–2024 were first retrieved ($n = 2,164$). From this pool, only those projects belonging to the humanities were selected based on the following criteria: 1) classified under the combined Humanities and Social Sciences category, 2) identifiable as a humanities project based on its title, keywords, and disciplinary affiliation stated in the abstract, and 3) submitted between 2021 and 2024. This procedure yielded a total of 100 abstracts.

From this set of 100 abstracts, the pilot corpora (C1 and C2) and the test corpus (C3) were created using simple random assignment. The randomization unit was the individual abstract. The allocation was performed manually: each abstract was assigned a unique number, the numbered list was shuffled, and the required number of abstracts was drawn to form C1 (30), C2 (30 from the remaining), and C3 (the remaining 40). No computational random number generator or fixed seed was used. First, 30 abstracts were randomly selected for C1 (used for iterative prompt development and error analysis in Phase 1). Second, from the remaining 70 abstracts, another 30 were randomly selected for C2 (reserved as an internal validation corpus to test the final prompt on unseen data in Phase 2). Finally, the remaining 40 abstracts automatically constituted C3 (used as the main test set for the formal evaluation against the independent human annotator in Phase 3). It was verified that there was no textual overlap between C1, C2, and C3, ensuring that each abstract was assigned to exactly one corpus.

The abstracts cover five major subdisciplines: (1) linguistics and language studies (32% of the corpus); (2) literary studies (21%); (3) history and archaeology (20%); (4) philosophy and ethics (16%); (5) art history and cultural studies (11%). The study assumes that rhetorical move structures in grant proposal abstracts are homogeneous across humanities when the funding agency and genre are held constant. This assumption is based on two arguments. First, Connor's (2000) cross-disciplinary study of grant proposals found that variation between humanities and sciences is more significant than variation within humanities subdisciplines. Second, the RSF application guidelines prescribe a uniform abstract structure for all applicants, explicitly requiring the same sections (relevance, objectives, methods, expected outcomes).

3.3. Move taxonomy

For annotation, Matzler's (2021) five-move framework (Territory, Niche, Goal, Means, Benefits) was modified into a seven-move taxonomy. Specifically, 'Territory' was renamed CONTEXT, 'Niche' was renamed GAP, and the broad 'Means' category was split into three distinct moves — TASKS (specific steps/sub-objectives), METHODS (procedures and techniques), and MATERIALS (data, resources, or equipment), while GOAL and BENEFITS were retained from the original scheme. This modification was motivated by the error analysis during prompt development, which revealed that splitting 'Means' into distinct subcategories can improve annotation accuracy.

As in Yu et al. (2024), the labeling process adhered to two principles: first, labels were assigned at the sentence level; second, sentences containing multiple moves received multiple labels (with the primary move listed first). Table 3 presents the final seven-label taxonomy with full definitions.

Table 3

Moves in grant proposal abstracts	
Label	Description
CONTEXT	Provides background information, describes a real-world problem relevant to the research, or presents established scientific knowledge in the area being studied
GAP	Identifies a missing piece of information, a problem, or a limitation in current knowledge or existing solutions
GOAL	Clearly states the main objective or aim of the research project
TASKS	Lists the specific steps or sub-objectives that will be undertaken to achieve the overall GOAL
METHODS	Explains the procedures and techniques that will be used to conduct the research and achieve the TASKS
MATERIALS	Describes the resources, data, or equipment that will be used in the research
BENEFITS	Describes the potential positive outcomes and real-world applications of the research

While the target corpus consisted of Russian-language abstracts, all instructional components of the prompt, including move definitions, annotation rules, and formatting guidelines, were composed in English. This choice was motivated by three considerations. First, ChatGPT-4 is primarily pre-trained on English-dominant datasets. Second, using English for meta-instructions aligns the methodology with foundational studies on LLM-assisted move analysis, facilitating cross-study comparability. Third, it reduces the risk of terminological ambiguity when translating abstract rhetorical concepts into Russian. However, to enable the model to identify specific lexical cues in the Russian source texts, the examples and the lists of typical Russian linguistic markers for each move were provided in Russian within the prompt.

3.4. Prompt development

Following established prompting principles (Liu et al. 2023, Yu et al. 2024, Yu 2024), the prompt was refined through four stages. The full text of the final prompt is provided in Appendix 1.

Stage 1: Initial zero-shot prompt. The first prompt was a zero-shot instruction without examples: *‘Identify rhetorical moves in the following grant proposal abstract. Use these labels: CONTEXT, GAP, GOAL, MEANS, BENEFITS. Label each sentence.’* To establish a benchmark against which the model’s early outputs could be measured, the author first manually annotated the 10 pilot abstracts from C1. This reference annotation evaluated ChatGPT’s predictions during the prompt-development stages. When the initial zero-shot prompt was tested against this reference, overall sentence-level accuracy was 52%. Error analysis revealed two problems. First, the model frequently confused CONTEXT and GAP: for example,

sentences that the reference annotation identified as GAP were often mislabeled as CONTEXT. Second, the model under-annotated multi-move sentences, identifying only the primary move and missing the secondary one (in 89% of cases).

Stage 2: Initial few-shot prompt with basic examples. Taking into account the errors at Stage 1, a prompt was created incorporating clear instructions, precise explanations for each label, and simple examples. The prompt included one illustrative example per move and a general instruction to handle multiple moves. The Stage 2 prompt was tested on all 30 abstracts of C1. Error analysis against the reference annotation revealed three problems: CONTEXT/GAP confusion (in 23% of GAP sentences); Omission of TASKS in sentences with METHODS (in 68% of multi-move sentences containing both TASKS and METHODS); Failure to apply the <UNCLEAR> label.

Stage 3: Refined prompt with error corpus and enriched examples. To address these three problems, an error analysis was conducted. The author compared the Stage 2 outputs on all 30 abstracts of C1 against the reference annotation. All sentences where the model’s predicted labels did not exactly match the reference were extracted and compiled into a purpose-built error corpus (n = 142 sentences across the 30 abstracts). Each error was manually coded according to the type of mismatch, using a simple classification scheme: (a) *label substitution* (one move predicted instead of another), (b) *omission* (one or more moves missed in a multi-move sentence), or (c) *false application of the <UNCLEAR> label*. This coding revealed that three error types accounted for 87% of all mistakes: CONTEXT/GAP substitution (34%), omission of TASKS in sentences with METHODS (29%), and failure to apply <UNCLEAR> (24%). Based on these categories, five targeted refinements were introduced into the prompt (Table 4).

Table 4

Prompt refinements introduced at Stage 3 based on Stage 2 error analysis	
Problem observed	Refinement implemented
CONTEXT/GAP confusion	Added explicit contrastive examples showing CONTEXT (background knowledge) vs. GAP (missing information).
MEANS over-generalization	Split MEANS into TASKS (specific steps), METHODS (procedures/techniques), MATERIALS (data/resources).
Multiple moves missed	Instructed to label all moves in a sentence, with primary/secondary distinction using multiple labels.
Poor recognition of linguistic markers	Compiled a list of typical Russian markers for each move (e.g., «целью является» for GOAL, «материалом послужили» for MATERIALS).
Failure to mark ambiguous sentences	Reinforced instruction to use <UNCLEAR> for highly ambiguous cases.

Stage 4: Final two-shot prompt. The iterative process yielded a final prompt incorporating two illustrative examples per move, a complete list of linguistic markers, and explicit handling of multi-move sentences. The full prompt is provided in Appendix 1. The final prompt was tested on C2 (30 abstracts) before being applied to C3. Agreement between the model’s outputs and the author’s

reference annotation for C2 reached 86% at sentence level, a substantial improvement over Stage 1 (52%) and Stage 2 (71%).

3.5. Performance evaluation

To evaluate the annotation performance of ChatGPT compared to a human, a randomly selected subset of 20 abstracts with 260 sentences from C3 was used. The sample comprised half of the main test corpus (C3, $n = 40$), providing a representative cross-section of the data. This size was considered adequate for the comparative analysis, as it contained a sufficient number of instances for each of the seven rhetorical move categories to allow for reliable calculation of precision, recall, and F1 scores. It yields a sufficient number of rhetorical-move instances for reliable precision/recall calculation while remaining small enough to prevent annotator fatigue. Each sentence was fed individually into the ChatGPT model alongside the same standardized prompt. A human annotator provided with similar instructions as the LLM was involved in the study. The human annotator was a Candidate of Philological sciences with eight years of experience of teaching English for Academic Purposes. The comparison between ChatGPT' and the human's annotations was conducted by the author, who was not involved in the manual annotation process to avoid bias.

The effectiveness of the prompt was assessed on C3 comprising 40 abstracts with 480 sentences. To evaluate the effectiveness of the annotation process, sentence-level accuracy was used, calculated as the proportion of correctly annotated sentences relative to the total number of sentences. Every move present in the sentence was correctly identified with the correct label, and no extra moves were falsely assigned. Instance-level accuracy, in contrast, was defined as the proportion of individual move occurrences correctly identified across the entire corpus, regardless of whether the sentence as a whole was perfectly annotated. For the purposes of the present analysis, sentence-level accuracy is reported as the primary measure as it reflects the practical utility of the model for full-sentence annotation, while the Precision, Recall, and F1 scores are calculated at the instance level for each individual move category.

4. Results

This section presents a performance analysis of ChatGPT. Initially, the model was instructed by being provided with a developed and refined prompt (Appendix 1) and a sample of 40 abstracts from C3. Then, the accuracy of the annotated sentences generated by the model was assessed, considering an instance as accurately annotated only when all moves were correctly coded. As shown in Table 5, the results demonstrate a clear performance advantage for ChatGPT at the sentence level. Subsequently, the performance levels of the LLM and the human annotator were compared by the author.

Table 5

Instance-level accuracy obtained with ChatGPT-4.0	
	Measures
No. of correctly annotated sentences	422
Sentence-level accuracy (%)	88

For comparison with a human annotator, a subset of 20 abstracts (260 sentences) from C3 was used. Each sentence was processed by ChatGPT using the final two-shot prompt (see Appendix 1). For a comparative baseline, an independent human annotator, a Candidate of Philological Sciences with eight years of experience in teaching English for Academic Purposes, was involved. The human annotator was provided with an instruction that was functionally equivalent in content to the prompt. This instruction contained the definitions of all seven moves; the same lists of typical Russian linguistic markers for each move; the same illustrative examples; and the explicit instruction to label multiple moves within a single sentence when present, with the primary move listed first.

To establish a reference annotation against which both ChatGPT and the human annotator were evaluated, the author manually annotated all 260 sentences using the seven-label taxonomy. This reference annotation was created prior to and independently of both ChatGPT’s output and the human annotator’s work. The author’s annotation followed the same move definitions, linguistic markers, and multi-label rules as those later provided to both the model and the human annotator. The resulting datasets from both ChatGPT and the human annotator were assessed by the author.

To evaluate the effectiveness of the annotation process, two levels of accuracy were distinguished. Sentence-level accuracy was calculated as the proportion of sentences for which the annotation matched the reference in all respects, i.e., every move present in the reference was correctly identified with the correct label, and no extra moves were falsely assigned.

For each individual move label, class-wise (per-label) Precision, Recall, and F1-scores were calculated. They are derived from instance-level counts of True Positives (TP), False Positives (FP), and False Negatives (FN), where TP means that the annotator assigned a specific move label to a sentence, and that same label was present in the reference annotation for that sentence; FP means that the annotator assigned a move label to a sentence that was not present in the reference annotation for that sentence; and FN means that the reference annotation contained a move label for a sentence that the annotator failed to assign.

For sentences containing multiple moves, the instance-level rules were applied independently to each label. For example, if a reference sentence contained two labels, for example <GAP> <GOAL>, and ChatGPT predicted only <GOAL>, this counted as TP for GOAL, FN for GAP, and no FP (since no extra label was added). Precision for each label was calculated as $TP / (TP + FP)$, Recall as $TP / (TP + FN)$, and F1 as their harmonic mean. Macro- or micro-averaged scores across all moves are not reported, as this study focuses on category-specific performance rather than a single aggregate measure.

Table 6

Accuracy measures for 260 C3 sentences annotated by ChatGPT and the human annotator

Sentence-level performance	ChatGPT			Human annotator		
	Precision, %	Recall, %	F1, %	Precision, %	Recall, %	F1, %
CONTEXT	88.20	81.40	84.65	100.00	96.20	98.06
GAP	91.00	96.70	93.76	94.20	91.90	93.03
GOAL	100.00	100.00	100.00	92.86	85.95	89.27
TASKS	92.10	91.12	91.6	100.00	98.78	99.39
METHODS	97.40	89.10	93.07	100.00	88.10	93.68
MATERIALS	100.00	94.17	96.99	100.00	97.73	98.85
BENEFITS	100.00	98.10	99.04	100.00	100.00	100.00

Table 6 summarises the performance of ChatGPT and the human annotator. Both achieved F1 scores above 90% for most move categories, suggesting that the two-shot prompt and the functionally equivalent instruction provided a workable basis for annotation by the model and the expert, respectively. A closer look at individual categories, however, reveals differences. For CONTEXT, the human annotator outperformed ChatGPT across all measures, with near-perfect precision (100.00%) and recall (96.20%) against ChatGPT’s 88.20% and 81.40%, respectively. The model’s lower recall points to a tendency to confuse background statements with research gaps, a distinction that appears to require more contextual sensitivity than the current prompt can supply. GAP proved to be one area where ChatGPT had the edge. Its recall (96.70%) exceeded the human’s (91.90%), though the human achieved higher precision (94.20% vs 91.00%). The F1 scores were close (93.76 vs 93.03), indicating that while both performed well, ChatGPT was somewhat more thorough in identifying gaps. GOAL was the clearest case of machine advantage. ChatGPT scored 100% on all three metrics, compared to the human’s 92.86% precision, 85.95% recall, and 89.27% F1. For TASKS, the pattern reversed. The human annotator produced near-perfect scores (precision 100.00%, recall 98.78%, F1 99.39%), while ChatGPT lagged behind (92.10%, 91.12%, 91.60%). The model’s difficulty here appears to lie in distinguishing task statements from methods, particularly when the two occur in the same sentence where the boundaries between moves become blurred. METHODS was the only category where the two annotators achieved virtually identical F1 scores (93.07 for ChatGPT, 93.68 for the human). The human had perfect precision (100.00%) but slightly lower recall (88.10%) than ChatGPT (89.10%). MATERIALS and BENEFITS were identified with high accuracy by both. For MATERIALS, the human held a modest advantage (F1 98.85 vs 96.99), while for BENEFITS, both performed near-perfectly, with the human reaching 100% across all metrics and ChatGPT close behind (F1 99.04). It should be noted that despite being explicitly instructed to identify and mark potentially ambiguous sentences with a <UNCLEAR> label, ChatGPT did not implement this directive.

4.1. Overall assessment

An abstract annotated by ChatGPT and the human annotator using the two-shot prompt is shown in Table 7.

Table 7

An example of output generated by ChatGPT and the human annotator with the two-shot prompt

ChatGPT

<CONTEXT> Варьируясь в текстах разных авторов, дистрибуция лексики может вызывать сложности понимания. <GOAL> В рамках проекта впервые предлагается оценивать дискурсивные характеристики текста на основе следующих параметров: лексическое разнообразие, лексическая плотность и индекс исключительности. <GOAL> Проект нацелен на разработку типологии указанных параметров 4 типов текстов. <METHODS> Для расчета значений параметров будет написана программа, устанавливаемая на профайлер RuLingva. <TASKS> Предлагается выявить диагностические критерии для идентификации «норм», а также описать пограничные случаи. <METHODS> Оценка лексического разнообразия будет осуществляться на основе нормализованного классического параметра TTR, оценка плотности – на основе расчета доли полных слов в тексте, а индекс «исключительности» – на основе нормализованного количества однократных употреблений полных слов в тексте на 100 словоформ. <MATERIALS> В качестве материала будет использован Учебный корпус русского языка, аннотированный корпус русской детской литературы. <MATERIALS> Для внутриязыкового контрастирования выбраны документы международного права. <MATERIALS> Для валидации типологии лексических параметров в рамках проекта будет собран Дополнительный корпус, в который войдут тексты Российской электронной школы, Международные правовые акты и тексты, рекомендованные для внеклассного чтения. <BENEFITS> Типология лексических параметров может использоваться в лексикологии, дискурсии для описания и моделирования текстов. <BENEFITS> В прикладной лингвистике лексическое разнообразие, плотность и индекс исключительности найдут применение при оценке номинативной плотности текстов различных типов. <BENEFITS> Отдельную область применения составляет использование допустимых референтных значений лексических параметров текстов при выборе текста для целевой категории читателей. <BENEFITS> Результаты могут представлять интерес для диагностирования склонности личности к различным нестандартным сценариям поведения.

Human annotator

<CONTEXT> Варьируясь в текстах разных авторов, дистрибуция лексики может вызывать сложности понимания. <GAP><GOAL> В рамках проекта впервые предлагается оценивать дискурсивные характеристики текста на основе следующих параметров: лексическое разнообразие, лексическая плотность и индекс исключительности. <GOAL> Проект нацелен на разработку типологии указанных параметров 4 типов текстов. <TASKS><METHODS> Для расчета значений параметров будет написана программа, устанавливаемая на профайлер RuLingva. <TASKS> Предлагается выявить диагностические критерии для идентификации «норм», а также описать пограничные случаи. <METHODS> Оценка лексического разнообразия будет осуществляться на основе нормализованного классического параметра TTR, оценка плотности – на основе расчета доли полных слов в тексте, а индекс «исключительности» – на основе нормализованного количества однократных употреблений полных слов в тексте на 100 словоформ. <MATERIALS> В качестве материала будет использован Учебный корпус русского языка, аннотированный корпус русской детской литературы. <MATERIALS> Для внутриязыкового контрастирования выбраны документы международного права. <MATERIALS> <TASKS> Для валидации типологии лексических параметров в рамках проекта будет собран Дополнительный корпус, в который войдут тексты Российской электронной школы, Международные правовые акты и тексты, рекомендованные для внеклассного чтения. <BENEFITS> Типология лексических параметров может использоваться в лексикологии, дискурсии для описания и моделирования текстов. <BENEFITS> В прикладной лингвистике лексическое разнообразие, плотность и индекс исключительности найдут применение при оценке номинативной плотности текстов различных типов. <BENEFITS> Отдельную область применения составляет использование допустимых референтных значений лексических параметров текстов при выборе текста для целевой категории читателей. <BENEFITS> Результаты могут представлять интерес для диагностирования склонности личности к различным нестандартным сценариям поведения.

The analysis demonstrated that ChatGPT correctly labelled sentences containing one rhetorical move. However, it failed to identify the second move in the sentence. For instance, the following sentence contains two moves — GAP and GOAL — correctly identified by the human annotator. ChatGPT captured only GOAL.

- (1) *В рамках проекта впервые предлагается оценивать дискурсивные характеристики текста на основе следующих параметров: лексическое разнообразие, лексическая плотность и индекс исключительности.*

[The project proposes for the first time to assess the discursive characteristics of a text based on the following parameters: lexical diversity, lexical density and the index of exclusivity.]

The example shows that ChatGPT is not always able to fully grasp the rhetorical complexity of academic writing. It can identify the stated objective (GOAL) but fails to recognize the motivation or justification (GAP).

The following sentence contains TASKS and METHODS moves. Again, unlike the human annotator, ChatGPT identified only one move — METHODS.

- (2) *Для расчета значений параметров будет написана программа, устанавливаемая на профайлер RuLingva.*
[To calculate the parameter values, a program will be written that will be installed on the RuLingva profiler.]

ChatGPT's failure to recognize TASKS could be explained by the fact that AI models are often best at identifying explicit statements and may struggle to infer implicit meanings. The TASKS move is implied in the sentence. Additionally, the way the sentence is structured, with the primary focus on the "program," might have led the AI to focus on the methodological aspect.

4.2. Assessment of individual labels

CONTEXT: According to the comparison conducted by the author, ChatGPT has higher precision (88.20%) than recall (81.40%), suggesting that it identifies many sentences as CONTEXT, but a larger proportion of them were actually GAPS. Here is an example of GAP identified as CONTEXT.

- (3) *В связи с этим, важной проблемой переводоведения является комплексное описание передовых технологий нейронного машинного перевода, включая такие аспекты данной проблемы, как объективная оценка качества перевода, выполняемого системами ИИ с усовершенствованными нейросетевыми алгоритмами.*
[In this regard, an important problem in translation studies is a comprehensive description of advanced technologies of neural machine translation, including aspects of this problem such as an objective assessment of the quality of translation by AI systems with advanced neural network algorithms.]

The human annotator achieves near-perfect accuracy here, indicating that identifying CONTEXT sentences might require more nuanced understanding than ChatGPT has.

GAP: ChatGPT has very high recall (96.70%), but slightly lower precision (91.00%), suggesting that it is good at finding most of the GAP sentences, but it incorrectly labels some sentences as GAP. Below is an example of CONTEXT identified as GAP:

- (4) *Образ будущего страны и состояние российского общества в скором времени зависят от молодежи, поэтому результаты исследований сформированных у нее установок являются основой для проведения и совершенствования эффективной внутренней молодежной политики, что обуславливает АКТУАЛЬНОСТЬ проекта.*
[The image of the future of the country and the state of Russian society will soon depend on the youth. Therefore, the results of research into their attitudes are the basis for implementing and improving an effective national youth policy, determining the RELEVANCE of the project.]

The precision and recall of the human annotator are more balanced, showing a more nuanced capacity to identify GAP sentences.

GOAL: ChatGPT achieves the highest scores (100%) for precision, recall, and F1. This could indicate that GOAL sentences are very clearly defined and easily identifiable based on their lexical markers (*планируется, имеет цель, целью является, цель заключается*). The lower scores of the human annotator could indicate that the humans are adding more complex understanding when it may not exist.

TASKS: ChatGPT's recall (91.12%) shows it captures not all actual TASK sentences and its slightly lower precision (92.10%) suggests some false positives. Here is an example of the Goal move marked as the TASK move:

- (5) *При реализации проекта планируется создание Лаборатории перспективных лингвополитологических исследований на базе СПбПУ.*

[The project implies the creation of a laboratory of advanced linguistic and political studies at SPbPU.]

The human annotator again shows a very high level of performance.

METHODS: ChatGPT has a respectable precision (97.40%), but a lower recall (89.10%). This suggests that while it is usually correct when it identifies a sentence as a METHOD, it misses some actual METHOD sentences. For example, the next sentence was not identified as METHOD move. ChatGPT marked it as TASK.

- (6) *Кейсы блогеров планируется рассмотреть с позиций лингвоаксиологии (1) в сопоставлении друг с другом как совокупность неодинаковых ценностных конфигураций, отраженных в их текстах, и (2) в единстве, обусловленном общностью лингвокультуры.*

[The bloggers' cases will be considered in terms of linguaxiology (1) in comparison with each other as a set of different value configurations reflected in their texts, and (2) in unity determined by shared linguistic culture.]

The human annotator has a higher recall, suggesting they are better at identifying all METHOD sentences.

MATERIALS and BENEFITS: Both ChatGPT and the human annotator have very high scores across all metrics, indicating that these sentence types are relatively easy to identify.

5. Discussion

To explore the efficiency of using LLMs for identifying rhetorical moves in grant proposal abstracts, a comparative study involving ChatGPT-4 and a human annotator experienced in academic writing was conducted. The findings of the study revealed that ChatGPT can be effectively trained to correctly identify rhetorical moves through clear definitions, linguistic markers, and examples of each rhetorical move. The study showed that the more accurate and clearer the definitions of the moves, the more accurate ChatGPT's rhetorical move analysis became. Training the model with illustrative examples of the moves also contributed to correctly recognizing the linguistic markers associated with each

move. Instructing the model to perform sentence-level analysis helped significantly reduce the tendency to overlook several moves in one sentence. It was also found that experimenting with different prompts and evaluating the results is crucial to improve ChatGPT's performance.

The results of the present study demonstrate that the two-shot prompt was effective enough in instructing the LLM. ChatGPT showed high accuracy in identifying the rhetorical moves in grant proposal abstracts, with F1 scores mostly above 90%. Most correctly identified moves were GOAL, MATERIALS, and BENEFITS, which suggests that the linguistic markers of these moves are clear and make them recognizable by the model. CONTEXT, GAP, and TASKS exhibited lower performance, which might be due to the fact that these categories have more diverse or less clear linguistic markers, requiring a deeper analysis of the context, which is a challenge for the LLM. The analysis also revealed that ChatGPT experienced difficulties in identifying multiple rhetorical moves within one sentence, effectively labeling only one of the moves.

These difficulties can be attributed either to limitations of LLMs or to other factors. First, the observed errors may stem from imperfections in the prompt design to recognize implicit meanings or multiple moves. The final prompt included an instruction to handle multi-move sentences and provided one example with two moves (<GAP> <GOAL>). However, it did not provide examples of sentences containing three or more distinct moves, nor did it illustrate how to order multiple tags in such cases. It is therefore plausible that more diverse examples would have improved accuracy. Second, some of the discrepancies between ChatGPT and the human annotator may be explained by vagueness in the source texts, which suggests that the model's performance can be improved by including examples of implicitly expressed MOVES in the prompt, not just those with clear linguistic markers. For example, while *актуальность* / 'relevance' is a positive statement of why the topic matters, some grant applicants formulate it as a GAP, i.e., they state what has not yet been studied (*Актуальность исследования сложного предложения на материале собственно карельского, ливвиковского и людиковского наречий обусловлена тем, что синтаксис карельского языка ... в целом исследован гораздо менее детально, чем морфологический, лексический и фонетический уровни* / 'The relevance of the study of a complex sentence based on the material of the Karelian, Livvik and Ludic dialects is due to the fact that the syntax of the Karelian language ... has been studied in much less detail than the morphological, lexical and phonetic levels'). In such cases, the sentence explicitly signals a missing piece of knowledge, which would be labeled as a GAP, however it contains a CONTEXT marker *актуальность* / 'relevance'. This variation creates ambiguity for ChatGPT. Future prompt designs should include examples of both formulations to help the model distinguish them.

Third, the cultural-rhetorical specificity of the training data must be considered when assessing the model performance. ChatGPT-4 was pre-trained predominantly on English-language texts, which tend to follow a relatively linear move structure (e.g., CONTEXT → GAP → GOAL → MEANS → BENEFITS). Russian grant

proposal abstracts, however, sometimes deviate from this prototype. For example, the following abstract contains two explicit statements of the GOAL (sentences 1 and 2) and places the CONTEXT and the GAP after the second GOAL rather than before it: *Проект направлен на изучение влияния сакрализации и ресакрализации пространства на идентичность арабского населения Палестины... Целью проекта является выявление роли сакрального пространства... Актуальность проекта определяется значением святынь... Несмотря на обилие работ... отсутствуют исследования... / 'The project aims to study the influence of sacralization and resacralization of space on the identity of the Arab population of Palestine... The goal of the project is to identify the role of sacred space... The relevance of the project is determined by the significance of holy sites... Despite the abundance of works... research is lacking...'). This rhetorical organization (repeated GOALS and delayed CONTEXT and GAP) may create a challenge for AI. A human annotator familiar with this culture-specific feature recognizes these variations as acceptable. ChatGPT, however, expects a more linear, non-repetitive structure. This example illustrates that the model's difficulties may be due to a mismatch between its training distribution (Anglophone rhetorical norms) and the actual organization of Russian grant proposal abstracts. Adapting prompts to include examples of such culturally specific deviations would likely improve performance. These findings resonate with Ponton and Mantello's (2026) critical discussion of AI and authenticity: they argue that LLMs, trained on massive but culturally homogenous datasets, often lack the semiotic agency required to interpret culturally embedded texts, producing instead outputs that reflect the statistical norms of their training data.*

This study thus provides some theoretically significant insights into discourse analysis. First, it shows how LLMs may be trained to understand rhetorical moves using clear definitions, linguistic markers, and informative examples. This supports discourse-analytic approaches that see rhetorical structures as identifiable through linguistic features (Swales 1990, Hyland 2020). Second, while ChatGPT achieved high accuracy in identifying the GOAL, MATERIALS, and BENEFITS moves, it experienced difficulties in recognizing context-dependent moves, such as CONTEXT, GAP, TASKS, suggesting that some rhetorical functions require contextual interpretation, which aligns with van Dijk's (1977) theory of context. However, these difficulties may be mitigated by prompt refinement or adaptation to non-English rhetorical norms. Third, the comparison between human and AI annotation highlights the importance of considering text-internal ambiguity (vagueness of phrasing) and text-external factors (cultural training data) when interpreting model errors. Fourth, the human annotator's better performance in identifying multiple moves within a single sentence highlights a significant limitation of the LLM: difficulties it experiences in handling polyfunctional discourse structures. This supports Fairclough's approach that emphasizes the dynamic, social practice-dependent nature of text (Fairclough 1992).

6. Conclusion

The study assessed the efficacy of using ChatGPT for annotating rhetorical moves in grant proposal abstracts. The research revealed that two-shot instructions can effectively guide ChatGPT in this process. It was also shown that AI demonstrated a different degree of accuracy in identifying different types of moves. While some rhetorical moves were easily identifiable, others posed some challenges to the LLM.

The key finding of this study indicates that the model's performance is highly contingent on three interrelated factors: 1) the clarity and completeness of the prompt design, 2) the degree of explicitness of rhetorical moves in the source texts, and 3) the alignment between the model's training data and the rhetorical conventions of the target genre. Consequently, despite its high performance, ChatGPT should be viewed as a supportive tool for move analysis rather than a substitute for human expertise. A hybrid approach, in which the AI performs initial annotation and the human annotator reviews and corrects ambiguous cases, represents the most valid and efficient methodology. The finding also suggests that the internationalization of AI-powered discourse tools requires adaptation to diverse rhetorical traditions. An equally important finding is that ChatGPT's difficulties may be eliminated through prompt engineering and fine-tuning.

Despite its useful findings and implications, the study has some limitations. First, the research drew on data from only one field of knowledge — the humanities, and its findings may not apply to other fields. Second, future research could expand the sample size of analyzed texts to verify the conclusions made in the present research. Second, the analysis relies on just one LLM — ChatGPT, specifically the gpt-4o-2024-05-13 version — and one human expert. This means we cannot determine whether the observed differences are specific to this particular model iteration or to this particular annotator. Different LLMs (e.g., Claude, Gemini, DeepSeek) might yield different annotation patterns, and a different expert might produce different benchmark results. Moreover, the use of a single human annotator precludes the calculation of inter-annotator reliability, which would have strengthened the validity of the reference annotation. Third, we did not examine the variability of ChatGPT's outputs across repeated runs of the same prompts. Given the model's stochastic nature submitting the same sentence multiple times could yield different move assignments. Fourth, the study assumes homogeneity of rhetorical move structures across the five included humanities subdisciplines. While this assumption was motivated by prior cross-disciplinary evidence, it was not empirically tested through subdisciplinary comparative analysis within our corpus. It is therefore possible that disciplinary conventions affect move distribution in ways not captured here, and this should be treated as a limitation of the present design. Fourth, while this study focused only on grant proposal abstracts, exploring AI identification and annotation performance for other academic genres would illuminate how genre conventions influence the annotation results. Including a different number of examples in the prompt could also reveal their impact on the model's performance. The analysis could also be extended by involving other LLMs.

References

- Afful, Joseph, Eugene K. Hesse, Ebenezer Agbaglo & Emmanuel Bonsu. 2022. Move analysis of grant recommendation letters written by senior academics in a Ghanaian university. *Taiwan International ESP Journal* 13 (2). 39–74. [https://doi.org/10.6706/TIESPJ.202212_13\(2\).0002](https://doi.org/10.6706/TIESPJ.202212_13(2).0002)
- Alyousef, Hesham S. 2021. Structure of research article abstracts in political science: A genre-based study. *SAGE Open* 11 (3). 1–8. <https://doi.org/10.1177/21582440211040797>
- Baltezarević, Radoslav, Lazar Stošić & Olga Mikhailova. 2026. AI-assisted academic writing: Balancing linguistic enhancement with legal and ethical oversight. *Russian Journal of Linguistics* 30 (1). 105–123. <https://doi.org/10.22363/2687-0088-44754>
- Basturkmen, Helen. 2012. A genre-based investigation of discussion sections of research articles in Dentistry and disciplinary variation. *Journal of English for Academic Purposes* 11 (2). 134–144. <https://doi.org/10.1016/j.jeap.2011.10.004>
- Bhatia, Vijay. 1993. *Analysing Genre: Language Use in Professional Settings*. Longman.
- Bhatia, Vijay. 2008. Genre analysis, ESP and professional practice. *English for Specific Purposes* 27 (2). 161–174.
- Biber, Douglas, Ulla Connor & Thomas Upton. 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. John Benjamins Publishing.
- Boginskaya, Olga. 2025a. Rhetoric of promotion and persuasion in grant proposal abstracts. *Integration of Education* 29 (3). 555–572. <https://doi.org/10.15507/1991-9468.029.202503.555-572>
- Boginskaya, Olga. 2025b. Disciplinary influences on research gap identification. *Research Result. Theoretical and Applied Linguistics* 11 (1). 50–71. <https://doi.org/10.18413/2313-8912-2025-11-1-0-3>
- Charles, Maggie & Karin Whiteside. 2024. Seeking research funding in a peripheral context: A learner corpus genre study of grant proposal summaries. *Journal of English for Academic Purposes* 71. 101431. <https://doi.org/10.1016/j.jeap.2024.101431>
- Connor, Ulla. 2000. Variation in rhetorical moves in grant proposals of U.S. humanists and scientists. *Text* 20 (1). 1–28. <https://doi.org/10.1515/text.1.2000.20.1.1>
- Connor, Ulla & Anna Mauranen. 1999. Linguistic analysis of grant proposals: European Union research grants. *English for Specific Purposes* 18. 47–62. [https://doi.org/10.1016/S0889-4906\(97\)00026-4](https://doi.org/10.1016/S0889-4906(97)00026-4)
- Cotos, Elena, Sarah Huffman & Stephanie Link. 2017. A move/step model for methods sections: Demonstrating rigor and credibility. *English for Specific Purposes* 46. 90–106. <https://doi.org/10.1016/j.esp.2017.01.001>
- Crookes, Graham. 1986. Towards a validated analysis of scientific text structure. *Applied Linguistics* 7 (1). 57–70.
- de Groot, Elizabeth. 2008. *English Annual Reports in Europe*. LOT.
- Dijk, Teun A. van. 1977. *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*. Longman.
- Fairclough, Norman. 1992. Discourse and text: Linguistic and intertextual analysis within discourse analysis. *Discourse & Society* 3 (2). 193–217.
- Feng, Dezheng. 2019. Interdiscursivity, social media and marketized university discourse: A genre analysis of universities' recruitment posts on WeChat. *Journal of Pragmatics* 143. 121–134.
- Feng, Hui. 2006. A corpus-based study of research grant proposal abstracts. *Perspectives: Working Papers in English and Communication* 17 (1). 1–24. https://www.researchgate.net/publication/228628027_A_corpusbased_study_of_research_grant_proposal_abstracts
- Feng, Hui & Ling Shi. 2004. Genre analysis of research grant proposals. *LSP and Professional Communication* 4 (1). 8–32.
- Flowerdew, Lynne. 2016. A genre-inspired and lexico-grammatical approach for helping postgraduate students craft research grant proposals. *English for Specific Purposes* 42. 1–12. <https://doi.org/10.1016/j.esp.2015.10.001>

- Gilardi, Fabrizio, Meysam Alizadeh & Maël Kubli. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences* 120 (30). e2305016120. <https://doi.org/10.1073/pnas.2305016120>
- Gray, Bethany, Elena Cotos & Jordan Smith. 2020. Combining rhetorical move analysis with multidimensional analysis. In Ute Römer, Viviana Cortes & Eric Friginal (eds.), *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise*, 137–168. John Benjamins.
- Guangyuan, Yao & Liu Zhaoxia. 2025. GPT as book reviewer: A move and syntactic complexity analysis of GPT-generated versus scholar-written academic book reviews. *Journal of English for Academic Purposes* 76. 101533. <https://doi.org/10.1016/j.jeap.2025.101533>
- Hu, Guangwei & Yongbing Liu. 2018. Three-minute thesis presentations: A cross-disciplinary study of genre moves. *Journal of English for Academic Purposes* 35. 16–30. <https://doi.org/10.1016/j.jeap.2018.06.004>
- Hyland, Ken. 2004. *Disciplinary Discourses: Social Interactions in Academic Writing*. University of Michigan Press.
- Hyland, Ken. 2020. Genre and discourse analysis in language for specific purposes. In Carol A. Chapelle (ed.), *The concise encyclopedia of applied linguistics*, 507–513. Singapore: Wiley Blackwell.
- Hyland, Ken. 2024. Genre-based instruction and corpora. *TESOL Quarterly* 58 (3). Article 3. <https://doi.org/10.1002/tesq.3273>
- Hyland, Ken & Polly Tse. 2012. ‘She has received many honours’: Identity construction in article bio statements. *Journal of English for Academic Purposes* 11 (2). 155–165. <https://doi.org/10.1016/j.jeap.2012.01.001>
- Jiang, Feng & Ken Hyland. 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>
- Khedri, Mohsen, Eatidal Hasan & Konstantinos Kritsis. 2022. Rhetorical structure and persuasive features of advertising: An intercultural analysis of English and Arabic online advertisements. *Russian Journal of Linguistics* 26 (3). 596–624. <https://doi.org/10.22363/2687-0088-24591>
- Kim, Minjin & Xiaofei Lu. 2024. Exploring the potential of using ChatGPT for rhetorical move-step analysis: The impact of prompt refinement, few-shot learning, and fine-tuning. *Journal of English for Academic Purposes* 78. 101422. <https://doi.org/10.1016/j.jeap.2024.101422>
- Kuzman, Taja, Igor Mozetic & Nikola Ljubescic. 2023. Automatic genre identification for robust enrichment of massive text collections: Investigation of classification methods in the era of large language models. *Machine Learning and Knowledge Extraction* 5 (3). Article 3. <https://doi.org/10.3390/make5030059>
- Liu, Huifeng, Lilliatl Ismail & Noraini Ahmad. 2024. Rhetorical move-step analysis of argumentative essays by Chinese EFL undergraduate students. *Theory and Practice in Language Studies* 14 (7). 2063–2072. <https://doi.org/10.17507/tpls.1407.13>
- Locke, Lawrence F., Waneen W. Spirduso & Stephen J. Silverman. 2014. *Proposals that Work*. SAGE.
- Matzler, Peter. 2021. Grant proposal abstracts in science and engineering: A prototypical move-structure pattern and its variations. *Journal of English for Academic Purposes* 49. 100938. <https://doi.org/10.1016/j.jeap.2020.100938>
- Millar, Neil, Brian Batalo & Brian Budgell. 2022. Trends in the use of promotional language (hype) in abstracts of successful National Institutes of Health grant applications, 1985–2020. *JAMA Network Open* 5 (8). e2228676. <https://doi.org/10.1001/jamanetworkopen.2022.28676>
- Myers, Greg. 1990. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. University of Wisconsin Press.
- Ozyumenko, Vladimir I. & Tatiana V. Larina. 2025. Artificial intelligence in translation: Advantages and limitations. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics] 24 (1). 117–130. <https://doi.org/10.15688/jvolsu2.2025.1.10>

- Ponton, Douglas & Peter Mantello. 2026. Simulated selves: Creativity, authenticity semiotic agency in AI companionship. *Russian Journal of Linguistics* 30 (1). 124–144. <https://doi.org/10.22363/2687-0088-42427>
- Russian Science Foundation [Electronic resource]. <http://www.rsf.ru> (accessed 25 September 2025).
- Swales, John M. 1981. *Aspects of Article Introductions. Language Studies Unit*. University of Aston.
- Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Tardy, Christine. 2011. ESP and multi-method approaches to genre analysis. In Diane Belcher, Ann Johns & Brian Paltridge (eds.), *New directions in English for specific purposes research*, 145–173. University of Michigan Press.
- van Herck, Ruben, Sofie Decock & Birgit Fastrich. 2022. A unique blend of interpersonal and transactional strategies in English email responses to customer complaints in a B2C setting: A move analysis. *English for Specific Purposes* 65. 30–48. <https://doi.org/10.1016/j.esp.2021.08.001>
- Vergaro, Carla. 2004. Discourse strategies of Italian and English sales promotion letters. *English for Specific Purposes* 23 (2). 181–207.
- Wang, Ying. 2025. Examining promotional strategies and trends in successful grant application abstracts in humanities: Moves and appraisal resources. *English for Specific Purposes* 78. 70–84. <https://doi.org/10.2139/ssrn.4837093>
- Wu, Ying. 2020. A corpus-based comparative genre analysis of CEO's letters from Chinese and American companies. *Journal of Applied Linguistics and Language Research* 7 (3). 1–13.
- Yu, Dengsheng. 2022. *Cross-cultural Genre Analysis: Investigating Chinese, Italian and English CSR Reports*. Routledge.
- Yu, Dengsheng. 2025. Towards LLM-assisted move annotation: Leveraging ChatGPT-4 to analyse the genre structure of CEO statements in corporate social responsibility reports. *English for Specific Purposes* 78. 33–49.
- Yu, Dengsheng & Marina Bondi. 2017. The generic structure of CSR reports in Italian, Chinese, and English: A corpus-based analysis. *IEEE Transactions on Professional Communication* 60 (3). 273–291. <https://doi.org/10.1109/TPC.2017.2702040>
- Yu, Dengsheng, Marina Bondi & Ken Hyland. 2024. Can GPT-4 learn to analyse moves in research article abstracts? *Applied Linguistics*. Article amae071. <https://doi.org/10.1093/applin/amae071>

Appendix 1. Final two-shot prompt used for rhetorical move annotation

Please learn the following contents. The move structure of the grant proposal abstract refers to the categorical composition including the following moves: CONTEXT, GAP, GOAL, TASKS, METHODS, MATERIALS, or BENEFITS.

The following defines and illustrates the rhetorical moves commonly found in the genre of grant proposal abstract.

CONTEXT

Description: Provides background information, describes a real-world problem relevant to the research, or presents established scientific knowledge in the area being studied.

Linguistic markers: *В настоящее время...* , *На сегодняшний день ...* , *В последние годы...* , *Известно, что ...* , *В современной науке...* , *Актуальность данной проблемы обусловлена ...* , *Существующие исследования показывают...* , *возросла потребность в...*

Examples:

<CONTEXT> Варьируясь в текстах разных авторов, дистрибуция лексики может вызывать сложности понимания и служить мерой успешности речевого акта

<CONTEXT> На сегодняшний день искусственный интеллект играет важную роль в развитии переводческой отрасли.

GAP

Description: Identifies a missing piece of information, a problem, or a limitation in current knowledge or existing solutions.

Linguistic markers: *Однако, недостаточно изучены... , Несмотря на многочисленные исследования, остается неясным вопрос о... , В настоящее время отсутствуют данные о... , Проблема [явление] остается малоизученной ..., Существующие методы имеют ряд недостатков... , Ограничением существующих исследований является..., Данные подходы не позволяют в полной мере решить проблему...*

Examples:

<GAP> При этом изучение процессов моделирования образа России в средствах массовой информации не получило полноценного научного освещения.

<GAP> Новизна исследования обусловлена не изученным ранее материалом - медиаконтентом органов исполнительной власти РФ.

GOAL

Description: Clearly states the main objective or aim of the research project.

Linguistic markers: *Целью данной работы является... , Данное исследование направлено на..., Основной задачей исследования является..., В рамках данной работы планируется ..., Предметом исследования является..., Исследование имеет целью...*

<GOAL> Целью исследования является исследование проблемы доступности и понятности текстов официальных документов во взаимодействии государственных органов и граждан.

<GOAL> Цель проекта — разработать модель интердискурсивного лингвополитологического анализа стратегий продвижения национальных интересов России.

TASKS

Description: Lists the specific steps or sub-objectives that will be undertaken to achieve the overall goal.

Linguistic markers: *Enumeration markers Во-первых..., Во-вторых... , Задачи исследования включают в себя: ...; action verbs in the infinitive form: Провести анализ..., Осуществить синтез..., Разработать модель ..., Экспериментально проверить..., Сопоставить полученные данные...*

<TASKS> Реализация проекта предполагает решение следующих задач: 1. Проведение масштабного социолингвистического эксперимента. 2. Формирование списка перцептивно сложных явлений.

<TASKS> Комплексное исследование когнитивных механизмов и дискурсивных стратегий, сопряженных с осмыслением и преодолением социокультурных угроз, планируется осуществить в нескольких аспектах: создание корпуса исторических и современных текстов, репрезентирующих социокультурные риски; определение и типология базовых социокультурных концептов, являющихся когнитивной основой дискурсивной манифестации угроз; выявление и типология доминирующих нарративных стратегий, моделирующих представления о происхождении, смысле и возможных последствиях предполагаемых опасностей.

METHODS

Description: Explains the procedures and techniques that will be used to conduct the research and achieve the tasks.

Linguistic markers: *В работе использованы следующие методы: ..., Для достижения поставленной цели применялись..., Исследование проводилось с использованием ..., предполагается применить комплексную методологию*

<METHODS> При решении поставленной задачи предполагается применить комплексную методологию современного дискурс-анализа

<METHODS> Это потребует совершенствования мультидисциплинарных способов анализа исторических и современных источников, основанных на совмещении методов когнитивной лингвистики, исторического анализа дискурса, нарратологии и нейросемантики.

MATERIALS

Description: Describes the resources, data, or equipment that will be used in the research.

Linguistic markers: *Материалом для исследования послужили... , В качестве данных использовались..., Для проведения эксперимента применялось следующее оборудование: ...*

<MATERIALS> Кейсы блогеров планируется рассмотреть с позиций лингвоаксиологии.

<MATERIALS> Материалом станут данные масштабного социолингвистического опроса.

BENEFITS

Description: Describes the potential positive outcomes and real-world applications of the research.

Linguistic markers: *Результаты данного исследования могут быть использованы для..., Практическая значимость работы заключается в..., Теоретическая значимость исследования состоит в..., Разработка новых методов..., Совершенствование существующих технологий..., Повышение эффективности..., Оптимизация процессов..., В дальнейшем результаты исследования могут быть использованы для...*

<BENEFITS> Результаты могут представлять интерес для диагностирования склонности личности к различным нестандартным сценариям поведения.

<BENEFITS> Полученные данные могут быть полезны при разработке молодежной политики.

Using this instruction, label each sentence in the text with <CONTEXT>, <GAP>, <GOAL>, <TASKS>, <METHODS>, <MATERIALS>, <BENEFITS>. If one sentence contains more than one move, identify all moves present. Label the primary move first, followed by the secondary move(s), using the same tags. For example: <GAP> <GOAL>. If a sentence is highly ambiguous, you may assign an <UNCLEAR> label.

The linguistic markers provided as examples are not limited to these lexical items. Here is an example of the move recognition: <CONTEXT> *Варьируясь в текстах разных авторов, дистрибуция лексики может вызывать сложности понимания.* <GOAL> *В рамках проекта предлагается оценивать дискурсивные характеристики текста.*

Article history:

Received: 05 September 2025

Accepted: 15 April 2026

Bionote:

Olga A. BOGINSKAYA is Doctor Habil., Professor at the Department of Foreign Languages, Irkutsk National Research Technical University. Her research interests lie in the field of professional and academic discourse studies. She has publications in peer-reviewed academic journals in the field of linguistics such as *Discourse Studies*, *Journal of English for Academic Purposes*, *ESP Today*, *Revista de Llengua et Drett*, *Slavia Centralis*, among others.

e-mail: olgaa_boginskaya@mail.ru

<https://orcid.org/0000-0002-9738-8122>

Сведения об авторе:

Ольга Александровна БОГИНСКАЯ — доктор филологических наук, профессор кафедры иностранных языков Института лингвистики и межкультурной коммуникации Иркутского национального исследовательского технического университета. Область исследовательских интересов — профессиональный и академический дискурс. Имеет публикации в ряде высокорейтинговых лингвистических журналов: *Discourse Studies*, *Journal of English for Academic Purposes*, *ESP Today*, *Revista de Llengua et Drett*, *Slavia Centralis* и др.

e-mail: olgaa_boginskaya@mail.ru

<https://orcid.org/0000-0002-9738-8122>