

# RAAMove: A Corpus for Analyzing Moves in Research Article Abstracts

Hongzheng Li<sup>\*</sup>, Ruojin Wang<sup>\*</sup>, Ge Shi<sup>†</sup>, Xing Lv<sup>\*</sup>, Lei Lei<sup>\*</sup>, Chong Feng<sup>‡</sup>,  
Fang Liu<sup>\*</sup>, Jinkun Lin<sup>‡</sup>, Yangguang Mei<sup>†</sup>, Lingnan Xu<sup>‡</sup>

<sup>\*</sup>School of Foreign Languages,

Key Laboratory of Language, Cognition and Computation Ministry of Industry and Information Technology,  
Beijing Institute of Technology

<sup>†</sup>Faculty of Information Technology, Beijing University of Technology

<sup>‡</sup>School of Computer Science, Beijing Institute of Technology

<sup>\*</sup>8, Liangxiang East Road, Fangshan District, 102488, Beijing

<sup>‡</sup>5, Zhongguancun South St., Haidian District, 100081, Beijing, China

<sup>†</sup>100, Pingleyuan, Chaoyang District, 100124, Beijing, China

{lihongzheng,wangruojin,lvxing,leilei,fengchong,liufang,ljkun,xln}@bit.edu.cn

{shige,meiyangguang}@bjut.edu.cn

## Abstract

Move structures have been studied in English for Specific Purposes (ESP) and English for Academic Purposes (EAP) for decades. However, there are few move annotation corpora for Research Article (RA) abstracts. In this paper, we introduce RAAMove, a comprehensive multi-domain corpus dedicated to the annotation of move structures in RA abstracts. The primary objective of RAAMove is to facilitate move analysis and automatic move identification. This paper provides a thorough discussion of the corpus construction process, including the scheme, data collection, annotation guidelines, and annotation procedures. The corpus is constructed through two stages: initially, expert annotators manually annotate high-quality data; subsequently, based on the human-annotated data, a BERT-based model is employed for automatic annotation with the help of experts' modification. The result is a large-scale and high-quality corpus comprising 33,988 annotated instances. We also conduct preliminary move identification experiments using the BERT-based model to verify the effectiveness of the proposed corpus and model. The annotated corpus is available for academic research purposes and can serve as essential resources for move analysis, English language teaching and writing, as well as move/discourse-related tasks in Natural Language Processing (NLP).

**Keywords:** move structure, corpus, research article, abstracts

## 1. Introduction

Effective communication in research articles relies on the use of "moves", which are discursal semantic units serving specific functions in written or spoken discourse, and linked to the writer's purpose (Swales, 1990, 2004). A move represents text frames that have their own meaning and function (Connor U., 2007).

In the realm of scientific research, Research Articles (RAs) play an important role in fostering academic exchange and promoting scientific progress. Within the RA, the abstract serves as an indispensable section, conveying the core ideas and highlights of the entire article in a concise yet convincing manner. Good abstracts should include well-organized move structures to promote the importance and values of the article. Just as Hyland (2007) argued, "One way that writers claimed significance was by opening their abstracts with a promotional statement."

Here is an abstract example<sup>1</sup> from the field of Natu-

ral Language Processing (NLP). The abstract contains clear and basic moves in square brackets, making the abstract highly readable and effectively showing the author's ideas and contributions.

**[Background]** Large-scale pretrained language models have achieved SOTA results on NLP tasks. **[Gap]** However, they have been shown vulnerable to adversarial attacks especially for logographic languages like Chinese. **[Purpose]** In this work, we propose RoCBert: a pretrained Chinese Bert that is robust to various forms of adversarial attacks like word perturbation, synonyms, typos, etc. **[Method]** It is pretrained with the contrastive learning objective which maximizes the label consistency under different synthesized adversarial examples. .... **[Result]** Across 5 Chinese NLU tasks, RoCBert outperforms strong baselines under three blackbox adversarial algorithms without sacrificing the performance on clean testset.

Move analysis has great potential for instructing, assessing abstract writing, aiding English for Specific Purposes (ESP) and English for Academic Purposes (EAP) learners and writers (Hyland, 2008;

<sup>△</sup>The first two authors contributed equally to this work.

<sup>1</sup>The abstract example is from Su et al., ACL 2022

Moreno and Swales, 2018; Swales, 2019), which has attracted global attention for decades, with numerous studies investigating moves, grammatical and linguistic features in sections of RAs such as Introduction and Method (Cotos et al., 2017; Lu et al., 2021; Alsharif, 2023). Some previous studies have also focused on moves and linguistic forms in abstracts (Pho, 2008; Suntara and Usaha, 2013; Darabad, 2016), and some work on move detection and recognition (Dayrell et al., 2012; Ding et al., 2019).

However, previous studies have predominantly focused on the analysis of moves within specific domains or journals, with few addressing multi-disciplinary RAs. For example, some works have discussed the rhetorical moves and their variations in Applied Linguistics journals (Pho, 2008; Samar et al., 2014; Yoon and Casal, 2020). To the best of our knowledge, there is a limited amount of research on the automated annotation of discourse moves in academic writing, and there is also a scarcity of large-scale annotated corpora specifically designed for move annotation within Research Article (RA) abstracts. This insufficiency poses a barrier to meeting the growing demand for comparative research in move analysis across diverse domains and tasks in Natural Language Processing (NLP), including the automatic identification of discourse moves and Computer-Assisted Language Learning (CALL).

To address this research gap, our long-term goal, which centers on intelligent English writing assistance and text generation for intelligent and digital language education, supports our mid-term goal of automatic move structure identification. In this pursuit, we introduce the RAAMove<sup>2</sup>, a large-scale multi-domain corpus of MOVE structures in the RA Abstracts from three scientific disciplines: Artificial Intelligence (AI), Communication Engineering (CE), and Mechanical Engineering (ME), where the latter two can be collectively referred to as Engineering. The corpus is constructed in two stages: In the first stage, the data is manually annotated to form a high-quality dataset; in the second stage, a simple but efficient BERT-based model is trained to achieve automatic annotation. Manual correction is then performed to modify the model's error annotation. We also conduct automatic move identification experiments to testify the built corpus. The results underscore the effectiveness of the annotation model and the value of the corpus.

The contributions of our paper are as follows:

- We develop a multi-domain move structure annotation corpus for analyzing moves in RA abstracts. This can benefit non-English speakers

---

<sup>2</sup>The RAAMove corpus is available at <https://github.com/ljk1228/RAAMove>

and students, helping them understand move structures and improve their writing skills. In addition, it can also serve as an essential and valuable resource for language teaching and learning, as well as for other tasks and applications related to move and discourse analysis.

- We suggest a revision of move structure categories based on Hyland's established move classification (Hyland, 2000), which is demonstrated to be better suited for the annotation of moves within scientific Research Article (RA) abstracts and engineering practice.
- We propose an innovative BERT-based automatic annotation model that incorporates word-level saliency attribution. The model is beneficial for facilitating the expansion of the corpus size while maintaining annotation quality.

The remaining of this paper is organized as follows: Section 2 discusses related work, Section 3 describes the scheme, and Section 4 presents the construction of the corpus in detail. Section 5 gives the corpus statistics. Section 6 conducts the experiments and analysis, and Section 7 offers a conclusion.

## 2. Related Work

### 2.1. Scholarly Dataset and Corpora

Corpora of academic texts contain scholarly writing such as research papers, essays and abstracts published in academic journals, conference proceedings, etc. Building scholarly corpora is beneficial for promoting scientific progress and have practical values in many application scenarios, attracting widespread attention from the academic community.

The number of scholarly datasets has increased significantly in recent years. Rohatgi et al. (2023) argue that some of them are task-specific corpora which include selective information of scientific papers (Hsu et al., 2021; Hou et al., 2021; Cachola et al., 2020); some others are open research-aimed datasets. These include corpora built based on *ACL Anthology*, such as ACL OCL corpus (Rohatgi et al., 2023), S2ORC (Lo et al., 2020), ARC (Bird et al., 2008) and AAN (Radev et al., 2009). These datasets usually contain comprehensive metadata and full text of scientific papers, providing researchers with accessible and essential data resources.

### 2.2. Move Structure Annotation

Analysis of arguments and rhetorical move structures in articles has a long history. The Create-A-Research-Space (CARS) model proposed by Swales' pioneering work (Swales, 1990, 2004) and move categories (an abstract can consist of up to

five possible moves)(Swales and Feak, 2009) had widespread influence and gave rise to a plethora of studies of academic discourses in terms of their move structures, but also had some problems. Other improved theoretical approaches like move analysis (Kanoksilapatham, 2003), Argumentative Zoning (AZ) scheme (Teufel et al., 1999) and revised AZ versions (Teufel et al., 2009; Teufel, 2010) propose useful guidelines for annotating moves in scientific articles.

Recently some datasets have been built for studying the move structures. Alliheedi et al. (2019) focuses on the semantic roles and rhetorical moves in the Methods section of biochemistry articles. Liu (2016) constructed a dataset of abstracts selected from the Applied Linguistics Journal; Dayrell et al. (2012) constructed a corpus of abstracts in the fields of Physical Sciences and Engineering, Life and Health Sciences. Cortes (2013) examined the use of lexical bundles and their correlation with move steps with a corpus of 1,300 published RA introductions. Tovar Viera (2020) investigated the rhetorical organization of RA abstracts published in native and non-native English-speaking countries by building an abstract corpus.

Considering that the number of corpora specifically for the moves in abstracts is still small, we believe that the creation of a comprehensive, large-scale corpus is indeed essential. Therefore, the following sections will discuss the scheme and methodology behind our construction of an academic corpus, drawing inspiration from prior theoretical and practical achievements.

### 3. Scheme

This section attempts to answer the following questions:

- Which rhetorical move theories could guide the construction of a corpus for analyzing moves in RA abstracts?
- In what manner can these theories find greater relevance in our work?

In light of this, this section begins by introducing Hyland's move classification (Hyland, 2000), which serves as the theoretical foundation underlying our research, then proceeds with a pilot study, and finally, modifies this classification to meet the needs of constructing a corpus for analyzing moves in current RA abstracts. Unless otherwise stated, Hyland's move classification will be referred to as *H-2000* below.

#### 3.1. Hyland's Move Classification

The rhetorical move structures appearing in RAs have been thoroughly analyzed in prior studies. Hyland (2000), nonetheless, discerned that the move

classification observed in full-length RAs might not be entirely suitable for analyzing RA abstracts. Consequently, he delved into a concentrated yet exhaustive investigation, with a particular focus on categorizing the move structures within RA abstracts. His analysis encompasses 1,400 texts across eight academic disciplines, which can be further categorized into pure sciences, applied sciences, humanities, and social sciences, deliberately covering a wide range of academic practices. This classification stands as a significant contribution for its critical insight into the intricacies and nuances of abstract writing practices and conventions in various disciplines, and thus, has been adopted as the theoretical foundation of this work.

An overview of H-2000 is presented in Table 1.

#### 3.2. Pilot Study

Conducting a pilot study, as emphasized by Biber et al. (2007), is deemed essential in corpus-based genre analysis to ascertain the study's feasibility. In this context, this pilot study was undertaken to assess the applicability of H-2000 for RA abstracts within the disciplines of NLP and Engineering. Furthermore, it sought to uncover any unforeseen issues that might hinder our exploration of these abstracts' move structures and construction of the move corpus.

Our pilot study examined a total of 40 abstracts, with 20 selected from each of the aforementioned disciplines. The data were collected from two highly influential top-tier publications: *the International Journal of Heat and Mass Transfer*<sup>3</sup>, as well as the *Proceedings of the 59th ACL-IJCNLP (Volume 1: Long Papers)*<sup>4</sup>. These abstracts were manually annotated with H-2000 move categories as illustrated in Table 2.

H-2000 closely aligns with the moves identified in the abstracts we annotated, particularly within the Introduction, Purpose, and Method categories. Noticeable distinctions, however, become evident in the Product and Conclusion categories. His definition of Product encompasses not only the statement of findings and results but also the presentation of arguments. However, we did not encounter a single sentence that fulfilled the communicative function of presenting arguments during our pilot study. It is also worth noting that directly applying the Conclusion category to analyze the move structures of abstracts in these fields may appear inappropriate. Consider the following instances:

**[Example 1]:** These results can provide a better understanding of surfactants and guide the practical preparation of multicomponent fluids for boiling

<sup>3</sup><https://www.sciencedirect.com/journal/international-journal-of-heat-and-mass-transfer>

<sup>4</sup><https://aclanthology.org/events/acl-2021/#2021acl-long>

| Move         | Function   |
|--------------|--|
| Introduction | Establishes context of the paper and motivates the research or discussion.   |
| Purpose      | Indicates purpose, thesis or hypothesis, outlines the intention behind the paper.                                    |
| Method       | Provides information on design, procedures, assumptions, approach, data, etc.  |
| Product      | States main findings or results, the argument, or what was accomplished.   |
| Conclusion   | Interprets or extends results beyond scope of paper, draws inferences, points to applications or wider implications. |

Table 1: Hyland’s classification of rhetorical moves in RA abstracts

| Move   | AI    |      | Engineering |     |
|--------|-------|------|-------------|-----|
|        | Freq. | %    | Freq.       | %   |
| Intro. | 17    | 85%  | 16          | 80% |
| Pur.   | 20    | 100% | 19          | 95% |
| Met.   | 19    | 95%  | 19          | 95% |
| Pro.   | 11    | 55%  | 16          | 80% |
| Con.   | 11    | 55%  | 13          | 65% |

Table 2: Frequency of moves identified based on Hyland’s classification in sample abstracts. Where the five moves correspond to those in Table 1.

heat transfer enhancement.

**[Example 2]:** We release source code for our models and experiments at <https://github.com/xxx>. Sentences resembling example 1 may not seamlessly fit within the Conclusion category, as this move typically emphasizes the interpretation and extension of the results rather than pointing out potential effects on related studies and practical applications. Sentences like example 2 often appear in the concluding sections of abstracts in the field of NLP, denoting a willingness to release research code or datasets to assist colleagues and foster future research. These sentences also fail to align with any of the H-2000 categories.

Moreover, sentences falling under the Introduction category in H-2000 appear to serve diverse communicative purposes. Consider the examples 3 and 4, where the former introduces the task, and the latter outlines deficiencies in recent studies:

**[Example 3]:** Undermining the impact of hateful content with informed and non-aggressive responses, called counter-narratives, has emerged as a possible solution for having healthier online communities.

**[Example 4]:** Although such studies have made an effort to build hate speech / counter-narrative (HS/CN) datasets for neural generation, they fall short in reaching either high-quality and/or high-quantity.

The findings from this pilot study underscore the need for a more detailed move structures’ classification. It also reveals significant differences between the abstracts in these two disciplines, highlighting the importance of establishing domain-specific move corpora to better tailor future applications.

### 3.3. Modifying the Move Categories

Considering the primary objective of our research, which is the automated identification of rhetorical moves in RA abstracts, conducting a more precise categorization and annotation of moves can facilitate a more thorough examination of their organizational structure and underlying logic. In this context, our pilot study prompted four modifications to H-2000. The modified classification is outlined in Table 3. To begin, we proposed two distinct categories to replace the Introduction category: Background and Gap. Within our framework, Background serves the purpose of stating the research area while providing historical, theoretical, or empirical context. Gap, however, concentrates on presenting prior studies and emphasizing their limitations, thereby justifying the study’s necessity and persuading the reader of its relevance and contribution to the field. Subsequently, we redefined the move category Product as Result to enhance its clarity in function. Finally, we restated the communicative function of Conclusion and introduced two new move categories: Implication and Contribution. The existing Conclusion category in H-2000, given its current scope, falls short of encompassing more specific functions when used as a label. Therefore, we introduced these two categories to amplify the granularity of move classification. These new categories cover sentences that serve the purposes of “drawing inferences that have not been explicitly stated in the abstracts” and “stating the theoretical and practical value of this paper.” This enrichment enables a more accurate understanding and analysis of abstracts’ moves and organizations.

## 4. Construction of the Corpus

### 4.1. Data Selection and Preprocessing

#### 4.1.1. Data Selection

Our corpus comprises abstracts carefully sampled from leading journals and internationally recognized conferences spanning the academic disciplines of AI and Engineering, in which the AI discipline includes NLP and Computer Vision(CV) domains; Engineering includes Communication Engineering (CE) and Mechanical Engineering (ME). All of which have been selected with the invaluable guidance of experienced professors. The sources

| Move         | Function   |
|--------------|--|
| Background   | States the research area and provides any historical, theoretical, or empirical related information.         |
| Gap          | Establishes a niche: indicates a gap, adds to what is known, presents positive justification (Swales, 2004). |
| Purpose      | Indicates purpose, thesis or hypothesis, outlines the intention behind the paper.                            |
| Method       | Provides information on design, procedures, assumptions, approach, data, etc.                                |
| Result       | States main findings or results or what was accomplished.  |
| Conclusion   | Summarizes the results or extends results beyond scope of paper.   |
| Implication  | Draws inferences which has not been explicitly stated.   |
| Contribution | Points out the theoretical and practical value.  |

Table 3: Enriched move classification

(Table 4), are distinguished by their exceptional Journal Impact Factors and Q1 rankings in the Journal Citations Reports (JCR), unequivocally affirming their preeminent global standing.

Our choice of disciplines was guided by careful consideration. The critical status of Artificial Intelligence and Engineering in applied sciences informed our choice to engage in these fields. As researchers with a background in Natural Language Processing, our strategy was to start with the areas where we are most familiar with. We also identified and selected Computer Vision (CV) in AI, as well as Communication and Mechanical Engineering, as fundamental branches of Engineering to serve as data sources for our corpus construction. We will expand our corpus to cover more domains in the future.

#### 4.1.2. Data Preprocessing

The ACL Anthology<sup>5</sup> is the key resource that archives up-to-date conference papers in the CL and NLP domain. According to ACL Anthology's policy, all the materials are open-access and available for research purposes. The website provides the metadata file (i.e., *bib*) that can be downloaded directly. Considering the main conferences have the largest number of long papers, the length of abstracts in long papers typically ranges from 150 to 250 words, and the formal academic style is also the most standardized, we decided to preprocess the metadata file to extract abstracts from long papers in ACL main conferences held from 2020 to 2022. For the CV domain, abstracts in the CV track proceedings of AAAI 2022 are extracted.

For journals, we first searched the name of each journal with the Web of Science (WOS) platform, and then the basic information of the articles in the journal including title, author, and abstract would be retrieved. After exporting and saving the retrieved results, we can collect all the abstracts data.

We performed sentence preprocessing on the abstract contents by dividing them into one sentence per line with the punctuation marks at the end of the sentences (such as periods) to facilitate the annotation later.

<sup>5</sup><https://aclanthology.org/>

## 4.2. Annotation Process

### 4.2.1. Annotation Guidelines

The annotations were performed with the open-source and crowdsourcing text annotation tool *doccano*<sup>6</sup>. Based on the pilot study mentioned before, we defined 8 basic labels of move structures and their abbreviation for the annotation as shown in Table 5. Besides, the labels are also visually distinguishable through the use of distinct colors in the annotation platform.

Principally, each complete sentence is assigned a single label, and annotators are asked to annotate the most suitable label for each sentence. Nevertheless, in case a sentence appears lengthy and contains two or more subsentences with distinct move structures, it is crucial to accurately label all the moves within that sentence. Here is an example:

[BAC]While neural networks with attention mechanisms have achieved superior performance on many natural language processing tasks, [GAP]it remains unclear to which extent learned attention resembles human visual attention.

### 4.2.2. Manual Annotation

To ensure high-quality annotation, our annotator team comprises four senior teachers and two graduate students specializing in linguistics and English language teaching. It is worth noting that, from the beginning of the annotation and construction of the corpus, the annotators held weekly discussions to collaboratively address challenges and difficulties that arose during the annotation process. This was especially crucial when it came to assigning appropriate labels to those controversial sentences. Through these discussions, the annotators gained a deeper understanding of the guidelines and the expressive characteristics of each abstract. The discussions foster a shared understanding and knowledge exchange within the expert annotators, enhance their comprehension of moves and professional skills in annotation and collaborative work, and allow for the documentation and learning from errors, leading to greater consistency and accuracy. In this way, the conventional Inter-Annotator Agreement (IAA) testing was not employed during the

<sup>6</sup><https://github.com/doccano/doccano>

| Discipline                | Journal/Conference   |
|---------------------------|--|
| Artificial Intelligence   | the Annual Meeting of the Association for Computational Linguistics (ACL)      |
| Artificial Intelligence   | Technical Track on CV on the AAAI Conference on Artificial Intelligence (AAAI) |
| Mechanical Engineering    | Journal of Mechanical Design   |
| Mechanical Engineering    | International Journal of Heat and Mass Transfer                                |
| Communication Engineering | IEEE Journal on Selected Areas in Communications                               |

Table 4: Selected journals and conferences for annotation

| Label        | Abbreviation |
|--------------|--------------|
| Background   | BAC          |
| Gap          | GAP          |
| Purpose      | PUR          |
| Method       | MTD          |
| Result       | RST          |
| Conclusion   | CLN          |
| Implication  | IMP          |
| Contribution | CTN          |

Table 5: Annotation labels and their abbreviation

annotation.

Using the annotation platform (Figure 1), annotators can easily complete the annotation process by selecting an entire sentence and then choosing a pre-defined label. Once annotated, the label and a horizontal line matching the label's color will appear below the sentence. If any modifications are necessary, annotators can simply click on the label beneath the sentence and choose a new one. The ID of abstracts, abstract texts and corresponding annotations can be exported and saved together in JSON format, as shown in the following example.

```
{ "id": 20,
  "data": "Words can have multiple
senses. Compositional distribu-
tional models of meaning have been
argued to deal well with finer
shades of meaning variation known
as polysemy, but are not so well
equipped to handle word senses that
are etymologically unrelated, or
homonymy.",
  "label": [[0, 31, "BAC"], [32, 265,
"GAP"]] }
```

#### 4.2.3. Automatic Annotation

Despite the fact that manual annotations ensure high-quality, gold-standard data, they appear to be labor-intensive and time-consuming. To expedite the annotation process and scale up the corpus, we propose a BERT-based (Devlin et al., 2019) automatic annotation model with saliency attribution trained on the manual annotations.

The model treats annotation as a multi-label recognition and classification problem. It learns semantic features from the annotated data, predicts move categories for unlabeled abstracts, and ultimately,

assigns each sentence with a potential move label.

Under certain circumstances, some words in the sentences may have direct impacts on predicting move labels. For instance, the appearance of the word "result(s)" may result in categorizing the sentence as either Result (RST) or Conclusion (CLN). Our proposed BERT model leverages saliency attribution to enhance move structure recognition (Lin et al., 2023).

To address such issues, our model integrates saliency attribution to improve the accuracy of move structure recognition. Initially, each sentence in the abstract is assigned a move label concerning the sentence's overall meaning. Subsequently, each word within the sentences is regarded as a feature, and its contribution (saliency value) to a particular label is calculated. These saliency values ultimately serve as supporting evidence for enhancing move structure recognition. To sum up, the input of the BERT model contains four embedding vectors: namely Token Embeddings, Segment Embeddings, Position Embeddings, and Saliency Embeddings. An example of move saliency attribution is given in Figure 2. In this figure, a saliency value is assigned to each word to measure its impact on the overall semantic meaning of a move.

Once the automatic annotations were uploaded to the platform, annotators conscientiously reviewed them and corrected any errors present in the annotations. During our annotation practice, we found that there are some common trends in automatic annotation errors. Consider, for example, the frequent misidentification between PUR and MTD moves in model annotations. This occurs under two primary circumstances. One significant source of error is found in instances where abstracts begin directly with PUR moves without including any preceding background information (i.e., BAC and GAP), leading to an incorrect classification of these as MTD.

Another prevalent misclassification is observed when authors of RAs commence their methodology descriptions with the phrase "we propose." In such cases, models tend to erroneously categorize these instances as PUR instead of MTD.

Integrating feedback from annotators, the model was iteratively refined to enhance its performance, enabling the rapid expansion of the corpus while maintaining annotation quality.

Neural abstractive summarization models are able to generate summaries which have high overlap with human references.

The screenshot shows a web interface for document annotation. A dropdown menu is open, displaying a list of move type labels: BAC, GAP, MTD, PUR, RST, CLN, IMP, and CTN. The background text is partially obscured but includes phrases like 'optimized for factual correctness, a critical metric in real-world applications', 'network where we evaluate the factual correctness of a generated summary by fact-checking it automatically', 'tion extraction module.', 'which optimizes a neural summarization model with a factual correctness reward via reinforcement learning.', and 'summarization of radiology reports, where factual correctness is a key requirement.'

On two separate datasets collected from hospitals, we show via both automatic and human evaluation that the proposed approach substantially improves

the factual correctness and overall quality of outputs over a competitive neural summarization system, producing radiology summaries that approach the quality of human-authored ones.

Figure 1: Screenshot of the doccano annotation platform

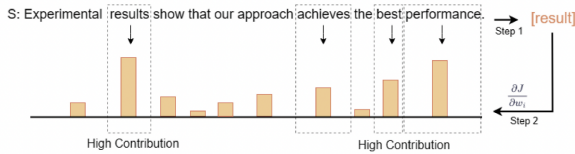


Figure 2: An illustration of move saliency attribution

## 5. Corpus Statistics

### 5.1. Distribution of Move Types

Under our framework of move analysis and implementing the aforementioned annotation procedure, all moves within the corpus have been identified and assigned to one of the eight predefined move types. We have annotated a total of 2,670 abstracts in the field of AI (1,340 in NLP and 1,330 in CV fields) and 2,003 abstracts in the Engineering field (1,003 in ME and 1,000 in CE).

Table 6 presents an overview of the 33,988 annotated move instances within the abstracts.

| Label        | Frequency     | %            |
|--------------|---------------|--------------|
| BAC          | 6,466         | 19.02        |
| GAP          | 3,272         | 9.63         |
| PUR          | 4,874         | 14.34        |
| MTD          | <b>11,526</b> | <b>33.91</b> |
| RST          | 3,732         | 10.98        |
| CLN          | 3,006         | 8.84         |
| IMP          | 282           | 0.83         |
| CTN          | 830           | 2.44         |
| <b>Total</b> | <b>33,988</b> | <b>100</b>   |

Table 6: Frequency and distribution of moves identified in our corpus

The distribution of move structures provides com-

pellent evidence of notable variations in the prevalence of move types within the corpus, characterized by pronounced disparities in both the frequency and proportion of these moves. Of particular significance is the substantial distinction between the move **MTD**, which holds the highest proportion at 30.68% and significantly surpasses all other categories, indicating that most of the abstracts from various fields all focus on describing the methods and approaches, this distribution is also consistent with intuitive judgment, since methods are indeed one of the most attractive parts of abstracts. The least frequently occurring move **IMP**, which accounts for even less than 1% of the total instances.

### 5.2. Occurrence of Move Types

It is pivotal to emphasize that move types are not equally well represented in the abstracts. Hence, we provide statistical examinations regarding the occurrence of these moves within the abstracts. To elaborate, when a specific move is identified in an abstract, regardless of its frequency, we consider it as **occurring** in that abstract. In addition, given the similar structural organizations of abstracts in CE and ME, we have combined them into the category of “engineering” and then proceeded to examine the occurrence of moves within the abstracts in these fields in comparison to the occurrence of moves in abstracts within the realm of AI. The moves’ occurrence in these two disciplines is presented in Table 7.

Table 7 reveals some commonalities and differences in the occurrence of the moves across these two disciplines. For example, in both of the two fields, the move PUR occurs the most, and IMP occurs the least, PUR, MTD and BAC are top three

| Move | AI           |              | Engineering  |              |
|------|--------------|--------------|--------------|--------------|
|      | #            | %            | #            | %            |
| BAC  | 2,003        | 75.02        | 1,528        | 76.29        |
| GAP  | 1,518        | 56.85        | 891          | 44.48        |
| PUR  | <b>2,333</b> | <b>87.38</b> | <b>1,901</b> | <b>94.91</b> |
| MTD  | 2,245        | 84.08        | 1,873        | 93.51        |
| RST  | 1,540        | 57.68        | 953          | 47.58        |
| CLN  | 1,192        | 44.64        | 1,079        | 53.87        |
| IMP  | 112          | 4.19         | 159          | 7.94         |
| CTN  | 544          | 20.37        | 215          | 10.73        |

Table 7: Occurrence and distribution of each move type identified across the two fields in our corpus

moves ranking the occurrence. On the other hand, the proportion of PUR and MTD in the engineering field is far greater than those in the AI field, especially PUR. This discrepancy further emphasizes the potential divergence in move structures within abstracts across various disciplines, underscoring the necessity of investigating move structures in a variety of disciplines. This observation also indicates that moves themselves might differ in that some are obligatory in abstracts, some are optional but commonly present, while others seem to be optional and rare.

|                     | AI      | Engineering |
|---------------------|---------|-------------|
| #Sent.              | 17,391  | 16,597      |
| Average #Sent.      | 6.51    | 8.29        |
| #Words              | 381,734 | 406,244     |
| Average #Words      | 142.97  | 202.82      |
| Average #Move types | 4.38    | 4.29        |

Table 8: The average number of sentences, words, and move types in each abstract within the two fields

Table 8 provides a statistical overview of the sentence counts and occurrence of move types in each abstract from these two distinct disciplines. Abstracts' sentence counts in the field of NLP range from 2 to 15, with an average of 6.5 sentences. Furthermore, the number of move types' occurrence unveils that the majority of these abstracts comprise between 3 to 6 moves, and surprisingly, none of the abstracts contains the complete set of 8 move types. Abstracts in the engineering discipline exhibit longer sentence counts on average, with 8.3 sentences, indicating their lengthy composition. The number of move types in abstracts from these two disciplines, however, appears to be similar. These observations bolster our previous discussion on the potential obligatoriness or optionality of move types within an abstract.

## 6. Experiments and Analysis

In this section, we conduct move structure recognition experiments based on the annotated corpus

to testify the effectiveness of the proposed model. We compare the results with baseline BERT and BERT+Context, where BERT+Context incorporates position information and the preceding and following sentences as additional contextual features. The performance is evaluated by Precision (P), Recall (R), and (Micro) F1.

We also compare the identification performance of our model with that of ChatGPT by manual evaluation.

### 6.1. Move Structure Recognition

The dataset and the results are shown in Table 9 and Table 10 respectively.

| Data         | #Sentences |
|--------------|------------|
| Training set | 7,147      |
| Test set     | 1,787      |

Table 9: Dataset statistics

| Method       | P (%) | R (%) | F1 (%)       |
|--------------|-------|-------|--------------|
| BERT         | 74.06 | 79.58 | 76.72        |
| BERT+Context | 74.55 | 81.23 | 77.60        |
| Our          | 75.01 | 82.34 | <b>78.53</b> |

Table 10: Results of move structure identification

### 6.2. Comparison with ChatGPT

The results indicate that our proposed method outperforms the baseline results, achieving the highest Micro F1 score (78.53%), representing an improvement of 1.81% over BERT and 0.93% over BERT+Context, highlighting the model's ability to not only account for the sequential relationship between contextual information and moves but also emphasize crucial structures in the sentences for accurate prediction and annotation of specific move types.

In this part, we compare the proposed model with ChatGPT. Our aim is to investigate whether Large language model (LLM) can achieve outstanding performance in move structure identification task and assess whether the proposed model could excel beyond the capabilities of LLM.

We randomly selected 30 annotated abstracts from the AI and Engineering disciplines in our corpus, and employed both our proposed model and ChatGPT (gpt-3.5-turbo model) to identify the move structures.

The process of conducting experiments through the gpt-3.5-turbo API is detailed as follows. Initially, a batch processing function capable of receiving multiple sentences as input and providing move structure recognition results for each sentence was defined. Following that, we formulated instructions for ChatGPT, as shown in Figure 3, to ensure the



The move structure of a scientific paper refers to the categorical composition of the linguistic rhetorical components of the academic discourse in the paper. Move recognition is essentially a classification problem in sentences. Now the moves are background, gap, method, purpose, result, conclusion, contribution, and implication. Here are a few examples of move recognition: Detecting emotion in text allows social and computational scientists to study how people behave and react to online events. [background] However, developing these tools for different languages requires data that is not always available. [gap] This paper collects the available motion detection datasets across 19 languages. [purpose] We train a multilingual emotion prediction model for social media data, XLM-EMO. [method] The model shows competitive performance in a zero-shot setting, suggesting it is helpful in the context of low-resource languages. [results] We release our model to the community so that interested researchers can directly use it. [contribution] Below I will give you some sentences, these sentences are from scientific papers, please complete the step recognition.

“Input sentences”

What is the move of this sentence?

Figure 3: Instructions for ChatGPT

professionalism and accuracy of the experiment. Eventually, the move structure recognition results were obtained from ChatGPT.

Subsequent to the acquiring of both models' results, we then asked three experts from each discipline, who were kept unaware of the identifications' generating models, to manually evaluate the results with rating scores range from 0 to 100.

Our model received an average score of 80, whereas ChatGPT obtained an average rating of 65. The results indicate that our proposed model excels at grasping the syntactic structures and context within the text, enabling more effective move structure recognition through the utilization of these structures in tandem with contextual information. Furthermore, our model has the potential for further refinement and optimization to align with specific situations and requirements, which may lead to more promising performance in specific tasks and domains than LLMs.

## 7. Conclusion

In this study, we develop RAAMove, a multi-domain corpus specifically designed for the annotation of move structures within research article abstracts. Based on a modified theoretical foundation derived from Hyland's move classification, this paper delves into the scheme and construction process of the corpus. Notably, the construction process entails a combination of manual annotation and human-guided machine annotation. Through the development of clear annotation guidelines and the engagement of crowdsourced expert annotators, we construct high-quality data. Furthermore, we lever-

age the capabilities of a BERT-based model to automate data annotation, incorporating valuable feedback from human experts. This collective endeavor not only ensures the quality of the corpus but also expands its scale, making it a valuable resource for researchers.

Our preliminary experiments provide empirical evidence of the efficacy of the proposed corpus and its associated model. These findings emphasize the practical utility of RAAMove in various domains, including English language teaching, writing, and research article (RA) analysis. RAAMove is capable of facilitating tasks ranging from discourse and move analysis to move identification and intelligent-assisted correction of research articles. Moreover, we will release the corpus to the public in the near future, ensuring its accessibility to the academic community and beyond.

In the future, we will continue to extend the annotation of move structures to abstracts across diverse domains, future enriching the depth of our linguistic analysis. In addition, we are considering the prospect of move annotation in other sections of research articles, such as the Introduction and Method sections. We are dedicated to furthering the boundaries of move structure analysis and look forward to the coming new developments in this field.

## 8. Acknowledgements

This work was supported by National Social Science Fund of China (NSSF, 23BY166). The authors thank the anonymous reviewers for their constructive comments and suggestions.

## 9. Bibliographical References

- Mohammed Alliheedi, Robert E. Mercer, and Robin Cohen. 2019. Annotation of rhetorical moves in biochemistry articles. In *Proceedings of the 6th Workshop on Argument Mining*, pages 113–123, Florence, Italy. Association for Computational Linguistics (ACL).
- Maryam Alsharif. 2023. Rhetorical move structure in business management research article introductions. *Journal of Language and Linguistic Studies*, 18(4).
- Douglas Biber, Ulla Connor, and Thomas As Upton. 2007. *Discourse on the move : using corpus analysis to describe discourse structure*. John Benjamins Pub. Co.
- Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. 2008. The ACL Anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel Weld. 2020. TLDR: Extreme summarization of scientific documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4766–4777, Online. Association for Computational Linguistics (ACL).
- Kanoksilpatham B. Connor U., Upton T. A. 2007. Introduction to move analysis. In *Discourse on the move: Using corpus analysis to describe discourse structure*.
- Viviana Cortes. 2013. The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for academic purposes*, 12(1):33–43.
- Elena Cotos, Sarah Huffman, and Stephanie Link. 2017. A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes*, 46:90–106.
- Ali M Darabad. 2016. Move analysis of research article abstracts: A cross-disciplinary study. *International Journal of Linguistics*, 8(2):125–140.
- Carmen Dayrell, Arnaldo Candido Jr., Gabriel Lima, Danilo Machado Jr., Ann Copestake, Valéria Feltrim, Stella Tagnin, and Sandra Aluisio. 2012. Rhetorical move detection in English abstracts: Multi-label sentence classifiers and their annotated corpora. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 1604–1609, Istanbul, Turkey. European Language Resources Association (ELRA).
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, volume 1, pages 4171–4186.
- Liangping Ding, Zhixiong Zhang, and Huan Liu. 2019. Factors affecting rhetorical move recognition with svm model. *Data Analysis and Knowledge Discovery*, 3(11):16–23.
- Yufang Hou, Charles Jochim, Martin Gleize, Francesca Bonin, and Debasis Ganguly. 2021. TDMSci: A specialized corpus for scientific literature entity tagging of tasks datasets and metrics. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 707–714, Online. Association for Computational Linguistics (ACL).
- Ting-Yao Hsu, C Lee Giles, and Ting-Hao Huang. 2021. SciCap: Generating captions for scientific figures. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3258–3264, Punta Cana, Dominican Republic. Association for Computational Linguistics (ACL).
- Ken Hyland. 2000. *Disciplinary discourses: Social Interactions in academic writing*. Longman.
- Ken Hyland. 2007. *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.
- Ken Hyland. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, 27(1):4–21.
- Budsaba Kanoksilapatham. 2003. *A corpus-based investigation of scientific research articles: Linking move analysis with multidimensional analysis*. Ph.D. thesis, Georgetown University.
- Jinkun Lin, Hongzheng Li, Chong Feng, Fang Liu, Ge Shi, Lei Lei, Xing Lv, Ruojin Wang, Yangguang Mei, and Lingnan Xu. 2023. Move structure recognition in scientific papers with saliency attribution. In *China Conference on Knowledge Graph and Semantic Computing*, pages 246–258. Springer.
- Xia Liu. 2016. *Constructing a model for the automatic identification of move structure in english research article abstracts*. Ph.D. thesis, Beijing Foreign Studies University.

- Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel Weld. 2020. S2ORC: The semantic scholar open research corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4969–4983, Online. Association for Computational Linguistics (ACL).
- Xiaofei Lu, Jungwan Yoon, and Olesya Kisselev. 2021. Matching phrase-frames to rhetorical moves in social science research article introductions. *English for Specific Purposes*, 61:63–83.
- Ana I Moreno and John M Swales. 2018. Strengthening move analysis methodology towards bridging the function-form gap. *English for specific purposes*, 50:40–63.
- Puong Dzung Pho. 2008. Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10(2):231–250.
- Dragomir R. Radev, Pradeep Muthukrishnan, and Vahed Qazvinian. 2009. The ACL Anthology network corpus. In *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries (NLP4DL)*, pages 54–61, Suntec City, Singapore. Association for Computational Linguistics (ACL).
- Shaurya Rohatgi, Yanxia Qin, Benjamin Aw, Niranjana Unnithan, and Min-Yen Kan. 2023. The ACL OCL corpus: advancing open science in computational linguistics. *CoRR*, abs/2305.14996.
- Reza Ghafar Samar, Hossein Talebzadeh, Gholam Reza Kiany, and Ramin Akbari. 2014. Moves and steps to sell a paper: a cross-cultural genre analysis of applied linguistics conference abstracts. *Text & Talk*, 34(6):759–785.
- Watinee Suntara and Siriluck Usaha. 2013. Research article abstracts in two related disciplines: rhetorical variation between linguistics and applied linguistics. *English Language Teaching*, 6(2):84–99.
- John Swales. 1990. *Genre analysis: English in academic and research settings*. Cambridge University Press.
- John Swales. 2004. *Research genres: Explorations and applications*. Cambridge University Press.
- John Swales and Christine B. Feak. 2009. *Abstracts and the writing of abstracts*. University of Michigan Press.
- John M Swales. 2019. The futures of eap genre studies: A personal viewpoint. *Journal of English for Academic Purposes*, 38:75–82.
- Simone Teufel. 2010. The structure of scientific articles: Applications to citation indexing and summarization. (*No Title*).
- Simone Teufel, Jean Carletta, and Marc Moens. 1999. An annotation scheme for discourse-level argumentation in research articles. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 110–117, Bergen, Norway. Association for Computational Linguistics (ACL).
- Simone Teufel, Advait Siddharthan, and Colin Batchelor. 2009. Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 1493–1502.
- Rodrigo Tovar Viera. 2020. Rhetorical move structure in abstracts of research articles published in ecuadorian and american english-speaking contexts. *Arab World English Journal (AWEJ) Volume*, 10.
- Jungwan Yoon and J Elliott Casal. 2020. Rhetorical structure, sequence, and variation: A step-driven move analysis of applied linguistics conference abstracts. *International Journal of Applied Linguistics*, 30(3):462–478.