Chinese Research Article Introductions: Move Analysis and Linguistic Features

Yan Zhou

Abstract—Research on academic genre has received widespread attention from linguists around the world. In the big data era, the demand of knowledge mining promotes the emergence and development of bibliometrics, and formal description of scientific knowledge is also put on the agenda. The application of RA move analysis to the field of knowledge mining is a new perspective of bibliometrics. At present, there are few studies on move analysis of Chinese RA introductions, and fewer studies are oriented to discourse computing. This study analyzes the move structures of Chinese RA introductions and finds out some linguistic features corresponding to each move. It is found that nearly half of the Chinese RA introductions follow the CARS model, but there are also unconventional structures. The linguistic features, including citation, negation, connectives and sentence subjects, are helpful for the automatic identification of move structures of Chinese RA introductions.

Index Terms—Chinese RA introduction, move analysis, linguistic features.

I. INTRODUCTION

Research on academic genre has received widespread attention from linguists around the world. A research article (RA) can be separated into several discoursal or rhetorical units called moves by Swales [1]-[3]. There has been long-term interest in the move analysis of different parts of English RA, such as abstract [4], introduction [5]-[6], literature review [7], methods [8], results and discussion [9], etc. Most of these studies focus on the functions of each move, or the socio-cultural and cognitive psychological mechanism hidden behind. Over the last decade, with the rapid development of natural language processing (NLP), automatic recognition of English RA moves has made great progress. For example, automatic identification of moves in the abstract part [10] and introduction part [11] have been successfully used in the field of automated writing evaluation and knowledge mining, which are both quite helpful for English learners around the world.

Currently, the number of global Chinese language learners has been gradually increased. Therefore, research on Chinese for Academic Purposes has become more important than ever before. In this research field, move analysis of Chinese RAs starts late and is obviously insufficient, partly because the move structures of Chinese RAs are not matching to those of English RAs. There are three major concerns in this area. The first is to make use of move analysis to improve Chinese language learners’ academic writing skills [12]. The second is to make comparative or contrastive analysis about the discourse structures of English and Chinese RAs, and explore the deep root of the differences [13]-[15]. The third is to investigate the linguistic features of Chinese RAs on the whole for two reasons: one is to find out the register features of Chineses RAs; the other is to help language learners better master Chinese academic writing strategies [16]. However, although there has been some progress in the field of Chinese natural language processing, no automatic identification model of Chinese RA moves has been found. Even few studies have been carried out quantitatively on the linguistic features corresponding to each move of Chinese RAs, which will be helpful for the development of automatic identification model of Chinese RAs. This paper intends to probe into the move structures and linguistic features of the introduction part of Chinese RAs, which is hoped to be a pilot study for a detailed investigation of the full text in the future.

The introduction part of RA is at the beginning of a research article, being responsible for creating a research space [1]-[3]. It serves as a road map of the full text, helping readers quickly understand the current situation and future trend of a certain research field. RA introductions have been regarded as an area to persuade readers and promote publication. The CARS model (Create-A-Research-Space) and its revised versions are the most influential model for RA introduction analysis. Researchers attempted to verify the validity and practicability of the model across languages and disciplines, and proposed different revised models. Among them, Niu [13] made a large-scale multi-level comparative analysis of English and Chinese RA introductions, and found that although most Chinese papers took Swale’s CARS Model, some still bear a prominent “Chinese style” which lacks the typical generic traits of RAs’.

II. GENRE ANALYSIS ON RA INTRODUCTION

Genre analysis is originally used in literary studies. Bakhtin is the first to introduce it to non-literature areas [17]. At present, genre analysis has been widely used to analyze all
text types, such as promotional discourse, academic papers, legal documents, and so on and so forth. Swales defines genre as “a class of communicative events, the member of which share some set of communicative purposes” [2] (p. 58). The expert members in a language community come to reach an agreement on the communicative purpose of a certain discourse, determining the overall structural features of the discourse, and guiding and limiting its content and style. In short, texts of the same genre show similar patterns in terms of research purpose, discourse structure, style and target readers.

Move is a rhetorical unit used to show the function of coherent communication. Its application in genre analysis can be traced back to Swales [1]. Researchers hold different opinions on how to judge moves. Swales believes that the judgment of moves depends on the understanding of the whole text. Context plays a crucial role in move analysis. Nwogu regards move as a discourse fragment composed of different linguistic features, such as vocabulary, proposition, illocutionary behavior, etc [18]. These features give discourse a consistent style. Nwogu believes that moves should be judged on the basis of identifying linguistic features. Only when researchers can not judge moves by the linguistic features, should they turn to context. In fact, in practice, the judgment of moves often depends on both linguistic features and context.

A. Move Analysis on RA Introduction

The CARS model proposed by Swales has three characteristics [2], [3]. First, each move is composed of several steps, including obligatory steps and optional steps. Second, the order of the moves and steps is usually found to be adjusted by different writers. Third, disciplinary differences can be observed from different move structures. Fig. 1 shows us the detailed explanation of CARS model [19] (p. 331).

Move 1—Establishing a research territory
a. by showing that the general research area is important, central, interesting, problematic, or relevant in some way (optional)
b. by introducing and reviewing items of previous research in the area (obligatory)

Move 2—Establishing a niche
by indicating a gap in the previous research or by extending previous knowledge in some way (obligatory)

Move 3—Occupying a niche
a. by announcing purposes or stating the nature of the present research (obligatory)
b. by listing research questions or hypotheses (probable in some fields, but rare in others)
c. by announcing principal findings (probable in some fields, but rare in others)
d. by stating the value of the present research (probable in some fields, but rare in others)
e. by indicating the structure of the RP (probable in some fields, but rare in others)

Fig. 1. Moves in research paper introduction (Swales & Freak, 2012: 331).

There are two critical problems in the study of moves. First, how to determine the minimum unit for analysis. In most cases, the minimum units are sentence, clause or phrase. One move is usually embedded in another move. For example, in the move analysis of RA abstract, the research method is often found in the research purpose or research result. At present, there is no discussion on the minimum unit for analysis in the study of moves of Chinese RAs. Researchers all probed into the move structures of Chinese RAs, but did not mention the minimum unit for analysis [13]-[15]. Second, when judging moves, should we follow the top-down approach, or bottom-up approach? Top-down approach is based on the CARS model without considering cross-language or interdisciplinary differences. Bottom-up approach is based on Nwogu’s idea and investigates move structures by summarizing linguistic features. However, linguistic features are usually too detailed to be sorted out systematically. Most studies on Chinese RAs adopt a top-down approach.

B. Linguistic Features of RA Introduction

1) Sequence

In RA move analysis, there are two characteristics of the linear structure: the move components and move order [20]. In most cases, the three moves in CARS model appear in normal order. Move 1 appearing at the beginning and Move 3 the end. Early studies find that less than 10% of RA introductions put Move 3 at the beginning [21]. Recent studies on Chinese RA introductions find that although there are move cycle and even unconventional structures, most papers start from Move 1 [15].

2) Citation

In the CARS model mentioned above, Move 1 includes two steps: a), showing that the general research area is important, central, interesting, problematic, or relevant in some way (optional); b) introducing and reviewing items of previous research in the area (obligatory). Writers arouse readers’ interest by summarizing previous studies, referring to classical theories and interacting with other researchers in this field. Therefore, although Step 1 is optional, Step 2 must appear in order to persuade readers and promote publication. Citation is a typical linguistic feature of Move 1.

Swales discusses citation from two aspects [2]. One discussion is the difference between integral citation and non-integral citation. “Integral citation” means that the name of the quoted author appears in the middle of the sentence, thus being a part of the sentence. “Non-integral citation” means that the name of the quoted author appears in parentheses at the end of the sentence. The other discussion is whether there are reporting verbs in a citation, such as “show, claim, establish”, etc.

3) Appraisal

Appraisal in academic discourse has been widely recognized [15], [22]. However, there are various forms of appraisal which are difficult to describe systematically. Most studies on appraisal in the RA introductions starts from negation. For example, words like “fail, conflate, violate, few, little, not” often appear in Move 2. Swales lists eight kinds of negation: negative or quasi-negative quantifiers, lexical negation, negation in the verb phrase, questions, expressed needs / desires / interests, logical conclusions, contrastive comment, and problem- raising [2].

Negation has already aroused grammarians’ interest in the last century. Negation can be distinguished into syntactic negation and morphological negation. Syntactic negation can be realized by adding explicit negators, such as “not, no, nobody, nothing, neither, nor, never, none”. Morphological negation can be realized by adding negative prefixes and suffixes, such as “im-, in-, un-, dis-, non-, -less”. Till now, there are few systematic studies on negation in Chinese RA introductions.
4) Connective

Connective is one of the important markers of local coherence in a text [23]. Halliday divides connectives into three types, namely, elaboration, extension and enhancement. For example, expressions like “in other words, that is to say, for example, for instance, at least, in particular, etc.” are used for elaboration, “moreover, in addition, on the contrary, instead, alternatively, etc.” are used for extension, “then, next, afterwards, before that, at the same time, in the end, in consequence, etc.” are used for enhancement [24]. In the study of Chinese discourse, Liang finds 30 common coherence relationships in Chinese narrative texts [25], and Li marks seven coherence relationships in Chinese news texts [26].

Connective is a typical feature in move analysis. For example, “however” and “but” are used to indicate adversative relationship and often appear at the beginning of Move 2. “Therefore”, “so” and “thus” which are used to indicate result often appear at the beginning of Move 3. The correspondence between moves and connectives is one of the concerns of move analysis.

5) Sentence subject

Pho first studies the relationship between sentence subject and move, and puts forward a subject classification system to analyze the move structure of RA abstracts. The subject system he proposes includes seven categories: (a) research objects and their attributes, such as “data, samples and participants”; (b) self-reference, such as “I, me, author, and researcher”; (c) other references, including specific names or citations, previous research in general, general topics and special research objects or outcomes; (d) audience; (e) writer’s own work (macro-research), such as “this study, this research, this paper, this article, and this investigation”; (f) writer’s own work (micro-research), such as “questionnaire, test, survey, interview, conclusion, discussion, argument, comparison and analysis”; (g) anticipatory “it” and existential “there” [27].

Although Pho’s study is aimed to investigate English RA abstracts, it is very enlightening for the analysis of Chinese RA introductions. For example, subjects like “本研究” (this study) and “本文” (this paper) often appear at the beginning of Move 3.

III. METHODOLOGY

A. Research Question

This study aims to answer two questions:
(A) What is the move structure of Chinese RA introduction?
(B) What are the linguistic features of each move?

B. Data Collection


The author sorted the articles according to the citation rate, selected the top 10 articles from each journal, and finally got a total of 110 articles. All the articles chosen were empirical ones with the sections of introduction, methodology, results and discussion, and conclusion. The author extracted the introduction part of each article and constructed a small corpus (43683 characters).

C. Procedure

The author chose clause as the minimum unit for analysis because many scholars hold the view that instead of sentence, clause is more appropriate for the analysis of Chinese language.

The present study adopted a combination of top-down and bottom-up coding methods. For the sake of convenience, the author decided to label the three moves in CARS model as M1, M2 and M3. All the RA introductions were carefully examined and read sentence by sentence and coded as M1, M2 and M3. To ensure the reliability of the annotation of moves, the author found a postgraduate student who majors in linguistics to annotate the corpus together and discussed any inconsistencies detected during the process. A coding reliability coefficient of 0.82 showed a high level of inter-rater reliability and agreement between annotators.

For word segmentation and part of speech tagging, the author chose the Chinese lexical analysis system ICTCLAS (Institute of computing technology, Chinese lexical analysis system) developed by the “Institute of Computing technology, Chinese Academy of Sciences”.

Finally, all the move structures and the linguistic features corresponding to each move were counted and analyzed.

IV. RESULTS AND DISCUSSION

A. Move Structure of Chinese RA introduction

Result shows that 48.2% Chinese RA introductions follow the order of M1-M2-M3, and 18.2% follow the order of M1-M3 without M2. That is to say, nearly half of Chinese RA introductions follow the conventional CARS structure. In addition to these two move structures, there are many other kinds of structures, including M1-M2-M1-M3, M1-M2-M1-M2-M3, M1-M2-M1-M2-M1-M2, M1-M2, M1-M3-M1-M2, M1-M2-M1-M2-M3-M2. Qian & Mu classified the move structures of Chinese RA introductions into two types, namely, conventional structures and unconventional structures [15]. Conventional structures are those that contain all the three moves, and the move order is normal, cyclic, and not greatly changed. Unconventional structures are those in which one move is missing or the order of moves is chaotic. The author used this classification to summarize our results as in Table 1, from which we can see the characteristics of move structures in Chinese RA introductions.

First, from the perspective of the move components, 77.27% of the introductions include all the three moves in the CARS model. However, 22.73% of the introductions lack one move, in which 18.18% lack Move 2. As we know, Move 2 is to “establish a niche by indicating a gap in the previous research
or by extending previous knowledge in some way”. This move is obligatory in CARS model but missing in some Chinese RA introductions. The result shows that Chinese writers tend to avoid directly indicating the gap in previous research.

<table>
<thead>
<tr>
<th>TABLE I: MOVE STRUCTURES OF CHINESE RA INTRODUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td>Conventional Structures</td>
</tr>
<tr>
<td>M1-M2-M3</td>
</tr>
<tr>
<td>M1-M2-M1-M1</td>
</tr>
<tr>
<td>M1-M2-M1-M2-M3</td>
</tr>
<tr>
<td>M1-M2-M2-M3</td>
</tr>
<tr>
<td>M1-M2-M1-M2-M1-M3</td>
</tr>
<tr>
<td>Subtotal</td>
</tr>
<tr>
<td>Unconventional Structures</td>
</tr>
<tr>
<td>M1-M3</td>
</tr>
<tr>
<td>M1-M2</td>
</tr>
<tr>
<td>M1-M2-M1-M2</td>
</tr>
<tr>
<td>M1-M2-M1-M2-M3-M2</td>
</tr>
<tr>
<td>Subtotal</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Second, from the perspective of the move sequence, all the examples in the corpus begin with Move 1. Meanwhile, moves can be circulated. Swales mentions that there is a great relationship between the length of the introduction and the move cycle [3]. The longer the introduction is, the more likely the move cycle appears. Although the moves circulate in different ways, 92.72% of them begin with Move 1 and end with Moves 3.

Third, in terms of the proportion of moves, Move 1 accounts for 66% of the total number of moves, being the largest percentage. Move 2 accounts for about 19% and Move 3 accounts for about 15%. English RA writing emphasizes the importance of Move 2 and Move 3, which shows that researchers need to find the research gap on the basis of reviewing previous studies, and then put forward their own research niche. Western scientists and philosophers believe that the advancement of science has usually been achieved through the interface of different viewpoints, so criticism between scholars is quite common. In the present study, Move 1 accounts for the largest percentage, whereas Move 2 and Move 3 are obviously insufficient. The result is consistent with many studies on Chinese RA introductions which show that Chinese writers pay much more attention to establishing a research territory, but avoid directly indicating a gap in previous research and establishing their own research territory.

B. Linguistic Features of Chinese RA Introduction

Table II summarizes the linguistic features corresponding to each move.

The typical feature of Move 2 is the combination of negators and connectives. In this study, negators appear only in Move 2, and most negators co-exist with connectives. For example, “...然而鲜有涉及多模态与二语语用学相结合的研究 (However, there are few studies involving the combination of multimodality and second language pragmatics.)” [28] The author combines the connective “however” with the negator “rare” to indicate the “research gap”. Other negators like “not yet, can not, not helpful, insufficient, few, little, puzzling, chaotic, etc.” can easily be identified in Move 2.

The typical feature of Move 3 is the sentences beginning with the subject “本文 (this paper)” or “本研究 (this study)”. Move 3 accounts for the least percentage in the total number of moves (15%). Most Move 3 in the present study only includes one sentence beginning with “this paper” or “this study”. Although there are 93 “I” and 11 “we” in the corpus, they did not appear in Move 3.

V. CONCLUSION

This study investigated the move structures and linguistic features of Chinese RA introductions, and found some interesting phenomena.

First, Chinese RA introductions are mostly composed of all the three moves in the CARS model, but the percentage is uneven. Move 1 accounts for 66%, Move 2 accounts for 19%, and Move 3 accounts for 15%. 18.18% of the introductions do not contain Move 2 which is very important for English writers who welcome different opinions and communication between scholars. It is normal for English writers to directly point out problems to each other [29]. Chinese writers often adopt the strategy of “saving face” by skipping a research gap [15]. Similarly, Move 3 accounts for the least, which also shows that Chinese writers do not state their opinions directly like western writers, but tend to pass by quickly to show a modest attitude.

Second, Chinese RA introductions have certain linguistic features, which could be given to computers as features to automatically identify Chinese RA moves. In the big data era, the convenient acquisition of massive scientific knowledge brings researchers not only a wide range of literature, but also the quick mastering of disciplinary knowledge and its current situation. The demand of knowledge mining promotes the emergence and development of bibliometrics, and puts the formal description of scientific knowledge on the agenda. At present, the knowledge mining software widely used in academia is CiteSpace developed by Professor Chen Chaomei of Drexel University in the United States [30], [31].
CiteSpace integrates two main methods to describe knowledge, namely, co-citation analysis and co-word analysis, focusing on the potential knowledge contained in scientific analysis. It is a visual citation analysis software developed under the background of scientometrics, data analysis and information visualization. This software is mainly aimed at finding out the relationship of the information contained in the literatures by analyzing the development trend of a subject field in a certain period. It also presents the relationship of information in the form of atlas, which is conducive to researchers’ macro grasp of a subject or field. However, CiteSpace does not involve in-depth and detailed text analysis, and could not present the details of disciplinary knowledge. Therefore, it could not really tap the most valuable knowledge in a research field. The linguistic features found in the present study could be given to computers as features to automatically identify the moves of Chinese RA introductions. If possible, the research method could be adopted in the analysis of a whole research paper. The incorporation of RA move analysis and subject knowledge mining can become a new perspective of bibliometrics.

**CONFLICT OF INTEREST**

The author declares no conflict of interest

**REFERENCES**


Copyright © 2022 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).