

The Study of Disciplinary Variations in Academic Discourse Anaphora under Accessibility Theory

Xing Junge*, Xing Fangyuan

Jilin University, Changchun, Jilin, 130012, China

*Corresponding author: 15090158401@163.com

Keywords: Discourse Anaphora, Accessibility, Disciplinary Language Variation

Abstract: Drawing on Accessibility Theory, this paper examined the differences in anaphoric expressions between academic discourses in literature and mathematics as represented in their respective journal articles. The comparative analysis of corpus data revealed that anaphoric expressions in both fields tended to adopt forms with higher accessibility. Further investigation into the mechanisms of anaphoric expression selection demonstrated that anaphora constrained its referent form and distance through the precision and immediacy of its reference. This study focused on the exploration of anaphoric mechanisms in different disciplines, aiming to comprehensively reveal the expression mechanism of academic language, promote the further standardization of academic discourse, and provide reference for beginners in academic writing.

1. Introduction

Anaphora is "the author/speaker can use the same or different linguistic expressions to indicate that it is the same as the referent of the expression that appears for the first time"^[1], and its antecedent must appear before the anaphora in the context of the text, and there is a discursive or logical correspondence with the anaphora.

With the development of cognitive linguistics, more and more scholars advocate the analysis of discourse anaphora from the perspective of cognitive psychology. For example, Xu Jiujiu (2003, 58) pointed out that there are three main types of anaphora in modern Chinese: pronoun anaphora, zero-shaped anaphora, and noun anaphora^[2], Chen Ping (1987, 363-378)^[3], Xu Yulong (2003, 321-328), and Xu Jiujiu (2005, 195-208+). 287) Ma Bosen (2008, 23-28) agree with this view and conduct in-depth research on pronoun anaphora and zero anaphora from different perspectives, while Xiong Xueliang and Liu Donghong (2007, 75-79) explore the phenomenon of abstract entity anaphora in argumentative texts^[4].

The research also focuses on the important role of anaphora form, discourse position, and anaphora in discourse organization. Most of the related studies such as Chomsky's "Jurisdiction and Constraint Theory"^[5], Givón's "Topic Continuity Model"^[6], Levinson's "Horn's Hierarchical Relationship"^[7], and Ariel's Accessibility Theory^[8] all take anaphora as the starting point to explore the semantic relationship between anaphora and its true antecedent, while Pan Ningyu and Si Jie (2021, 79-86; 2023, 100-108) analyze anaphora and its selection characteristics in English-Chinese academic discourse based on the accessibility theory^{[9][10]}.

It is worth noting that most of the studies on anaphora are concentrated in English discourse, and there are relatively few studies on anaphora in Chinese, and most of them are narrative and story discourses, and relatively little attention is paid to discourses of special genres such as academic discourse. At the same time, in the selection of corpus, the relevant research pays attention to the correlation between the stylistic characteristics of discourse and the choice of anaphora, and selects academic papers in science and engineering and humanities to form a comparison, which expands the new perspective of academic anaphora research to a certain extent.

Academic language advocates objectivity and descriptiveness, and the particularity of its stylistic characteristics and the richness and diversity of anaphora make it extremely necessary to study academic discourse anaphora. Based on the previous research on Chinese academic discourse, this paper takes literature and mathematics as examples to explore the characteristics of anaphora in different scientific terms, and tries to further analyze the selection mechanism of anaphora, in order to have a deeper grasp of the phenomenon of anaphora in academic discourse.

In summary, this study can be divided into the following key questions: 1. Does disciplinary language affect the form of anaphora to express antecedents, the form of anaphora, the choice of anaphora distance, and the accessibility of anaphora expressions?

2. Research Methods

Based on the self-built anaphora corpus of academic papers, this paper presents the collective characteristics of the authors' anaphora expression of academic papers with a large sample study, and studies the anaphora phenomenon in literary journal papers and mathematical journal papers in combination with the accessibility theory.

The concept of accessibility originated in the field of psychology and generally refers to the ease with which a person can extract a unit of language or memory from the brain's memory system when speaking^[11]. The theory of accessibility is applied in the study of anaphora in discourse, which is reflected in the difficulty of readers to extract a certain concept or information in the context of a specific text. Ariel's referential accessibility follows this concept, according to his accessibility theory, the easier the antecedent information is extracted in the anaphora, the higher the accessibility, and vice versa. At the same time, the accessibility theory also involves the referential distance and the form of anaphora, and this paper compares and analyzes the differences in the anaphora phenomenon in different scientific terminology based on this theoretical framework, which can ensure the comprehensiveness of the argument to a certain extent.

In terms of corpus selection, this paper takes the two basic disciplines with the most significant differences between arts and sciences, literature and mathematics, as representatives. At the same time, in order to ensure the standardization of the corpus, the selected academic papers are all core journal papers of mature scholars in the two disciplines, and the cited data and downloads of CNKI are used as reference standards. In order to avoid the corpus being affected by the personal writing tendency of the creator of academic papers, and to ensure the objectivity and universality of the corpus to the greatest extent, this paper tries to avoid papers with the same author or the same research topic when selecting academic papers. Among them, considering that the abstract part of the academic paper and the content of the main body of the paper overlap to a certain extent, the corpus selected in this paper is limited to the body part of the paper. In addition, since the part of the paper that directly cites the views of other scholars and related literary works is not the content of the author's writing, this part is not included in the corpus.

In this paper, 30 eligible literary journal papers and 30 mathematics journal papers were selected, and the corpus was annotated by MAXQDA software, and after the initial screening and elimination of the corpus, the code annotation function of the software was used to label the corpus from three

levels: antecedent form, anaphora form and the distance between the two in the discourse. The total number of characters in Chinese journals is 254593, 10,059 marked corpus codes and 2,292 corpora are organized, and the total number of characters in mathematics journals is 295881, 5,580 marked corpus codes and 1,378 corpus are organized.

3. Anaphora comparison of terms in the arts and sciences

3.1. Comparison of the forms of anaphora objects

According to Ariel's accessibility theory, the more convenient the information extraction of the antecedent, the higher the accessibility, so the influence of the antecedent on accessibility lies in the complexity of the information contained in it and the difficulty of understanding.

In this paper, the anaphora is divided into nouns/phrases, clauses, sentences, and sentence groups according to the different manifestations of the anaphora in the discourse text. For the definition of clauses and sentences, this paper takes inter-sentence punctuation as the main criterion, and at the same time refers to the completeness of the semantic expression of sentences, and generally uses periods and semicolons as the markers of sentences. As for the distinction between sentences and sentence groups, this paper considers that more than one sentence is a sentence group (including a sentence and a clause acting as antecedents).

In this paper, symbols, formulas, and operation steps are used as antecedents in mathematics journal papers, symbols and their combinations are marked as phrases, theorem formulas are marked as sentences, operation steps of formulas are marked as clauses, and two or more operation steps are marked as sentence groups.

Table 1: The frequency of use of anaphora in different forms of literary journals and mathematics journals

Total	Concrete	Abstract	subject	Noun/Phrase	Clause	Sentence	Sentence Group	Total
2269	1553	716	literature	1521	229	289	201	2240
1356	437	919	mathematics	529	407	361	76	1373
3625	1990	1635	Total	2050	636	650	277	3613

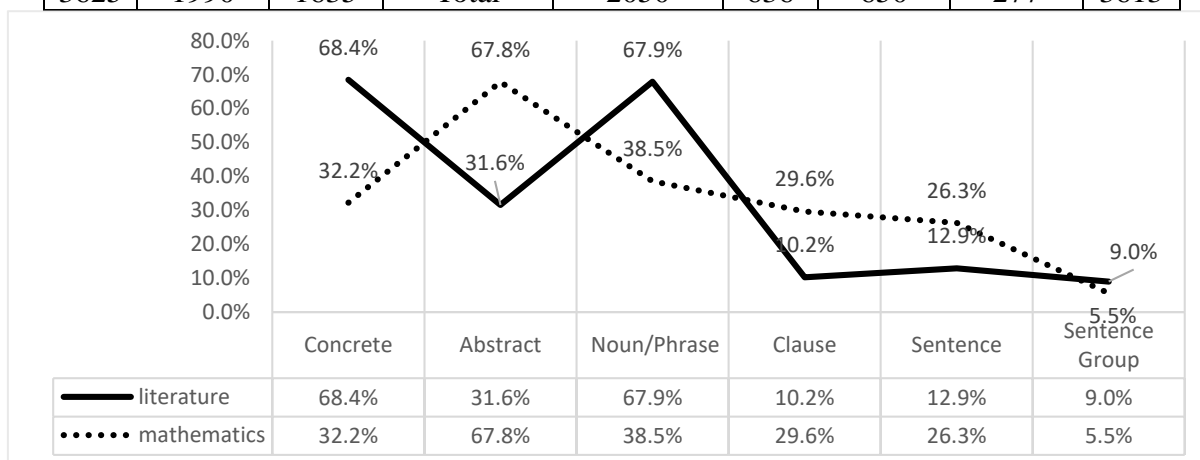


Figure 1: Frequency of use of anaphora in different forms of literary and mathematical journals

From the perspective of antecedent form, the antecedent in the anaphora of literary journals and mathematics journals is mainly in the form of noun/phrase, but there is still a big difference in the

frequency of use of this form in the anaphora of the two types of papers. The frequency of using nouns/phrases in the anaphora of literary journals is as high as 67.9%, and the frequency of using clauses, sentences and sentence groups is about 10%. In the anaphora of mathematics journal discourse, there is little difference in the frequency of noun phrases, clauses and sentences used in antecedents, and the distribution is relatively balanced, but sentence groups are rarely used as antecedents.

In addition, according to the difficulty of understanding and the degree of abstraction of the meaning of the anaphora object. Antecedents can be divided into concrete things and abstract entities, which need to be judged in the context of academic texts. It should be noted that there is a situation where the object of the anaphora is not clear in the academic discourse, and it is necessary to make associative speculation in combination with the context, so the form of the antecedent cannot be judged. The data error in Table 1 is caused by this, but it does not affect our overall comparison of the anaphora phenomenon between literary journal papers and mathematical journal papers, and we will not go into further study here.

The concept of "abstract entity anaphora" was proposed by Asher (1993)^[12], in which the abstract entity acts as the anaphora of the precursor, that is, the abstract entity anaphora. Abstract entities generally refer to concepts, propositions, facts, events, etc., which contain rich and complex information, and correspondingly, the manifestations of such antecedents are mostly clauses, sentences or sentence groups. In contrast to the abstract entity anaphora is the concrete thing anaphora, whose antecedent is in most cases a discursive information that can be directly perceived, and is mostly a noun or noun phrase in the form of expression.

In the actual corpus annotation process, some proper nouns or noun phrases are often difficult to define due to abstract or strong generalization, in this regard, this paper stipulates that: (1) although proper nouns and conceptual nouns are rich in connotation, the meaning used in the sentence in which the anaphora is located is the formal meaning of the word in most cases, so such anaphora should be marked as concrete thing anaphora; (2) if the noun phrase contains a proposition can it be marked as an abstract entity anaphora, otherwise it is a concrete thing anaphora.

In addition, due to the phenomenon that formulas and calculation steps are often used as anaphora objects in mathematical journal papers, this paper stipulates that the simple theorem formula anaphora is annotated as concrete thing anaphora, and if the antecedent involves the reasoning operation of the formula, it is marked as abstract thing anaphora.

Combined with the data in Table 1 and Figure 1, it can be seen that the specific thing anaphora in literary journals accounts for the vast majority, with a frequency of 68.4%, while the relative mathematics journals have a lower frequency of concrete thing anaphora (32.2%), but the abstract entity anaphora is more frequent (67.8%). From the point of view of the abstraction of the meaning of the antecedent, the form of anaphora in literary journals and mathematics journals is quite different.

On the whole, there are significant differences in the formal choice of antecedents in the anaphora of the two scientific terms. Academic discourse in literary journals tends to choose specific things as antecedents, and less in terms of form, and less semantically richer and more complex clauses, sentences and sentence groups. Most of the antecedents referred to in the academic discourse of mathematics journals are abstract entities, and most of the abstract entities are operation steps and theorem explanations, which conform to the language characteristics of mathematics disciplines.

3.2. Comparison of anaphora forms

In this paper, the division of anaphora forms refers to the standards of Ariel (1990) and Xu

Yulong (2018, 26-34), and adjusts according to the discourse characteristics of literary journal papers and mathematical journal papers. According to the morphology and semantic characteristics, anaphora is divided into demonstrative words, demonstrative phrases and definite descriptors. Among them, definite descriptors mainly include "adjectives (except indicative words) + nouns" structure and light rod nouns with anaphora characteristics.

In addition, mathematics journal papers mostly refer to the previous text in the form of "formula/theorem/inference + ordinal number", and this paper marks such echo form as a definite descriptor.

Combined with the accessibility theory, the influence of the demonstrative on accessibility is reflected in the accuracy and timeliness of the anaphora. The form of the demonstrative word is relatively simple, it does not contain the antecedent information, it is highly dependent on the context, and it is necessary to obtain the antecedent information according to its indicative, and the accuracy of the anaphora is relatively low; the definite descriptive words are mostly "indicative words + noun" structure, in which the indicative words contains the antecedent information of the anaphora; the structure of the indicative phrase is the most complex, the accessibility is strong, and the antecedent information is also richer, and even has a supplementary explanatory effect on the antecedent to a certain extent

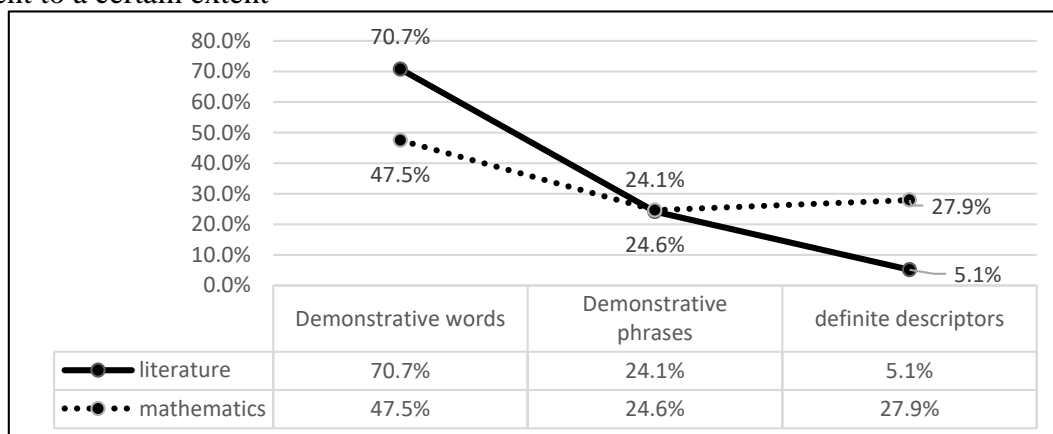


Figure 2: Frequency of use of different forms of anaphora in literary and mathematical journals

As shown in Figure 2, the demonstrative words in literary journals and mathematics journals is mainly used as the anaphora form, but the frequency of use of demonstrators in literary journals is 70.7%, which is much higher than that of mathematical journals. The frequency of choosing demonstrative phrases as anaphora in the two types of academic discourse echoes is very similar, and this form of anaphora is not greatly affected by the linguistic differences of disciplines. It is worth noting that although this kind of structure is simple in form, the accuracy of the anaphora is higher, and the nature and characteristics of the antecedent are indicated by the "formula/theorem/introduction" which is related to the fact that the structure of "formula/theorem/introduction + ordinal number" is often used to refer back to the previous text.

On the whole, the academic discourse of literary journals has a strong restriction on the choice of anaphora form, and tends to choose the simpler form of the indicator as the anaphora language, while the academic discourse of mathematical journals has a relatively weak restriction on the choice of anaphora form, and the various anaphora forms are relatively balanced.

3.3. Anaphora distance comparison

Ariel's accessibility theory refers to the "activation of memory"^[13] at the cognitive-psychological level, arguing that the extraction of anaphora information is the activation of that information in the

memory system. According to this point of view, the closer the anaphora and the antecedent in the academic anaphora, the easier it is to activate the antecedent information, and the higher the accessibility, and the opposite is low.

In this paper, the distance between the antecedent and the anaphora language is labeled with reference to Ariel's accessibility indication level parameters, and four types of text distribution types are marked: the same sentence, the previous sentence, the same paragraph, and the spanning paragraph. Among them, the antecedent of the same sentence is the noun, phrase, and clause in the sentence where the anaphora is located, the antecedent of the previous sentence is the noun, phrase, and clause in the previous sentence, or the whole sentence before the sentence in which the anaphora is located; the same paragraph anaphora is manifested as the distance between the anaphora and the antecedent in more than one sentence, and the anaphora of the two sentences is not classified as such; the spanning paragraph anaphora is manifested as the antecedent is located in the previous paragraph or the first few paragraphs of the paragraph in which the anaphora is located. It should be noted that in the same paragraph, if the antecedent is a sentence group and is located before the sentence in which the echo is located, and is separated by one sentence, it is also marked as the same paragraph instead of the previous sentence.

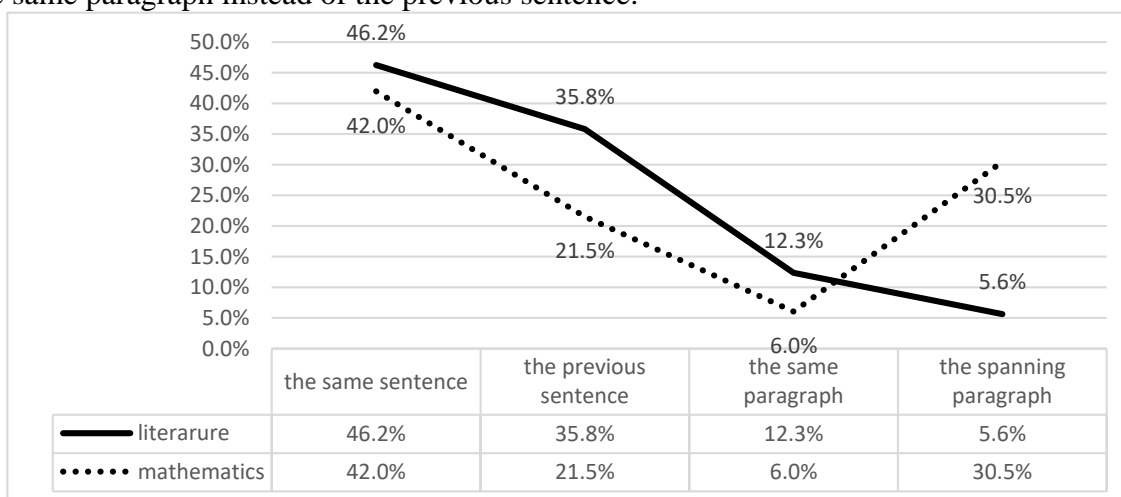


Figure 3: The frequency of use of anaphora distances for literary and mathematical journals

According to the statistics in Figure 3, both literary journals and mathematics journals tend to choose the form of anaphora with a closer anaphora in academic discourse. The frequency of choosing different anaphora distances in academic discourse in literary journals showed an obvious decreasing trend from near distance to distant distance, and the anaphora forms with close distance such as the same sentence and the previous sentence were also mostly selected in mathematics journals, but in particular, the frequency of using the middle span anaphora in academic discourse in mathematics journals was as high as 30.5%. This paper further combs the corpus and finds that the inter-segmental anaphora in mathematical journals mainly appears in the form of "formula/theorem/introduction + ordinal number", and its special structural characteristics make the accessibility of such anaphora high, and the long-distance anaphora of the span is also convenient for the extraction of antecedent information in this anaphora form.

3.4. The summary of anaphora expresses differences in literary journals and mathematics journals

Based on the above data analysis results, the difference of anaphora expression in literary journals and mathematics journals is significant.

From the perspective of the form of anaphora, the two types of academic discourses tend to choose a shorter form in the choice of antecedent form, and the difference is in the degree of abstraction of the meaning of the antecedent, literary journals mostly choose concrete things as the anaphora object, and mathematical journals mostly choose abstract entities as anaphora, which is consistent with the characteristics of the two disciplines, and the academic terminology of mathematics is strong in theory and the meaning is more abstract. This is an important manifestation of the influence of anaphora and its accessibility by the language of the subject.

From the perspective of anaphora form, the anaphora of academic discourse in literary journals and mathematical journals has the same tendency to choose anaphora, and the frequency of choosing the indicator phrase as the anaphora is very similar. It can be seen that the linguistic discourse environment of different disciplines is not strongly restrictive to the form of anaphora expression.

From the perspective of anaphora distance, the two types of academic discourses mostly use the same sentence, the previous sentence and other close anaphora, and the inter-paragraph anaphora in mathematics journals is also used more frequently due to the influence of their anaphora form.

4. The mechanism by which the expression is chosen by anaphora

Based on the above comparison results, this paper finds that the constraints of the language discourse environment of different disciplines on the choice of anaphora form are relatively weak, and the anaphora with more anaphora information and close anaphora is more accessible. Therefore, based on the same corpus, this paper further explores the mechanism of the selection of anaphora expressions in academic discourse.

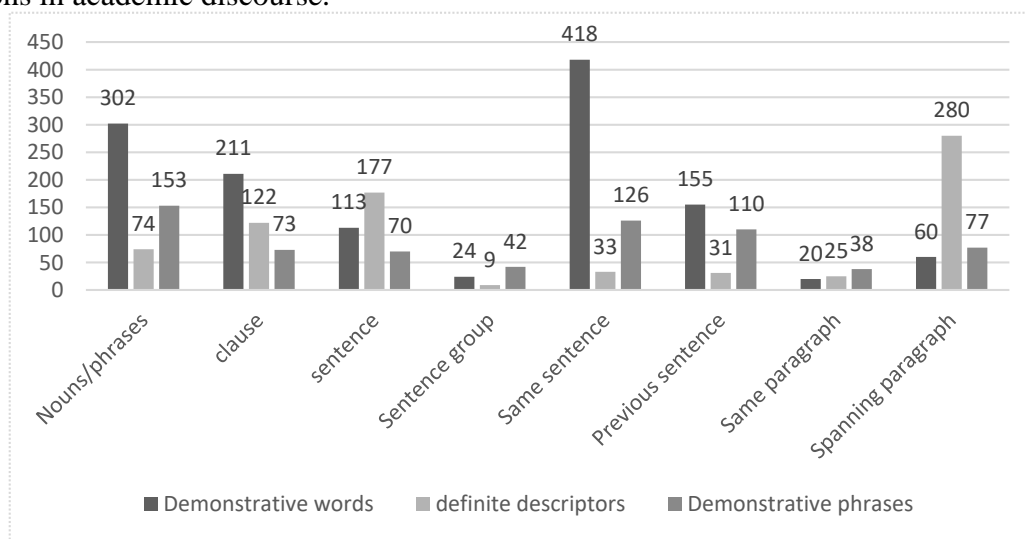


Figure 4: The degree of overlap between the anaphora form and other anaphora components

According to Ariel's accessibility theory, the more convenient the antecedent information extraction, the higher the accessibility, and the acquisition of antecedent information is a representation of the information activation of the memory system. In academic discourse, the relationship between anaphora and antecedent information lies in the indication of the position of the text, and the reference to its whole or part of its content. As shown in the following corpus:

The anaphora of "formula/theorem/inference + ordinal number" in mathematical journal papers can more fully reflect the relationship between anaphora and antecedent information. "formula/theorem/inference" refers to antecedent information, and the ordinal number indicates the antecedent position.

In the above, this paper finds that the influencing factors of anaphora accessibility in academic discourse are reflected in two aspects: first, the accuracy of the antecedent is echoed, that is, the more detailed the description and nature of antecedent by anaphora, the more convenient it is to extract antecedent information and the higher the accessibility of anaphora. The second is the timeliness and accuracy of the anaphora, that is, the passage distance between the anaphora and the anaphora object, generally speaking, the closer the passage distance, the clearer the antecedent information retained in the memory system, and the more convenient it is to activate and obtain this information.

Combined with the data analysis in Figure 4, the structure of the demonstrative word is the simplest, which almost does not contain the information such as the characteristic nature of the antecedent, and only plays the discourse function of indicating or referencing, and the accuracy of the anaphora is not high, so the antecedent in the form of the indicative word is mostly selected in the short form of noun, phrase, clause, etc., and the form of the same sentence or the previous sentence is also selected in the anaphora distance, so as to improve the timeliness of the anaphora, so as to enhance the accessibility of the anaphora as a whole.

Relatively speaking, the definite descriptive language with "adjectives (except indicative words) + nouns" as the main structure contains part of the characteristics and nature descriptions of the antecedent, which can form a better correspondence with the information of the antecedent, and it is more accurate and convenient to obtain the antecedent. The structure of the demonstrative phrase is "demonstrative + adjective + noun", which has the function of indicating the antecedent on the basis of the characteristic nature of the definite descriptive, and the accuracy of the anaphora is the highest, and the constraints on the form of the anaphora object and the distance of the anaphora are weak.

5. Conclusions

The characteristics of academic language are not arbitrary, but are designed to meet the needs of scientific methods, scientific arguments and theoretical interpretations, and thus require logical clarity and clear expression of academic discourse. The study of anaphora in academic discourse is also an attempt to explore the expression mechanism of academic language.

Based on the quantitative statistics of the self-built corpus, this paper compares the differences in anaphora in different scientific terms with literature and mathematics as representatives, and analyzes in detail the choice of anaphora form, anaphora form and article distance in the anaphora expression of two types of journal papers, and further explores the selection mechanism in combination with the anaphora accessibility theory. The results show that the frequency of using short sentences as antecedents in academic anaphora in literary journals and mathematical journals is higher, and the frequency of anaphora is relatively consistent in the choice of anaphora forms, and the frequency of indicating phrases is particularly similar. The difference is that the anaphora in academic discourse in literary journals is mostly concrete, while the abstract entity anaphora in mathematical journals is used more frequently. In addition, the frequency of cross-paragraph anaphora in academic discourse in mathematical journals is relatively high due to the unique anaphora form. This distinction reflects the influence of disciplinary language on anaphora expression. On the whole, the influence of the form of anaphora on the object and distance of anaphora under the accessibility theory is significant.

In the above, this paper analyzes the anaphora phenomenon in literature and mathematics in detail, and further explores the linguistic expression mechanism of academic anaphora based on the accessibility theory. However, it should be noted that the factors influencing the expression of anaphora are diverse and complex, and there is still room for improvement in the discussion of

cognitive linguistics from the perspective of language form.

Acknowledgements

This paper is one of the research results of the 2023 National Innovative Training Program for University Students "Research on the Indication and Expression of Master's Degree Dissertations" (Project No.: 202310183009).

References

- [1] Liao Qiuzhong. *Collected Works of Liao Qiuzhong* [M]. Beijing: Beijing Language Institute Press, 1992: 45
- [2] Xu Jiujiu. *Modern Chinese Chapter Anaphora* [M]. Beijing :China Social Sciences Press, 2003:58
- [3] Chen Ping. *Discourse analysis of Chinese zero-shaped anaphora* [J]. *Chinese Chinese*, 1987, (05), 363-378.
- [4] Xiong Xueliang, Liu Donghong. *Research on the anaphora of abstract entities in the argumentative text* [J]. *Journal of Sichuan University of Foreign Languages*, 2007, (01):75-79.
- [5] Chomsky N. *Lectures on Government and Binding* [M]. Dordrecht: Fories, 1981: 74-85.
- [6] Givón T. *Introduction* [M]. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1983: 5-41.
- [7] Levinson S. *Pragmatics and the Grammar of Anaphora: a Partial Pragmatic Reduction of Binding and Control Phenomena* [J]. *Journal of Linguistics*, 1987 (2):379-434.
- [8] ARIEL M. *Referring and Accessibility* [J]. *Journal of Linguistics*, 1988 (1): 65-87.
- [9] Pan Ningyu, Si Jie. *A study on the anaphora of English-Chinese academic discourse based on the perspective of accessibility* [J]. *Journal of Jiaxing University*, 2021, 33(05):79-86.
- [10] Pan Ningyu, Si Jie. *Commonality analysis of anaphora in English-Chinese academic discourse from the perspective of cognition* [J]. *Journal of Jiaxing University*, 2023, 35(05):100-108.
- [11] Xu Yulong. *English-Chinese referential word expression accessibility* [J]. *Foreign Language Teaching and Research*, 2000 (5): 321-328.
- [12] Asher, N. *Reference to Abstract Objects in Discourse* [M]. Dordrecht: Kluwer Academic Publishers, 1933.
- [13] ARIEL M. *Accessing Noun-Phrase Antecedents* [M]. London and New York: Routledge Press, 1990.