

A Multi-Dimensional Contrastive Analysis on English Abstracts in Chinese Master's Theses and International Core Journal Articles

Yifan Shi

School of English Language, Literature and Culture, Beijing International Studies University, Beijing, China

Zhencong Liu

School of English Language, Literature and Culture, Beijing International Studies University, Beijing, China

Abstract—As a signpost for the reader, the abstract contains the main body of a thesis and summarizes the most important information in an article. Thus, it is essential for assessing the quality and value of a thesis. This research aims to provide insight into the linguistic features of English abstract writing. Based on the multi-dimensional analysis approach proposed by Biber (1988), this study investigates the difference in English abstracts between the Chinese master's theses (CMT) and the international core journal article (ICJA). With the help of the MAT (Multi-dimensional Analysis Tagger 1.3) analysis tool, the corpus of 100 English abstracts from these 2 resources has been analyzed for comparison of the functional and lexical-grammatical features in 6 dimensions. The results show that there are differences between CMT and ICJA's English abstracts in dimension 4 (Overt Expression of Persuasion) and dimension 6 (On-line Informational Elaboration), and 25 linguistic features across these 6 dimensions. The results help learners identify the differences in dimensions and linguistic features between their English abstracts and those written by experts to improve their abstract writing and to write more native-like theses.

Index Terms—Multi-dimensional Analysis, abstract writing, master's theses, EFL graduate student, academic writing

I. INTRODUCTION

To get a master's degree, a master's thesis is the essential and last step in most universities around the world. For some L2 learners, it is extremely important to produce a high-quality English master's thesis. In a master's thesis, the abstract is the most important piece of work as it is one of the first things an examiner will look at. It is also the section of text that is reproduced in computerized databases of theses (Paltridge & Starfield, 2020). Therefore, to impress the examiners and show excellent language ability, the language used in the abstract needs to be paid attention to by English as a foreign language (EFL) graduate students. As important learning resources for EFL graduate students, articles from international core journals provide them with the most professional guidance, which are not only academic references for students but also models of academic writing. Through the learning of these theses, academic languages can be acquired by them, and linguistic features can be imitated to improve their academic writing ability. Besides, in some L2 academic English writing classes, the structure, purpose, and format of each chapter in a master's thesis are emphasized rather than vocabulary, phrases, sentences, and grammar. Therefore, some EFL graduate students find themselves struggling when they are writing academic English, and a large number of inappropriate uses of languages can be found in their master's thesis. Moreover, for some EFL graduate students, the English abstract is often an adjunct to the abstract in the native language of their thesis (Zhang et al., 2018). As a result, grammatical errors and incorrect usages widely exist in their English abstracts.

To improve the academic writing ability of EFL graduate students and help them with the writing of abstracts in their theses, a multi-dimensional analysis approach was used in this study to explore the difference in English abstracts between CMT and ICJA. Specifically speaking, compared with experts in this field, EFL learners' deficiencies in abstract writing are expected to be found, so as to facilitate academic English writing teachers and graduate students to have an in-depth understanding of the linguistic features of English abstracts and provide a reference for the teaching of academic English writing. As for the significance of this study, it aims to promote the learning and teaching of abstract writing. Based on the results of this study, EFL learners are able to identify the differences in dimensions and linguistic features between their English abstracts and those written by experts. Subsequently, training and imitation can be used to make up for their deficiencies in a certain dimension by learning and imitating the linguistic features in abstracts of ICJA to write a more native-like thesis.

II. MULTI-DIMENSIONAL ANALYSIS

Based on Halliday's (1988) register analysis method which is limited to a single text, the theoretical framework of Multi-feature / Multi-dimensional Analysis (MDA) was developed by Biber (1988) to identify underlying dimensions of linguistic variation from a wide range of spoken and written registers of English. Using factor analysis, he established MDA supported by a large number of texts, which effectively overcomes the shortcomings of traditional register research.

The theory uses quantitative and empirical methods to identify salience language co-occurrence features, and qualitatively interprets the functions of language features, forming the basic dimension of observing register variants. Based on the LLC (London Lund corpus) spoken English corpus and the LOB (Lancaster Oslo Bergen) written English corpus, Biber interpreted the dimensions among texts specified by the factor scores. Accordingly, 67 linguistic features were analyzed and 6 main functional dimensions have been extracted. Each linguistic feature has a "loading". Features with higher absolute value "loadings" on a factor are better representatives of the dimension underlying the factor. The linguistic features with absolute values of "loadings" greater than 0.45 in each dimension are typical linguistic features. When a text has several occurrences of the features with negative weights it will likely have few of the features with positive weights, and vice versa.

According to Biber (1988), the 6 main dimensions involve dimension 1: Involved versus Informational Production. Low scores on dimension 1 present high informational and integrative of a text; Dimension 2: Narrative versus Non-narrative Concerns which distinguishes narrative discourse, and high scores on this dimension indicate that the text is narrative; Dimension 3: Explicit versus Situation-dependent Reference which distinguishes explicit, context-independent reference and nonspecific, situation-dependent reference; Dimension 4: Overt Expression of Persuasion which overt marking of the speaker's point of view and high scores on this dimension indicates that the text is explicit; Dimension 5: Abstract versus Non-abstract Information which marks informational discourse that is abstract, technical, and formal versus other types of discourse; Dimension 6: On-line Informational Elaboration which distinguishes discourse that is informational but produced under real-time conditions from other types of discourse.

III. LITERATURE REVIEW

A. *Research on Multi-Dimensional Analysis*

Over the past three decades, MDA has been widely used in the field of linguistics because of "its unique macro research perspective, and it has expanded from oral and written registers to other registers" (Pan, 2022, p. 28). This framework has been mainly applied in the study of register variation. For example, Biber et al. (2002) made a Multidimensional Comparison between Speaking and Writing at the University. Registers of college writers with and without disabilities have also been analyzed using MDA approach (Gregg et al., 2002). In recent years, MDA research has been switched to non-traditional registers such as networks, television, and film. Sardinha and Pinto (2019) identified the dimensions of variation across American television programs. Brugman et al. (2021) collected transcripts of satirical news shows, regular news shows, and fiction shows to identify register dimensions, and patterns in linguistic features unique to genres, which we used to determine the presence of discursive integration. Pinto and Sardinha (2014) analyzed the discourse evolution of American films from 1930 to 2010, identified 7 dimensions, and used ANOVA to analyze the variation. Thus, it can be seen that the analysis of register variation based on MDA is paying more attention to new media and new registers. In addition to the study of variation among various registers, MDA has also been applied to studies of academic English discourse, for instance, Crosthwaite (2016) explored EAP instruction's impact on longitudinal linguistic variation in the direction of the established norms of an academic register, and recently, some scholars used MDA to make a comparative analysis of L1 and L2. Pan (2018) identified and interpreted 4 dimensions that capture the lexical and grammatical differences between L1 and L2 academic writing in applied linguistics. Friginal and Weigle (2014) identified the functional dimensions of L2 academic essays and analyzed linguistic variation in the corpus across parameters of time and average assessment scores. A small number of Chinese scholars compare the English abstracts of Chinese and international core journals from dimensions and linguistic features (Xie & Ma, 2021; Zhang et al., 2018), and the textual and linguistic changes in English abstracts have also been investigated from a diachronic perspective (Xie, 2020).

B. *L2 English Abstracts Writing*

The abstract is one of the most essential parts of a master's thesis, which has attracted the attention of many scholars, especially for L2 learners' English abstracts. As noted above, some contrastive studies have utilized MDA as an approach to compare the linguistic features of English abstracts written by native speakers and non-native speakers. For example, using the MDA approach, Zhang et al. (2018) compared the English abstracts of the most-cited articles between Chinese linguistics journals and international linguistics journals. They found that the abstracts of highly cited articles in Chinese and international journals do not differ significantly in the 6 dimensions, but some specific language features in the same dimension are used differently, and there are differences in the use of specific linguistic features in the same dimension. According to this study, it is worth noting that the abstracts of articles written by L2 and English-speaking researchers do not differ significantly in functional dimensions, but there are some differences in

linguistic features, which suggests that although the scholars' abstracts are similar in register, L2 authors need to learn further from the articles of native speakers in terms of linguistic features. In addition to linguistic features, scholars also compared abstracts genre (Feng & Wu, 2016), hedges (Qiu, 2015), pragmatic identity construction (Sun, 2015), and so on. Most of these studies are corpus-based, and both qualitative and quantitative research method was adopted to analyze the English abstracts of native speakers and non-native speakers. The purposes of these papers are similar, which are to find the differences between the two corpora and provide guidance for non-native speakers' Academic English writing. Moreover, the research object of these studies is mainly about the comparison between journals to improve the quality of abstracts in published papers. Although such research has been expanded to some extent, there is no multi-dimensional contrastive study on the abstracts written by EFL graduate students and international journals to improve L2 learners' English academic writing ability. Therefore, this study is expected to improve the research in this field from the perspective of academic writing teaching.

IV. METHODOLOGY

A. Research Questions

This research aims to analyze the functional and lexical-grammatical features difference in English abstracts between CMT and ICJA, to help the EFL graduate students with the writing of abstracts in their master's theses and provide a reference for the teaching of academic English writing. Research questions include: (1) What are the differences in functional dimensions of Chinese master's theses and international core journal articles? (2) What are the differences in lexical-grammatical features of Chinese master's theses and international core journal articles?

B. Corpus

In this study, two corpora were established by the researcher. The first corpus includes 50 abstracts of CMT in recent 5 years, which were downloaded from the China Master's Theses Full Text Database (CFMD). The selected corpus was drawn from the highly cited theses in CFMD in the last five years. The MAs who wrote these theses were all majoring in foreign linguistics and applied linguistics, and the topics of these papers include translation studies, teaching studies, linguistics studies, etc. The other corpus consists of 50 abstracts from the ICJA in the field of linguistics in 2022, including, Applied Linguistics, Journal of Pragmatics, and TESOL Quarterly. The total word count of the CMT corpus is 20221 and there are 8762 words in the corpus of ICJA. The CMT corpus has more words than the ICJA corpus, but MAT can standardize the data to ensure the comparability of the two corpora.

C. Methods

Firstly, the multi-dimensional annotation and analysis tool developed by Nini (2015) -- Multidimensional Analysis Tagger V1.3 was used to conduct automatic text annotation, feature extraction, and data analysis. With the help of this tool, 67 detailed language features and 6 functional dimensions of CMT and ICJA were counted respectively. Then, an independent sample t-test was carried out on the statistical results of the 2 corpora by using SPSS 26. Lastly, an analysis of the results was conducted with case studies to explore the differences between theses function dimensions and lexical-grammatical features.

V. RESULTS AND DISCUSSION

As shown in Table 1, an independent sample t-test on the dimension scores of English abstracts between CMT and ICJA shows that there are significant differences between English abstracts of CMT and ICJA in dimension 4 (Overt Expression of Persuasion) and dimension 6 (On-line Informational Elaboration) ($p < 0.05$), while there is no significant difference in dimension 1, 2, 3, and 5. In dimension 4, the dimension score of CMT is higher than ICJA, while in dimension 6, the dimension score of CMT is lower than ICJA. Besides, the dimension score of both CMT and ICJA are high in dimensions 3, 5, and 6 and low in dimensions 1, 2, and 4.

TABLE 1
T-TEST ON THE DIMENSION SCORE OF ENGLISH ABSTRACTS BETWEEN CMT AND ICJA

T-TEST ON THE DIMENSION SCORE OF ENGLISH ABSTRACTS BETWEEN CMT AND ICJA							
	CMT		ICJA				
	n=50		n=50		<i>t</i>	<i>df</i>	<i>p</i>
	mean	Std deviation	mean	Std deviation			
Dimension1	-17.667	5.011	-16.950	5.362	-0.692	98.000	0.491
Dimension2	-4.049	2.582	-3.068	3.218	-1.682	98.000	0.096
Dimension3	10.861	4.336	10.063	5.259	0.828	98.000	0.410
Dimension4	-3.415	2.604	-4.538	2.881	2.045	98.000	0.044
Dimension5	3.700	3.769	3.323	5.278	0.412	88.661	0.682
Dimension6	0.065	2.121	1.242	2.866	2.334	90.280	0.022

Based on the above data analysis, compared with ICJA, CMT scored higher on dimension 4, indicating that CMT's English abstracts are more persuasive and explicitly mark the author's point of view, while CMT scored lower on dimension 6, indicating that CMT's English abstracts are more elaborated. In addition, the dimension scores of English abstracts of CMT and ICJA are high in dimensions 3, 5, 6 and low in dimensions 1, 2, and 4, indicating that their English abstracts are more informative, explicit, abstract, and elaborated but less narrative and persuasive. According to the results of MAT, the closest text type of CMT and ICJT's abstracts are "learned exposition". Texts belonging to "learned exposition" are typically informational expositions that are formal and focused on conveying information and scores of them are low on dimension 1 and high on dimension 3 and 5 (Biber, 1989).

A. Contrastive Analysis on Dimension 1

Dimension 1 is the opposition between "Involved and Informational" discourse. There is no significant difference between the dimension scores of these two corpora ($p = 0.491 > 0.05$). The scores of these two corpora on this dimension are low, which indicates that abstracts are the type of texts that can be associated with a high informational focus and a careful integration of information.

TABLE 2
THE DIFFERENCES IN LINGUISTIC FEATURES IN DIMENSION 1 (INVOLVED VERSUS INFORMATIONAL PRODUCTION)

Features	CMT's mean	ICJA's mean	t	p
Average word length	2.649	3.518	-4.816	0.000
Emphatics	-0.059	-0.634	2.150	0.034
First-person pronouns	-1.002	-0.772	-4.690	0.000
Total other nouns	2.985	2.220	3.483	0.001
Pronoun it	-0.430	-0.946	3.347	0.001
Total adverbs	-2.419	-2.127	-2.033	0.045
Be as main verb	-1.683	-2.134	3.003	0.003
Private verbs	-0.661	-0.367	-2.100	0.038
Sentence relatives	4.845	1.760	2.708	0.008
WH-clauses	-0.290	0.818	-2.216	0.031
Attributive adjectives	2.042	3.104	-3.475	0.001

There are significant differences between CMT and ICJA on 11 linguistic features of dimension 1 ($P < 0.05$). Among them, the scores of CMT are lower than that of ICJA in 6 linguistic features, namely Average word length, First-person pronouns, Total adverbs, Private verbs, WH-clauses, and Attributive adjectives, whereas the scores of CMT are higher than that of ICJA in 5 linguistic features including Emphatics, Total other nouns, Pronoun it, Be as main verb, Sentence relatives.

First-person pronouns appear more frequently in texts with high interactivity. In the corpus of CMT, first-person pronouns appear less frequently than ICJA. Pronouns "we" appeared 4 times in the CMT corpus, while it occurred 35 times in the ICJA corpus. This indicates that EFL students prefer to avoid using first-person pronouns and tend to use words like "the author" or "this research" instead of "we" or "I". On the one hand, this may be because EFL students consider that avoidance of first-person pronouns may enhance the objectivity of the thesis. The emphasis on the objectivity of the thesis is the consensus of EFL academic writing instruction and the basic rule of the dissertation writing. On the other hand, this indicates that EFL students ignore the interaction with readers and place more emphasis on the objectivity of information. In contrast, international scholars use more first-person pronouns. For example, "we argue that..." "we suggest that..." "we examine..." "we explore..." etc. This shows that the use of first-person pronouns can be accepted by international journals and scholars. Although the use of first-person pronouns may diminish the informative nature of academic articles, the appropriate use of first-person pronouns does not cause the thesis to become unobjective.

Word length and type/token ratio similarly mark "a high density of information, and longer words also convey more specific, specialized meanings than shorter words" (Biber, 1988, p. 104). According to the results in Table 2, the average word length of ICJA is significantly longer than that of CMT, which indicates that the words of ICJA abstracts are more specific and specialized. The reason for this is that EFL students may have a limited vocabulary compared to international scholars. Therefore, more specialized word study is crucial for them, and teachers are expected to encourage students to use more long words in their thesis abstracts.

Be as main verb can be associated with a fragmented presentation of information, resulting in a low informational density. It is typically used to modify a noun with a predicative expression, instead of integrating the information into the noun phrase itself, for example, "the house is big versus the big house" (Biber, 1988, p. 106). According to Table 2, there are fewer uses of Be as main verb in ICJA, which highlight the information density and the informativity of the text. However, the use of Be as main verb often appears in EFL students' theses, for instance:

- (1) "Metaphor is a common language phenomenon in our lives." (CMT)
- (2) "...news report is full of hostility with ecological property of eco-destructive." (CMT)
- (3) "the eco-critical discourse analysis is a newly developed branch of the eco-linguistics, which..." (CMT)

Besides, Pronoun it is often found co-occurring with be verbs (e.g., it is meaningful to study...; it is necessary to enhance...; It is easier to express...). The use of these 2 features decreases the density of information, which reduces the

specialization and informativity of the thesis. Therefore, EFL students should try to avoid the use of *be* as the main verb, thus enhancing the informativity of the article.

Emphatics mark heightened feelings, and sentence relatives are used for attitudinal comments by the speaker. This feature is used for “involved discourse, marking high interpersonal interaction or high expression of personal feelings, such as, *sure, very, really, just, most, more, certainly, assuredly* and so on” (Biber, 1988, p. 106). English writers used emphatics for a variety of purposes: “to stress the significance and contributions of their findings, boost the current knowledge and scholarship, emphasize the results, etc” (Abdollahzadeh, 2011, p. 293). Emphatics appear less frequently in abstracts written by EFL students and scholars, but more emphatics are used by students in comparison to scholars, for example:

(4) The study found that Chinese and American news reports do have differences in language use. (CMT)

(5) The modal verbs are more likely to express the attitudes of the reporter. (CMT)

(6) ...proclaim resources of contraction resources to clearly show China’s diplomatic stance. (CMT)

The use of emphatics in CMT’s abstract enhances the authors’ emotional expression and interpersonal interaction, and the extensive use of emphatics reduces the objectivity of the abstract, which is not in line with the avoidance of subjectivity in academic texts. EFL students are expected to reduce the use of adverbs, modal verbs, and verbal when reporting results in their abstracts to enhance the objectivity of the text.

Private verbs (e.g., *reveal, indicate, imply, show, suppose*) are used for the overt expression of private attitudes, thoughts, and emotions. International scholars use more private verbs to make their points and convince their readers. Those verbs can engage the reader and help them picture what’s happening in the paper.

B. Contrastive Analysis on Dimension 2

Dimension 2 is the opposition between “Narrative and Non-Narrative Concerns”. There is no significant difference between the dimension scores of these two corpora ($p = 0.096 > 0.05$). The scores of these two corpora on this dimension are low, which indicates that the texts are non-narrative.

TABLE 3
THE DIFFERENCES IN LINGUISTIC FEATURES IN DIMENSION 2 (NARRATIVE VERSUS NON-NARRATIVE CONCERNS)

Features	CMT’s mean	ICJA’s mean	<i>t</i>	<i>p</i>
Past tense	-1.108	-0.846	-2.975	0.004
Perfect aspect	-0.864	-1.237	2.539	0.013
Public verbs	-0.709	-0.231	-2.044	0.044

There are significant differences between CMT and ICJA on 3 linguistic features of dimension 2 ($P < 0.05$), including, Past tense, Perfect aspect, and Public verbs. The negative scores for all 3 linguistic features indicate that these 3 features occur less in both corpora. Students use less Past tense, Public verbs, and more Perfect aspects compared to the experts’ abstracts.

Past tense and Perfect aspect forms mark actions in past time. They have been associated with “narrative/descriptive texts and with certain types of academic writing” (Biber, 1988, p. 223). In dissertation abstract writing, the tense of the verb often depends on the difference in moves. Hyland (2000) proposed a five-move structure to analyze academic abstracts, including Introduction, Purpose, Method, Product, and Conclusion. The study showed that in the abstracts of applied linguistics journal articles, the present tense usually occurred in the first, second, and fifth moves, while the past tense was often used in the third and fourth moves (Tseng, 2011). Therefore, according to the results of MAT, although abstracts are non-narrative texts with less use of the past tense, the past tense is essential when writing moves of Method and Product. In this study, fewer past tenses appeared in the abstracts written by EFL students compared to the ICJA abstracts, which may be because few students used the past tense at the fourth move (Product), while more international scholars used the past tense at this move, for example:

(7) The results of the research show that it helps students play their positive roles in learning... / The results from data analysis indicate that... / the author finds that Howard Goldblatt has a great awareness of... (CMT)

(8) The results demonstrated a linguistic positivity bias in academic writing. / Results showed that over two-thirds of posts are attached with hashtags... / Reciprocal self-disclosures were found to be central to constructing positive relations... (ICJA)

In contrast, the Perfect aspect appears more in the CMT summaries, probably because EFL students use more Perfect aspects in the first move for the introduction. Whereas, scholars are accustomed to reducing the length of the first move due to the word limit of international journal abstracts. In general, verb tenses in abstracts should be used flexibly under the condition that they follow the rules of academic writing, as multiple tenses may appear in the abstract. In addition to the verb tense in this dimension, there is a significant difference in the use of public verbs that are apparently used frequently with these other forms because they function as markers of indirect, reported speech (e.g., *admit, assert, declare, suggest, report, say*) (Biber, 1988, p. 109). Scholars tend to use more private verbs to demonstrate the views of previous researchers.

C. Contrastive Analysis on Dimension 3

Dimension 3 is the opposition between “Context-Independent Discourse and Context-Dependent Discourse”. There is no significant difference between the dimension scores of these two corpora ($p=0.410>0.05$). The scores of these two corpora on this dimension are high, which indicates that the texts are explicit and not dependent on the context.

TABLE 4
THE DIFFERENCES IN LINGUISTIC FEATURES IN DIMENSION 3 (EXPLICIT VERSUS SITUATION-DEPENDENT REFERENCE)

Features	CMT's mean	ICJA's mean	<i>t</i>	<i>p</i>
Nominalizations	2.034	2.652	-2.036	0.045
Phrasal coordination	6.012	4.489	2.100	0.038
Pied-piping constructions	-0.318	0.559	-2.270	0.027

There are significant differences between CMT and ICJA on 3 linguistic features of dimension 3 ($P < 0.05$), including, Nominalizations, Phrasal coordination, and Pied-piping constructions. Students use fewer Nominalizations, Pied-piping constructions, and more Phrasal coordination compared to the scholars' abstracts.

“The co-occurrence of phrasal coordination and nominalizations with the relativization features indicates that referentially explicit discourse tends to be integrated and informational” (Biber, 1988, p. 110). Nominalizations can expand idea units and integrate information into fewer words. As shown in Table 4, nominalization appears significantly more often in ICJA abstracts than in CMT, for example, “attainment” “enhancement” “implicitness” “evasiveness” “opposition” and so on. Nominalizations allow complete sentences to be simplified into more compact and informative noun phrases, which makes the abstract clearer and more explicit. Hence, more nominalizations should be used by EFL students to make the abstracts more explicit.

The form of Pied-piping constructions belongs to WH relative clauses, which can all be considered as “devices for the explicit, elaborated identification of referents in a text, and are used to specify the identity of referents within a text in an explicit and elaborated manner so that the addressee will have no doubt as to the intended referent” (Biber, 1988, p. 110). There are more Pied-piping constructions in ICJA, for instance:

(9) ...they are indicated as points of comparison on which a team may base their interpretation of the current situation.

(10) ...the elements with which they appear to affiliate trigger interactional misalignment...

(11) The findings underline the importance of contextualizing data and understanding the ecology in which teaching and learning of any subject or any language take place.

The reason for the low occurrence of Pied-piping constructions in CMT abstracts may be caused by the absence of similar syntax in Chinese, which leads many students to be unfamiliar with its rules. Accordingly, they tend to avoid using Pied-piping constructions due to the concerns of negative transfer from the misuse of it. In summary, EFL students are weak in using advanced vocabulary and grammar. To enhance the professionalism and explicitness of the texts, complex words and syntax need to be mastered by the students.

D. Contrastive Analysis on Dimension 4

Dimension 4 measures “Overt Expression of Persuasion”. There is a significant difference between the dimension scores of these two corpora ($p=0.044<0.05$). The dimension score of CMT is higher than that of ICJA, which indicates that the abstracts written by EFL students are more persuasive and explicitly mark the author's point of view.

TABLE 5
THE DIFFERENCES IN LINGUISTIC FEATURES IN DIMENSION 4 (OVERT EXPRESSION OF PERSUASION)

Features	CMT's mean	ICJA's mean	<i>t</i>	<i>p</i>
Necessity modals	-0.299	-0.864	2.448	0.017
Predictive modals	-0.959	-1.286	2.449	0.018
Suasive verbs	-0.315	0.255	-2.200	0.031

There are significant differences between CMT and ICJA on 3 linguistic features of dimension 4 ($P < 0.05$), including, Necessity modals, Predictive modals, and Suasive verbs. Students use more Necessity modals and Predictive modals, but fewer Suasive verbs.

Prediction modals are direct pronouncements that certain events will occur, e.g., will/would/shall; necessity modals are pronouncements concerning the obligation or necessity of certain events, that they should occur, e.g., ought/should/must. Suasive verbs imply intentions to bring about some change in the future, e.g., command, stipulate. These features can mark the assessment of likelihood and the speaker's attempts to persuade the addressee (Biber, 1988). According to Table 5, EFL students prefer to use Necessity modals and Predictive modals. In the CMT abstracts, “should” appears 31 times, and “will” appears 26 times, while in the ICJA these two words each occur only 2 times, for example:

(13) ...the author offers the following suggestions: teachers should strengthen the guidance of learning strategies... (CMT)

(14) ...the translator should sacrifice those features to ensure the readability of the translated text. (CMT)

(15) It is hoped that this thesis will arouse people's attention to ecological issues... (CMT)

As the above examples show, EFL students often use necessity modals to make suggestions when it comes to the moves of significance/importance discussion. Such an expression enhances the persuasive effect of the article, but it reduces the objectivity of the text to a certain extent. In contrast, scholars deliberately avoid the use of necessity modals and predictive modals when discussing the importance of research. Suasive verbs appear more frequently in this move of ICJA abstracts, for instance:

(16) ...stories suggest further implied meanings dealing with the relationships between characters and the authors' manipulations of person references. (ICJA)

(17) We argue that qualitative variation reflects the...and propose a model for experimental designs that can... (ICJA)

(18) Based on these findings, it is recommended that curricula be adjusted to include spaces for multilingual practices... (ICJA)

Based on the above examples, the author argues that, in contrast to necessity modals and predictive modals, suasive verbs enhance the objectivity of the text and convey the author's persuasion as well. Therefore, the use of such modal verbs should be used cautiously, and suasive verbs might be a better choice when discussing the importance and implications of the thesis.

E. Contrastive Analysis on Dimension 5

Dimension 5 is the opposition between "Abstract and Non-Abstract Information". There is no significant difference between the dimension scores of these two corpora ($p=0.682>0.05$). The scores of these two corpora on this dimension are high, which indicates that the texts provide information in a technical, abstract, and formal way.

TABLE 6
THE DIFFERENCES IN LINGUISTIC FEATURES IN DIMENSION 5 (ABSTRACT VERSUS NON-ABSTRACT INFORMATION)

Features	CMT's mean	ICJA's mean	<i>t</i>	<i>p</i>
Predicative adjectives	0.632	-0.468	2.779	0.007
Type-token ratio	-1.899	-4.489	10.086	0.000

There are significant differences between CMT and ICJA on 2 linguistic features of dimension 5 ($P < 0.05$), including, Predicative adjectives and Type-token ratio. Students use more Predicative adjectives compared to the scholars, and the Type-token ratio of CMT's abstracts is higher than that of ICJA.

Predicative adjectives might be considered more fragmented in their function (e.g., the horse is big), while attributive adjectives are highly integrative (e.g., the big horse). Attributive adjectives are used to further elaborate nominal information. They are a more integrated form of nominal elaboration than predicative adjectives or relative clauses since they pack information into relatively few words and structures (Biber, 1988). According to Tables 2 and 6, there are significant differences between CMT and ICJA in both the attributive and predicative adjectives. However, the results show that more predicative adjectives and fewer appear attributive adjectives in CMT, which leads to more fragmented abstracts in CMT, for instance:

(12) And there also exists imbalance and the distribution is not systematic in two sets of textbooks. (CMT)

(13) The Russian translation is strict and tidy, more sticking to the original text and showing more normalization. (CMT)

(14) Abundant audiovisual resources online can ease the pressure of learning and provide more ways to relax, so the use of emotional strategies is relatively frequent. (CMT)

Therefore, the use of adjectives and be verbs should be noted by EFL students because these linguistic features affect the informativeness and formality of the paper.

The type-token ratio stands for the number of different lexical items in a text and is a measure of vocabulary variation within a text, and non-technical informational discourse has a markedly higher lexical variety than abstract, technical discourse (Biber, 1988). In this study, Table 6 illustrates that the Type-token ratio scores of ICJA are lower than those of CMT, which may be because the number of abstract words in ICJA is less than that of CMT. Scholars used more precise vocabulary, which made their abstracts more technical and formal.

F. Contrastive Analysis on Dimension 6

Dimension 6 measures "On-line Informational Elaboration". There is a significant difference between the dimension scores of these two corpora ($p=0.044<0.05$). The dimension score of CMT is lower than that of ICJA, which indicates that the abstracts written by EFL students are more elaborate and planned.

TABLE 7
THE DIFFERENCES IN LINGUISTIC FEATURES IN DIMENSION 6 (ON-LINE INFORMATIONAL ELABORATION)

Features	CMT's mean	ICJA's mean	<i>t</i>	<i>p</i>
Demonstrative pronouns	-0.726	-0.295	-2.925	0.005
That verb complements	-0.220	0.526	-2.388	0.020
That relative clauses on subject position	0.228	2.748	-3.536	0.001

There are significant differences between CMT and ICJA on 3 linguistic features of dimension 4 ($P < 0.05$), including, Demonstrative pronouns, That verb complements, and That relative clauses on subject position. All these 3 linguistic features appear less frequently in the CMT than in ICJA, suggesting that EFL students emphasize the elaboration of abstracts.

Demonstrative pronouns (this, that, these, and those) are associated with informal, unplanned types of discourse (e.g., speech), and mark reduced lexical content and interpersonal involvement, confirming that aspect of the interpretation (Biber, 1988). However, this feature appears more frequently in the ICJA abstracts, for example:

(15) These findings contribute to literature on instructional activities... (ICJA)

(16) its functions are limited to those mentioned... (ICJA)

The reason students use it less may be because Chinese articles often have repetitive nouns and demonstrative pronouns are rarely used to refer to them, while this is not the case in English. Besides, that verb complements, that relative clauses on subject position and WH-clauses (which are similar to that clause, see dimension 1) also appear widely in the ICJA, such as, it shows/argues/reveals that..., it is shown/proposed that, what we found is that..., etc. This might be the result of the writing habits of native speakers of English. In contrast, the language of EFL students' theses is usually refined and revised several times, and students are very careful in the use of language to prevent grammatical errors. As a result, EFL students' abstracts demonstrate a higher level of elaboration.

G. Summary and Discussion

To summarize the findings and discussion of these 6 dimensions, the CMT and ICJA abstracts differed significantly in dimension 4 and dimension 6, which shows that, compared with ICJA, CMT is more persuasive and elaborated. In addition, there are significant differences in 25 linguistic features across 6 dimensions. CMT's scores are higher than ICJA's scores on 11 linguistic features, including Emphatics, Total other nouns, Pronoun it, Be as main verb, Sentence relatives, Type-token ratio, Perfect aspect, Phrasal coordination, Necessity modals, Predictive modals, and Predicative adjectives. While, ICJA's scores are higher than CMT's scores on 14 linguistic features, including Average word length, Nominalizations, First-person pronouns, Demonstrative pronouns, Past tense, Total adverbs, Private verbs, Public verbs, Suasive verbs, Attributive adjectives, Pied-piping constructions, WH-clauses, That verb complements and That relative clauses on subject position. The results suggest that in the thesis abstract of EFL graduate students, more long words, less Be as main verb, Emphatics should be used to enhance the informative of the text; more Nominalizations, Pied-piping constructions could be used to make the text more explicit; more Predicative adjectives should be used instead of Attributive adjectives; more Private verbs, Public verbs, Suasive verbs should appear more frequently; fewer Necessity modals, Predictive modals should appear in the text. There are also features found to be different from the traditional view of EFL academic writing, such as First-person pronouns, Demonstrative pronouns, That verb complements, That relative clauses on subject position.

These results shed light on the teaching and learning of EFL abstract writing, especially for graduate students. EFL students can not only evaluate the shortcomings of abstract writing of CMT according to the scores in the table, but also improve their writing professionalism by imitating the linguistic features used by international experts. For L2 writing teachers, some traditional academic writing views may not be true. To illustrate, first-person pronouns such as "I" and "we" are considered inappropriate in academic writing because of their emphasis on subjectivity, but this does not seem to be the case in ICJA.

The results of the study also have implications for teaching practice. This corpus-driven research can be applied to teaching EFL writing to improve students' academic English writing by instructing students to analyze their papers using the MDA approach. Similar studies have been carried out by scholars: Dong and Lu (2020) integrated corpus-based and genre-based approaches to teaching rhetorical structures in English as a Foreign Language academic writing course at a university in China. Based on the corpus, a move analysis of the introduction section of the selected articles was conducted by students with the help of AntMover (which is a moves analysis tool), to help them understand the rhetorical structures of research articles' introductions and the linguistic features associated with different rhetorical moves. The researchers believe that academic English writing teaching can also be combined with MAT, which allows students to better understand the functional and linguistic features of their theses and make up for the shortcomings in writing.

VI. CONCLUSION

This paper aims to analyze the differences in functional and lexical-grammatical features on English abstracts between CMT and ICJA to help EFL students find the differences between their theses and articles written by experts. Our findings show that the CMT and ICJA abstracts differed significantly in dimension 4 and dimension 6, which shows that, compared with ICJA, CMT is more persuasive and elaborated and there are significant differences in 25 linguistic features across 6 dimensions. The results show that for Chinese EFL graduate students, to make their abstracts more authentic, they have to pay special attention to tenses (Perfect aspect and Past tense), modals (Necessity modals and Predictive modals), nouns (Nominalizations, Phrasal coordination), adjectives (Emphatics, Predicative adjectives, and Attributive adjectives), verbs (Private verbs, Public verbs, Suasive verbs and Be as main verb), pronouns (Pronoun it, First-person pronouns and Demonstrative pronouns), sentence forms (Pied-piping constructions, Sentence relatives,

WH-clauses, That verb complements and That relative clauses on subject position), Average word length and lexical variety (Type-token ratio). The results help learners identify the differences in dimensions and linguistic features between their English abstracts and those written by experts to improve their abstract writing and to write more native-like theses.

Similar to many studies, our study is not exempt from limitations. One limitation is the difference in the size of the two corpora. The abstracts of the master's thesis are longer as students tend to present a more detailed summary to the reader. Therefore, the size of the CMT corpus is larger than that of ICJA. Although MAT has standardization function, the larger size difference reduces the persuasiveness of the results. Another limitation is that the size of the corpus is relatively small. To address these two limitations, in future studies, researchers will use two large corpora of similar size to conduct studies on academic texts. In addition, in future research, the researcher may combine the corpus with English writing teaching by instructing students to use MAT to annotate and analyze their theses to help them understand the functional and linguistic features of English academic writing.

REFERENCES

- [1] Abdollahzadeh, E. (2011). Poring over the findings: Interpersonal authorial engagement in applied linguistics papers. *Journal of Pragmatics*, 43(1), 288-297. <https://doi.org/10.1016/j.pragma.2010.07.019>
- [2] Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511621024>
- [3] Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3-43. <https://doi.org/10.1515/ling.1989.27.1.3>
- [4] Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *Tesol Quarterly*, 36(1), 9-48. <https://doi.org/10.2307/3588359>
- [5] Brugman, B. C., Burgers, C., Beukeboom, C. J., & Konijn, E. A. (2021). From The Daily Show to Last Week Tonight: A quantitative analysis of discursive integration in satirical television news. *Journalism Studies*, 22(9), 1181-1199. <https://doi.org/10.1080/1461670X.2021.1929416>
- [6] Crosthwaite, P. (2016). A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes*, 22, 166-178. <https://doi.org/10.1016/j.jeap.2016.04.005>
- [7] Dong, J., & Lu, X. (2020). Promoting discipline-specific genre competence with corpus-based genre analysis activities. *English for Specific Purposes*, 58, 138-154. <https://doi.org/10.1016/j.esp.2020.01.005>
- [8] Feng, E. Y., & Wu, L. (2016). Guo nei wai guang xue lei ke ji qi kan lun wen ying wen zhai yao ti cai dui bi fen xi [A contrastive genre analysis of English abstracts from Chinese and international journals in optics]. *Zhong guo ke ji qi kan yan jiu*, (02), 230-236.
- [9] Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. *Journal of Second Language Writing*, 26, 80-95. <https://doi.org/10.1016/j.jslw.2014.09.007>
- [10] Gregg, N., Coleman, C., Stennett, R. B., & Davis, M. (2002). Discourse complexity of college writers with and without disabilities: A multidimensional analysis. *Journal of Learning Disabilities*, 35(1), 23-38. <https://doi.org/10.1177/002221940203500103>
- [11] Halliday, M. A. (1988). On the language of physical science. In M. Ghadessy (Ed.) *Registers of written English: Situational factors and linguistic features* (pp. 162-178). Burns & Oates.
- [12] Hyland, K. (2000). *Disciplinary discourse: Social interactions in academic writing*. London: Longman.
- [13] Nini, A. (2015). *Multidimensional Analysis Tagger* (Version 1.3). Retrieved March 19, 2022, from <https://sites.google.com/site/multidimensionaltagger/>
- [14] Paltridge, B., & Starfield, S. (2020). *Thesis and Dissertation Writing in a Second Language: A Handbook for Students and their Supervisors*, 2nd ed. New York/London: Routledge. <https://doi.org/10.4324/9780203960813>
- [15] Pan, F. (2018). A multidimensional analysis of L1-L2 differences across three advanced levels. *Southern African Linguistics and Applied Language Studies*, 36(2), 117-131. <https://doi.org/10.2989/16073614.2018.1476162>
- [16] Pan, F. (2022). Duo wei du yan jiu san shi nian fa zhan zong shu he fang fa xiang jie [A review of 30-year development of multi-dimensional analysis and step-by-step procedures]. *Wai yu jiao xue li lun yu yan jiu*, (01), 26-34.
- [17] Qiu, G. Z. (2015). Zhong mei yi xue qi kan ying wen zhai yao mo hu xian zhi yu dui bi fen xi [A comparative analysis of hedging in English abstracts of Chinese and American medical journals]. *Zhong guo ke ji qi kan yan jiu*, (11), 1212-1216.
- [18] Sardinha, T. B., & Pinto, M. V. (2019). Dimensions of variation across American television registers. *International Journal of Corpus Linguistics*, 24(1), 3-32. <https://doi.org/10.1075/ijcl.15014.ber>
- [19] Sun, L. (2015). Zhong guo shuo shi xue wei lun wen ying wen zhai yao de yu yong shen fen jian gou yan jiu [Study of Chinese MA Students' Construction of Pragmatic Identities in English Thesis Abstracts]. *Wai yu yu wai yu jiao xue*, (05), 15-21. <https://doi.org/10.13458/j.cnki.flatt.004164>
- [20] Tseng, F. P. (2011). Analyses of move structure and verb tense of research article abstracts in applied linguistics journals. *International journal of English linguistics*, 1(2), 27. <https://doi.org/10.5539/ijel.v1n2p27>
- [21] Veirano Pinto, M. (2014). Dimensions of variation in North American movies. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional Analysis, 25 Years on: A Tribute to Douglas Biber* (pp. 109-146).
- [22] Xie, Q. & Ma, Y. Y. (2021). Zhong wai bo shi xue wei lun wen ying wen zhai yao de dui wei du dui bi fen xi [A Multi-dimensional Contrastive Analysis of English Abstracts in Dissertations by Chinese and Native PH.D. Students]. *Wai yu xue kan*, (04), 39-44. <https://doi.org/10.16263/j.cnki.23-1071/h.2021.04.006>
- [23] Xie, S. (2020). Multidimensional analysis of Master's thesis abstracts: a diachronic perspective. *Scientometrics*, 123(2), 861-881. <https://doi.org/10.1007/s11192-020-03408-6>

- [24] Zhang, Y. N., Sun, C. H., & Li, Y. (2018). Zhong wai yu yan lei qi kan gao bei yin lun wen yin wen zhai yao yu yan te zheng duo wei du fen xi [Multi-dimensional Analysis of the English Abstracts of the Most-Cited Articles in Linguistics Journals]. *Wai yu dian hua jiao xue*, (04), 64-71.

Yifan Shi, born in Beijing, China in July 1999, is currently a graduate student from the School of English Language, Literature and Culture in Beijing International Studies University, China. His current research interests include creative writing, academic writing, corpus linguistics, etc.

Zhencong Liu, born in Guangxi, China in September 1968, PhD in linguistics, now is a professor of the School of English Language, Literature and Culture in Beijing International Studies University, China. His current research interests include general linguistics, cognitive linguistics, cross-cultural communication, English teaching theory and practice, etc.