SciencePG
Science Publishing Group

A Comparative Study of a Multi-Dimension/Multi-Feature Approach Between Chinese Debate and Speech

Liu Yanchun[1], Hu Fengguo[1], Zhao Yi[2], Qiu Le[1]

[1]Faculty of Literature and Law, Communication University of China, Beijing, China

[2]Academy of Media and Public Affairs, Communication University of China, Beijing, China

Email address:
liuyanchun@cuc.edu.cn (Liu Yanchun)

Abstract: This study compares and analyzes Chinese debate and speech from the perspective of registers, which makes itself significant since it fills the gaps in the field. This paper is based on a self-built corpus and refers in Biber's Multi-dimension/Multi-feature Approach. The quantitative statistical results show that there are 44 significant differences among the 65 linguistic features of debate and speech registers, and that these feature differences, from the macroscopic view, can be summarized into 8 different dimensions, which, to be specific, are known respectively as Multi-directional Interaction VS Single-directional Communication, Demonstrative VS Narrative, Intense Confrontation VS Deliberate Storytelling, Centralized Focuses VS Dispersing Contents, Precise Expressions VS Diverse Methods of Expressions, Informative VS Affective, Specialized VS Universal, along with Literate Style VS Oral Style. Why there are these linguistic features and dimensional differences between the two registers is also explained. It could be fair to say that this study makes a breakthrough on the basis of Biber's research which mainly put its focus on linguistic features at the lexical, syntactic perspectives. And that is to advance comparing linguistic features between debate and speech registers to perspectives of phonetics and speeds. Yet it has its deficiency for failing to conduct factor analysis in the process of dimension induction due to the limited volume of text corpus.

Keywords: Debate, Speech, Multi-dimension, Multi-feature, Comparison

1. Introduction

Great achievements in studying debate and speech have been made in China, yet few of them are done from the perspective of register. So far, I have found no monographs of register studies on debate or speech, with relevant articles less than 20. Nor have I seen any works comparing registers between debate and speech. Hence, this study is somewhat significant since it fills the gaps.

Quantitative methods have yet rarely been taken into account in terms of Chinese register studies, with a few involved going no further by merely investigating single or a few linguistic features. In the 1970s, some linguists abroad argued that similarities and differences among registers cannot be fully disclosed by analyzing individual linguistic features. Both Ervin-Tripp [1] and Hymes[2] insisted that "co-occurrence" should be used as a linguistic feature for register comparison. Yet confined to corpus and techniques at

the time, it was not until the 1980s when Douglas Biber developed the multi-dimension/multi-feature approach that a better register comparison through multiple co-occurring linguistic features was accomplished.

Biber's multi-dimension/multi-feature approach is to identify several (usually dozens) linguistic features of different registers and have their frequencies compared statistically, which leads to a microscopic comparison among multiple features. Then these features would be generalized into several dimensions for macroscopic multi-dimensional investigation. And finally, functional explanations would be given according to situational characteristics. In his research process, Biber adopted qualitative methods for linguistic feature extraction, dimension identification and naming, as well as functional explanation in accordance with situational characteristics, and quantitative methods for judging

differences among linguistic features, analyzing co-occurring factors and calculating dimension scores. In this way, Biber's multi-dimension/multi-feature approach would not bring about a biased result in analyzing registers as before when only single or a few linguistic features were focused, but also make itself a combination of qualitative and quantitative methods, macro- and micro- analysis, which helps to avoid the limitations caused by adopting a single method or analyzing from a single angle.

Biber proposed the multi-dimension/multi-feature approach in the 1980s and applied it into comparative studies of different registers for the following years [3-11]. Multi-dimension/multi-feature approach was introduced into China in the 2000s (see Lei Xiuyun [12] and Wu Jiangsheng [13, 14]) and has been used since then for research, such as Ma Guanghui [15], Gui Shichun [16], Wen QiuFang [17], Hu Xianyao [18] etc. Most of them focused on English texts, though. Biber's multi-dimension approach was based on comparative analysis of English texts, which varies considerably from Chinese. Therefore, some linguistic features identified by Biber like derivation, that-clause, WH-clause with prepositional preposition etc. are not applicable for Chinese. Plus, investigation of multi-dimensions requires complicated statistical computations and software uses. As a result, few scholars in China have adopted this approach for register studies so far.

This paper attempts to comparatively study the registers between debate and speech via the multi-dimension/multi-feature approach, with emphasis on presenting an overall view of similarities and differences between debate and speech by comparing multiple linguistic features of them. The results could be used to improve apparent effects of register teaching in education, Chinese-English translation accuracy in translation, as well as operability of computer recognition for registers in information processing. Also, it could be assumed as a preliminary for multi-dimension studies.

The corpus used here was on the basis of the 4th Televised Speech Contest for Military University Students of China, which was themed with Military Revolution and included the speech session and the debate session. In details, recordings of 36 debaters in 6 debates, and speech recordings of 33 speakers were collected and transcribed. There is a total of 74019 Chinese characters (including punctuations) with debate 36536 and speech 37483, and of 46449 words with debate 21082 and speech 25367. Shi Xiaodong's Segmentation System for Chinese and SPSS 18.0 were adopted for word segmentation, comparison of means and independent samples t-test.

2. Comparison of Multiple Features Between Debate and Speech

Different situational characteristics and communicative purposes determined the frequencies of the same linguistic features in different registers. Through investigation and hypothesis testing of frequencies of linguistic features, a deep insight into similarities and differences of linguistic features among different registers is available. According to situational characteristics of debate and speech, along with communicative purposes, this paper identified 65 linguistic features and here are the results of quantitative comparisons.

Table 1. *Comparison of Debate, Speech and Normal Speeds.*

Debate		Normal	Speech
303 characters/min		244 characters/min	226 characters/min
Opening and Closing Statement	Rebuttal, Free Debate		
287 characters/min	313 characters/min		

2.1. Speed

Based on the fact that debate and speech are of different communicative purposes, in addition to our perception of reality, considerable differences between the speeds of them were expected. And a timing statistic was conducted for confirmation. The result showed that: they did bear much difference, with the average speed of debate reaching 303 characters per minute and speech 226 characters per minute[1], while the normal speed would be 244 characters per minute

[19].

2.2. Register Markers and Common Words

Debate register has markers, such as my fellow debaters （对方辩友） and Thank you, chairman（谢谢主席）, while speech doesn't have any. High-frequency words of debate are always related to reasoning, logistic, negation and stance, while those of speech are generally temporal and position nouns, toponyms or recollective words. Common words normally account for more than 10% of the two registers. For example, common words that are typical of debate register make up 11.40% of debate (the same words comprise merely 5.62% of speech); while common words that are typical of speech register make up 12.65% of speech (the same words comprise merely 4.72% of debate).

---

[1]This study tested the speed of the same persons in debate and speech (two people took part in speech and the opening and closing statement parts of debate) to eliminate bias caused by individual differences from statistical results. The result showed their speeds in speech are 226 characters per minute and 224 respectively, almost the same as the average speed in speech. And their speeds in the opening and closing statement in debate are 279 and 295, respectively, close to the average speed of opening and closing statement (which is 287 characters/ min). Thus, it is fair to consume that the statistic results of this study are unbiased and immune to individual differences.

Table 2. Comparisons of Register Markers and Common Words between Debate and Speech.

Register	Markers	Common Words		Percentage
Debate	对方辩友;谢谢主席;请问等. Translation: my fellow debaters; Thank you, chairman; May I ask that… etc.	逻辑，观点，事实，解释，如果，那么，因为，所以，首先，第一，不，没有，认为，当然，必定，应该，必将等 Translation: logistics, opinions, the truth, explain, if, then, because, therefore, first of all, at first, no, there is not, think, of course, necessarily, should, surely will etc.		11.40
Speech		地下，身上，手里，深入，眼中，去年，今日，此时，今朝，每天，如今，曾经，当初等 Translation: underground, (have sth.) on one, in hand, go deep into, in the eye of, last year, this day, at the moment, for the moment, every day, now, once, at that time etc.		12.65

2.3. HF Words

Table 3. Comparison of the Top 20 HF Words between Debate and Speech.

Debate	Frequency Sum of Top 20 HF Words	Percentage Sum of Top 20 HF Words	Words and Data	Word	的	是	对方	战争	辩友
				Frequency	1471	725	464	402	369
				Percentage	7.5	3.7	2.37	2.05	1.77
	6629	33.68		Word	我	技术	军事	个	呢
				Frequency	210	201	199	187	184
				Percentage	1.07	1.02	1.01	0.95	0.94
Speech	Frequency Sum of Top 20 HF Words	Percentage Sum of Top 20 HF Words	Words and Data	Word	的	我	了	以	在
				Frequency	1740	549	419	345	343
				Percentage	8.82	2.78	2.12	1.75	1.74
	5550	28.13		Word	就	他	这	个	中
				Frequency	149	137	136	122	119
				Percentage	0.76	0.69	0.69	0.62	0.6

Table 3. Continued.

Debate	Frequency Sum of Top 20 HF Words	Percentage Sum of Top 20 HF Words	Words and Data	Word	了	不	一	我们	在
				Frequency	362	305	281	246	243
				Percentage	1.85	1.56	1.43	1.25	1.24
	6629	33.68		Word	越来越	信息	就	说	发展
				Frequency	177	161	160	144	138
				Percentage	0.9	0.82	0.82	0.73	0.7
Speech	Frequency Sum of Top 20 HF Words	Percentage Sum of Top 20 HF Words	Words and Data	Word	是	着	我们	不	和
				Frequency	342	183	159	158	150
				Percentage	1.73	0.93	0.81	0.8	0.76
	5550	28.13		Word	那	有	地	上	与
				Frequency	116	109	98	95	81
				Percentage	0.59	0.55	0.5	0.48	0.41

HF words refer to the top several high-frequency words in a text, with words sorted by descending order of frequency. Comparison of HF words helps to learn the concentration, states of content words and function words, nouns and pronouns et al. of different registers. The information would be further referred to for defining the dispersion of word uses, informativity and context-dependency between registers. The results show that the sum of tokens of top 20 HF words comprised 33.68% of the total in debate, and 28.13% in speech. It reveals that debate register has more high-frequency types and they are more concentrated. Besides, HF words of debate are composed of 75% content words, which are mostly nouns. The proportion of speech is 55% with pronouns in the majority. That is to say, the debate register is more informative than speech register, while speech is more dependent on contexts.

2.4. Rare Words

Rare words refer to low-frequency words, with a case in point being those that occur once and thus share a frequency of 1. The number of types or words that occur once is 3020 in debate, twice larger than speech. And in each frequency segment less than 30, speech has more types than debate; while in each segment over 30, debate has more types than speech. The fact that speech has more rare words and debate more high-frequency words makes types in speech nearly twice higher than that in debate, while the total tokens of the two are much the same. More rare words equals to a more dispersed use of types. That speech is more dispersed than debate in type uses confirms the statistical result of HF words above.

Table 4. Comparison of Word Frequency Distribution between Debate and Speech.

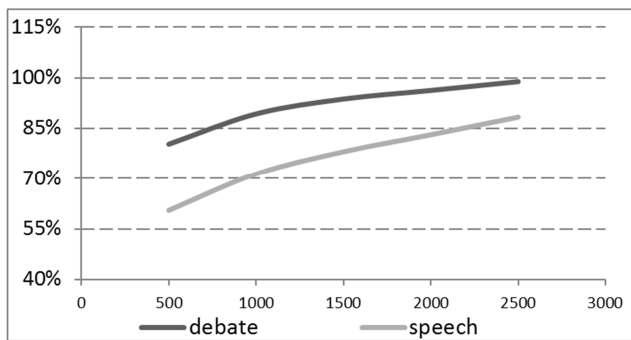| Register	Frequency	1	2-5	6-10	11-20	21-30	31-50	51-100	101-200	>201	Total Types	Total Tokens
Debate	Types	1364	839	228	131	50	46	37	19	12	2726	19611
Speech	Types	3020	1596	276	141	55	41	25	12	6	5172	19726

Figure 1. Comparisons of Coverage Ratio between Debate and Speech.

2.5. Coverage Ratio

According to Hou Min [20], coverage ratio is a percentage of specified subjects in the total amount of the examined corpus, namely the ratio of the cumulative sum of the frequency of each subject and its previous one to the total number of words of all subjects, with all subjects sorted by a descending order of frequency. The calculation formula is: $A_i = \sum_{k=1}^{i} \frac{n_k}{N} \times 100\%$,nk is the frequency of subject k, N the total of all subjects, A the cumulative frequency and Ai is the cumulative frequency of A1 to Ai. This study took 500 words as a pitch and divided debate and speech into 5 subsections respectively of 500, 1000, 1500, 2000 and 2500 words. It is known from Illustration 1 that the coverage ratios of debate in all subsections are higher than speech. More HF words lead to higher recurrence ratio and coverage ratio of types. That debate has a higher coverage ratio than speech confirms the statistical result of HF words and rare words.

2.6. Normalized TTR

Type/token Ratio or TTR is generally used as an index for lexical diversity. However, it is confined to the text size, because a larger text brings about a lower TTR. So we had the texts normalized to acquire normalized TTR. The detailed operation was to compound the texts of the debate and speech into two big texts respectively and divided them into smaller ones with each containing 1000 words (including punctuations, and the word count of the last text is its actual amount if it's less than 1000). Then we got 22 small texts of debate and 26 of speech, and we calculated the TTR of each. Finally, we compared the mean values and independent samples t-test to find out if there are considerable differences of the normalized TTR between debate and speech. It could be seen that the average of normalized TTR of speech is higher than that of debate, and Sig. value of t-test is 0.000, which confirms that speech is significantly different from debate in normalized TTR and much more diverse lexically.

Table 5. Comparison of Normalized TTR between Debate and Speech.

	Mean			Levene's Test for Equality of Variances		T-test for Equality of Means		
Normalize d TTR (%)	**Debate**	**Speech**		**F**	**Sig.**	**t**	**df**	**Sig.(2-tailed)**
	35.318	46.977	Equal variances assumed	3.522	.067	-5.551	46	.000
			Equal variances not assumed			-5.375	35.634	.000

2.7. Lexical Density

There are various ways to get lexical density and that the number of content words (or lexical words) divided by the total number of words is usually adopted. As is often the case, a higher lexical density makes a text more informative. The statistical result show that debate bears a slightly higher lexical density than speech, yet there is no much difference between them.

Table 6. Comparison of Lexical Density of Debate and Speech.

	Mean			Levene's Test for Equality of Variances		T-test for Equality of Means		
Lexical Density	**Debate**	**Speech**		**F**	**Sig.**	**t**	**df**	**Sig.(2-tailed)**
	82.088	80.661	Equal variances assumed	1.877	.179	1.373	37	.179
			Equal variances not assumed			2.06	12.434	.061

2.8. Word Length & Sentence Length

In this section, the word length and the average word length of words with syllables ranging from 1 to 7, along with average length of complete sentences and clauses[2] were investigated. Previous studies found that a register of language, if more literate, has a higher share of disyllables; while if more oral, it has a higher share of monosyllables [21]. And literate registers have an advantage over oral registers in terms of average word length and sentence length. The statistical result show that the proportion of disyllables in debate is slightly over that of monosyllables, meanwhile the situation in speech is to the contrary, in which the proportion of disyllables is much lower than that of monosyllables. The proportion of disyllables in debate is much higher than in speech, and the averages of word length, complete sentence and clause lengths are also slightly larger in debate. These linguistic features reflect an intensive literacy of debate register over speech.

2.9. Word-Class Distribution

30 items were identified for word-class investigation in

---

2 A complete sentence is the sentence that ends with period, question mark, or exclamation mark, and a clause is the segment that ends with comma, semicolon or colon in a sentence. The average lengths of complete sentences and clauses refer to the average length of a complete sentence and that of a clause.

accordance with the situational characteristics of debate and speech. Mean comparison and t-test saw significant differences of word-class distribution in 18 items, and none in the other 12. And word-class distributions of 9 classes are higher in debate than in speech and they are noun, verb, adverb, yuqi word, conjunction, interrogative pronoun, temporal word, particle, and "得" particle. Among them, the first 4 are of significant differences while the latter 5 are not. Besides, word-class distributions of 21 classes are higher in speech than in debate with 14 of them such as toponym, personal pronoun present significant differences, while the other 7 classes like person name and adjective don't.

Table 7. Comparisons of Word Length and Sentence Length between Debate and Speech.

Register	Percentage of Word Length							Average Word Length	Amount of Sentences and Sentence Length			
	1	2	3	4	5	6	7		Complete Sentence	Word/Complete Sentences	Clause	Word /Clauses
Debate	47.39	48.60	3.09	0.81	0.09	0.01	0.01	1.58	1035	18.95	2982	6.58
Speech	57.51	38.21	2.46	1.66	0.13	0.02	0.01	1.49	1066	18.50	3261	6.05

Table 8. Comparison of Word-Class Distribution between Debate and Speech.

Higher Proportion in Debate over Speech	Significant Differences	noun, verb, adverb, yuqi word
	Insignificant Differences	conjunction, interrogative pronoun, temporal word, particle, and "得" particle
Higher Proportion in Speech over Debate	Significant Differences	preposition, position word, toponym, location word, personal pronoun, numeral, quantifier, "了", "地" and "着" particles, idiom, locution, lettered-word, state word
	Insignificant Differences	pronoun, adjective, determiner, abbreviation, person name, onomatopoeia, "的" particle

2.10. Sentences

18 linguistic features like question sentence, complex sentence, negative, ellipsis and exclamation, parallelism and citation were investigated on the basis of situational characteristics and communicative purposes of debate and speech. It is obvious that debate has higher proportions of condition complex sentences, reason complex sentences, fanwen question[3], shewen question[4], WH-question, negation, citation, while speech has higher proportions of exclamation, metaphor or simile, analogy, parallelism and iteration. Also, debate has brief response and ellipsis that speech doesn't, and speech has metonym, antithesis, hyperbole and pun that debate doesn't.

Brief response and ellipsis in debate occur during the free debating stage. For instance, "Of course not." "OK." "Of course" "Thanks." "Certainly." "Definitely not." "Of course yes." "Good." "Great!" Brief response in debate differs from that in normal interactions, under which turn-taking usually takes place following the response. While in debate, more detailed explanations or demonstrations would be given before turn-taking. Here are some examples.

A1：请问对方辩友，一个乐队里没有乐器，是不是不能奏乐?而乐器是不是这个乐队的主导呢？

B1: 当然不是。但是一个乐队有了乐谱它就是乐队了吗？难道你要说乐谱是这个队的主导吗?那么指挥是什么？

C1: 能不能请对方辩友给我们列举出一个，哪怕是只有一个，哪项技术的运用使战争的时间越来越长了呢？

D1: 当然有。同样是占领一个国家的战争，德国进攻波兰只用了 35 天的时间，比美国近期几场战争的时间要短得多。难道对方辩友要说，正是因为那时候德国的技术要比现在的美国先进吗？

Translation:

A1: My fellow debater, cannot a band play if it doesn't have any instrument? And that makes the instruments the leader of the band?

B1: Of course not. However, does a music sheet make a band? Does it mean that the sheet is the leader of the band? Then what's the instructor for?

C1: Could the fellow debater name even one case, if not many, where the employment of certain technique length the war time?

D1: Sure. Under the similar circumstance of occupying a country, Germany spent 35 days conquering Poland, much shorter than the several battles America involved recently. In this case, are you arguing that Germany owned advanced techniques then than America does now?

In the examples above, both B1 and D1 used ellipsis and brief response. The complete expression of B1 should be "Of course instruments are not the leaders of the band.", and that of D1 should be "Surely there are cases where the employment of certain technique length the war time." But the complete form seems to be too verbose and less powerful in free debating. As a result, debaters tend to use ellipsis and brief response, and explain or argue in details afterwards. It agrees with the specific situational characteristics and communicative purposes of debate. For instance:

(1) 桂林的山太柔了，它承受不起巨大的伤痛；桂林的水太弱了，它承载不住绵长的忧愁。

(2) 仰望蓝天，我豪情依旧；面对挑战，我壮志满怀。在这场深刻的军事变革中，我将用理想、信念和忠诚标定人生的坐标，用责任、素质和奉献托起腾飞的翅膀，实现人生的第二次起飞！

(3) 我知道，只要霸权主义存在，阴云就会时时遮住和平的阳光；我知道，如果我们的新军事变革不能持续快速地推向前进，我们就难以积蓄起抗击强权的力量。

(4) 在他不大的办公室里，在堆积如山的资料和奖杯奖状中间，从一幅他亲手书写的近代爱国诗人丘逢甲的《春愁》里，你会找到答案："春愁难遣强看山，往事惊心泪欲潸。四百万人同一哭，去年今日割台湾。"

---

3 A *fanwen* question is a figure of speech in the form of a question that is asked to express attitudes while requires no answer because it is already known.

4 A *shewen* question is a figure of speech in the form of a question that is asked to attract attention and it's answered by the questioner.

Translation:

Mountains in Guilin are too supple to bear great pains, and its waters too fragile to take perpetual sorrows.

I am proud when looking up to sky and ambitious when faced with challenges. In this profound military revolution, I will calibrate the coordinates of life with ideal, faith, and loyalty, lift the soaring wings with responsibility, competency and dedication, and have my life had a second take-off.

I am aware that, if only the hegemony exists, the sunshine of peace would be overshadowed by dark clouds from time to time; and that if our new military revolution cannot be pushed forward constantly and rapidly, it would be hard for us to gather strength to fight against the great powers.

In his humble office, among those piled documents, trophies and award certificates, you could find the answer from a calligraphy work of his own that is titled with Spring Sorrow by the patriotic poet Qiu Fengjia: Spring sorrow is so overwhelming that I forced myself to enjoy the scenery; the past is too startling for me to hold back my tears. Four million people wept at the same time, for that Taiwan was occupied this day of last year.

Both (1) and (2) used antitheses in the first sentences, "第二次起飞" (a second take-off) in example (2) is a pun, "阴云" (dark clouds) in (3) is a metonymy and "四百万人同一哭" (Four million people wept at the same time) in (4) is a hyperbole. Such figures of speech are missing in debate.

Table 9. Sentence Comparisons between Debate and Speech.

Sentence	Fanwen	WH-question	Condition Complex	Reason Complex	Shewen	Negation	Citation	Brief Response	Ellipsis
% in Debate	21.04	11.11	7.63	6.76	3.82	2.37	1.53	0.37	0.10
% in Speech	2.35	3.47	3.75	3.56	1.41	0.96	0.84	0.00	0.00
Sentence	Exclamation	Metaphor or Simile	Analogy	Parallelism	Iterations	Metonym	Antithesis	Hyperbole	Pun
% in Debate	0.87	4.12	2.19	0.19	0.10	0.00	0.00	0.00	0.00
% in Speech	20.08	8.16	4.50	2.91	2.72	1.78	0.84	0.66	0.47

3. Multi-Dimension Differences Between Debate and Speech

Multi-feature analysis offers a thorough and comprehensive view of similarities and differences among different registers, yet which seem to take efforts to grasp overall with so many items and being too microscopic. Besides, some linguistic features generally co-occur as an indicator for a certain function. Factor analysis could help researchers to find out these co-occurred linguistic features and reduce the dimensions of register differences. Dimension reduction, to be exact, is decreasing dozens of linguistic features to several ones so that register differences could be better understood on the whole. Biber et al. [11], in their comparative studies of spoken and written registers in American universities, conducted dimension reduction by inducing the 90 linguistic features into 4 dimensions. So they had a macroscopic comparative study of registers. Another thing that should be noticed is that dimensional differences of registers are not static. When comparing different registers, Biber usually concluded several different dimensions on the basis of specific registers. Also, some linguistic features do not always occur in the same dimension. One single or several features occasionally co-occur in different dimensions and indicate different functions. 65 linguistic features of Chinese debate and speech were presented in the second part of this paper and 44 of them appear significant differences. On this basis, together with other insignificant differences, we summarized 8 dimensions of the register differences between debate and speech.

3.1. Multi-Directional Interaction VS Single-Directional Communication

Debate has much more verbs, adverbs, fanwen questions,

shewen questions, and WH-questions than speech, as well as brief responses and ellipses, all of which are typical characteristics of interactive registers. Debate is interactive in nature. Its interactivity is beyond debaters and also embodied in the simple interaction between debaters and the chairman (like "OK, time's up!" "Thank you, chairman."), which makes debate multi-directional interactive. And speech is basically single-directional communicative, thus not equipped with the linguistic features of interactive registers.

3.2. Demonstrative VS Narrative

In Chinese speech, the frequencies of personal pronouns, "了" particle of past and perfect tense, and "着" particle of progressive tense are much higher than in debate. The result agrees with what Biber [6] wrote about linguistic features in narrative dimensions. Besides, the frequencies of prepositions, position words, location words, person names and toponyms in Chinese speech are higher than in debate. These features jointly indicate that speech is narrative. On the other hand, debate has a higher lexical density, and more nouns, yuqi words, conjunctions, reason complex sentences, condition complex sentences and citation. These features make debate demonstrative by being informative, explicitly expressive, intense in speaking and logic-emphasized, also citing extensively to strengthen its arguments.

3.3. Intense Confrontation VS Deliberate Storytelling

The speed of debate is quite faster than speech and the normal speed of speaking, which plus much more negations, fanwen question, shewen question, WH-questions and yuqi words, make debate confrontational. Negation, interrogating and questioning are generally used for confrontation, while higher speed, more verbs, adverbs and yuqi words have something to do with the inherent requirement of refutation and the pursuit of intense effects. On the contrary, speech has

a slower speed, more pronouns, adjectives and person names, more "的"particle. And its prepositions, toponyms, location words, personal pronouns, metaphors or similes and exclamations are way too many over debate. Those agree with its communicative purposes to express thoughts or emotions, and arouse empathy by narrating and portraying. Only via deliberate storytelling with its reasonable narrative rhythm and slower overall speed, can the audience have a better understanding of the thoughts and emotions of the speaker.

3.4. Centralized Focuses VS Dispersing Contents

Lower normalized TTR, less word types and rare words, more HF words, and higher coverage ratio of word types indicate that debate is centralized in word type uses with higher recurrence ratio of same types, which is a reflection of centralized contents of debate register. In contrast, higher normalized TTR, more word types and rare words, less HF words, and lower coverage ratio of word types, along with lower recurrence ratio of same types indicate that speech is relatively dispersing in its content.

3.5. Precise Expressions VS Diverse Methods of Expressions

Debate, with argumentation being predominant, requires rigorous logics and explicit expression, while avoids ambiguity. That explains why debate has much more nouns, condition complex sentences, reason complex sentences and citations than speech and doesn't use figures of speech like metonymies or puns. The case is different in speech, though. To make their speeches more interesting and graphic, speakers usually make use of multiple methods: Firstly, lexical diversity (more rare words and higher normalized TTR), which is related to the content dispersion of speech, and to the selection and pursuit of lexical diversity of the speaker. Secondly, more dispersed distributions of different syllables, which for debate are mostly mono-, di- and tri-syllables, and for speech are composed of more quadri- and quinque-syllables that make words more variable. Thirdly, idioms, locutions, abbreviations, lettered-words, state words, onomatopoeias generally appear much more frequently in speech than in debate. Finally, speech has more exclamations, metaphors or similes, analogies, parallelism, and recurrences, along with more metonymies, antithesis, hyperboles and puns that don't show up in debate.

3.6. Informative VS Affective

Debate, with content words taking a high share of its HF words, more nouns and conjunctions, higher lexical density, and larger averages of word lengths, complete sentence lengths and clause lengths, are more informative than speech. Speech, with its high proportions of adjectives, exclamations, parallelism, recurrence, hyperbole, is relatively more affective.

3.7. Specialized VS Universal

Terms occurred frequently in debate, with quite a few just in

HF words, like war, military, technology, information, let alone in the whole texts, such as asymmetry, nonlinearity, operation modes, military technologies, military structure, leading elements, system engineering, and mechanized revolution. These terms make the debate register specialized in a way. Though "military revolution" is also the topic of speech, such military terms barely occurred in speech. Speakers basically tell stories that happen in daily life and use daily vocabularies, which make the language of speech universal to the public.

3.8. Literate Style VS Oral Style

Debate adopts many terminologies to make it more literate, which is also confirmed by its slightly higher proportion of disyllables over monosyllables. The proportions of monosyllables and quantifiers are higher in oral style registers than literate style registers, see Liu Yanchun [21-23].In speech, the proportion of monosyllables is much higher than that of disyllables, so are the proportions of monosyllables and quantifiers over debate. That makes speech more of oral style or colloquial than debate, which is also confirmed by the fact that the proportion of disyllables is much higher in debate, plus that the lexical density, averages of word lengths, complete sentence lengths and clause lengths of debate are slightly higher over speech.

4. Functional Explanations for Differences of Linguistic Features and Dimensions between Debate and Speech

Biber et al. [11] believed that linguistic features, instead of being arbitrary, are in accordance with the specific situation of different registers. Hence, only the combination of linguistic features and situational characteristics can basic dimensions of registry varieties be better explained and described. Biber offered a frame for analysis of situational characteristics, which is universally applicable for any registers to analyze the situational characteristics, or to be exact, to analyze the relationship between participants, channels, production circumstance, settings, communicative purposes and topics. In terms of these items, analyses of how linguistic features and dimensional differences of Chinese debate and speech correspond with the situational characteristics are given below.

4.1. Participants

This item could be investigated from the aspects of the addressor and the addressee. In speech, the addressor is a single person, and the addressee is a group of people, who may not be provided with the involving knowledge. It requires the speech to be less professional in the contents and the language so that audiences could understand and accept the speaker's thoughts and emotions as much as possible. That's why dictions of speech are basically from daily life, hardly

involving terms. And that gives the speech the feature of being universal. However, in the case of debate, the addressor and the addressee are mostly debaters of both sides (the chairman and judges also need to learn about the proposition in advance), who are prepared and ready for arguments to defense the viewpoints of his side. That's to say, they don't need to avoid being professional. Instead, more professionally a debater behave, more persuasive his arguments is. That explains the occurrences of terms in debate, which gives the debate the feature of being specialized.

4.2. Relationships Between Participants

Judging from the relationship between participants, speech is basically away from interactive (it is occasionally interactive, yet the interactions are pretty brief and merely used to arouse emotions or attract attentions of the audience), and debate is highly interactive, especially during the free debating phase. So there are some typical interactive linguistic features in the debate register, like addressing, responding and ellipsis etc. Also, because the two parties of debate have confrontation against their opinions, debate has a high speed and many negative words and sentences, fanwen and shewen questions. In speech's case, the speaker tends to slow down and tell stories deliberately so audiences would understand his thoughts and emotions.

4.3. Channels

Both debate and speech are in spoken forms, so the averages of word lengths, complete sentence length, and clause lengths of them are relatively smaller on the whole, and both of them have the feature of being colloquial. However, debate is a little more of literate style than speech because of its rigorous logics, standard languages, and large numbers of terms, conjunctions, logic words and disyllables, along with its slightly higher lexical density over speech.

4.4. Production Circumstance

Though in spoken forms, debate and speech can be well prepared like being planned and written in advance. That's why repetition, pleonasm or pauses etc. that are common in spoken language are missing from both. And they are both careful with dictions and strict with structures. Besides, to make the speech more variable and attractive, the speaker would employ multiple techniques of expressions and figures of speech, for example, controlling the recurrence ratio of words, using extensive rare words, collating different syllables, using idioms, locutions, abbreviations, lettered-words, onomatopoeias as well as metaphors or smiles, analogies, antitheses, hyperboles and puns and so on.

4.5. Setting

The setting refers to the physical context of the communication – the time and place. Though both sides of the debate, speaker and audience of the speech share the time and space, the contents of debate and speech actually are not necessarily related to the scene. Speech, with its contents focusing on recollection and reconstruction of the past, has many temporal words, toponyms, and position words that related to the past, along with recollective words, as well as more "了" particle of past and perfect tenses, and "着" particle of progressive tense over debate, all of which makes speech more narrative. And debate, with its contents focusing on the proposition, has more related terms and logic word relevant to ways of argumentation, and thus appear to be more demonstrative.

4.6. Communicative Purposes

As debate aims to support the viewpoints of their own side and refute the opposite, it has large numbers of register markers like "my fellow debaters", nouns, conjunctions, reason complex sentences, condition complex sentences, citations, and words that used to express standpoints and logical relations, and it appears to be highly informative and demonstrative. As speech aims to express thoughts and emotions, and arouse empathies through reconstructions of stories and imitation of objects, it has many related personal pronouns, person names, toponyms, position words, particles of "了" and "着" and appears to be highly narrative. Besides, large quantities of adjectives, state words, metaphors or similes, exclamations, parallelisms, and recurrences also make the speech more affective.

4.7. Topics

Topic is an open-ended category that can be described at many different levels. It is possible to distinguish among very general topical domains, such as science, religion, politics, and sports, but any text will have its own specific topics. In this paper, debate is generally about confrontational argumentations between two sides on a proposition and emphasizes thorough analyses and demonstrations on the argument, which in linguistic features are large amounts of nouns, terms, HF words, conjunctions and logic words, small amounts of rare words, along with low normalized TTRs. The situation is different in speech. Despite the same theme – military as debate, speech is mostly about storytelling, with its contents focusing on touching stories about causes and lives. And because stories differ from speaker to speaker, speech covers content of wide ranges, which in linguistic features are small amounts of HF words, large amounts of rare words and types, high normalized TTRs, and dispersed information.

5. Conclusion

This paper attempts to adopt Biber's multi-dimension/multi-feature approach for a comparative study of Chinese debate and speech. In terms of linguistic features, instead of indiscriminately taking Biber's in his register studies, we identified 65 linguistic features for investigation on the basis of specific situational characteristics and communicative purposes of Chinese debate and speech, along with reality perception of differences between the two. It is in accordance with what

Biber proposed as "the situational characteristics are more basic" and "registers are determined by their situational characteristics". What's more, speed in phonetic level, some figures of speech in rhetoric level plus sentence types are also included here as features. It is an extension of Biber's multi-features analyses, in which only vocabularies, characteristics of words, grammatical features and syntactic structures were investigated. In terms of practice methods for multi-feature investigation, except for some that are not testable, most linguistic features were tested with comparison of means and hypothesis testing, which basically implemented the objective identification of significant differences. However, in terms of multi-dimension analysis, the volume of the text corpus was insufficient for factor analysis, which caused failure of the objective recognition of identifying co-occurrences of linguistic features. Namely, this study didn't fully achieved what Biber advocated as quantified dimensional analysis. This is a tentative application of Biber's multi-dimension/multi-feature approach; it gives a deeper understanding of the application conditions and the statistical techniques this approach requires, and lays the foundation for further applying it into overall comparative studies of Chinese registers.

Notes

References

[1] Ervin-Tripp, Susan. On sociolinguistic rules: alternation and co-occurrence. In John J. Gumperz and Dell Hymes (eds.) Directions in Sociolinguistics, New York: Holt, Rinehart and Winston 1972, pp. 213-250.

[2] Hymes, Dell. Foundations in Sociolinguistics: An Ethnographic Approach. Philadelphia: University of Pennsylvania Press 1974.

[3] Biber, Douglas. A model of textual relations within the written and spoken modes. Unpublished Ph.D. dissertation, University of Southern California 1984.

[4] Biber, Douglas. Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. Linguistic. Vol. 23, No. 2, 1985, pp. 337-360.

[5] Biber, Douglas. Spoken and written textual dimensions in English: Resolving the contradictory findings. Language. Vol. 62, No. 2, 1986, pp. 384-414.

[6] Biber, Douglas. Variation across Speech and Writing. Cambridge: Cambridge University Press 1988.

[7] Biber, Douglas. On the complexity of discourse complexity: A multidimensional analysis. Discourse Processes. Vol. 15, No. 2, 1992, pp. 133-163.

[8] Biber, Douglas. Dimensions of Register Variation: A Cross-linguistic Comparison. Cambridge: Cambridge University Press 1995.

[9] Biber, Douglas. Corpus-based analyses of discourse: Dimensions of variation in conversation. In Vijay Bhatia, John Flowerdew & Rodney H. Jones (eds.) Advances in discourse studies, London: Routledge Taylor & Francis Group, 2008 pp.100-114.

[10] Biber, Douglas, Susan Conrad, Randi Reppen, Pat Byrd & Marie Helt. 2002 Speaking and writing in the university: A multi-dimensional comparison. TESOL Quarterly. Vol. 36, No.1, 2002, pp. 9-48.

[11] Biber, Douglas, Susan Conrad. Registers, Genre, and Styles. Cambridge: Cambridge University Press 2007.

[12] Lei, Xiuyun, YangHuizhong. Corpus-based Approaches and the MD/MF Model in Stylistic Study of English for Academic Purposed. Contemporary Linguistics. Vol. 3, No.2, 2001, pp. 143-151.

[13] Wu, Jiangsheng. A Review of Multi-Dimensional Approach to Register Variation. Journal of PLA University of Foreign Languages. Vol. 24, No.3, 2001, pp. 6-9.

[14] Wu, Jiangsheng. A academic e-mails stylistic features analysis based on Multi-dimensional approach. Foreign Languages and Their Teaching. Vol. 179, No.2, 2004, pp. 53-57.

[15] Ma, Guanghui. Contrastive analysis of linguistic features between EFL and ENL essays. Foreign Language Teaching and Research. Vol. 34, No.5, 2002, pp. 345-349.

[16] Gui, Shichun. A corpus-based analysis of the register of English linguistics. Foreign Language Teaching and Research Press 2009.

[17] Wen, Qiufang. A Study on Changes of Register Features in Learner English. Journal of Foreign Languages. Vol. 32, No.4, 2009, pp. 2-10.

[18] Hu, Xianyao. A corpus-based multi-dimensional analysis of the stylistic features of translated Chinese. Foreign Language Teaching and Research. Vol. 42, No.6, 2010, pp. 451-458.

[19] Zhang, Wujiang. Average speed of TV commercials and its understanding of audiences. Modern Communication. Vol. 182, No.9, 2011, pp.148-149.

[20] Hou, Min. Introduction of Some Terms Concerned with Language Resource Construction and Language Monitoring. Terminology Standardization & Information Technology. Vol. 58, No.2, 2010, pp.30-33.

[21] Liu, Yanchun. A Comparative Study of Broadcasting Programs Stylistics based on Dependency Treebank. Language Teaching and Linguistic Studies. Vol. 159, No.1, 2013, pp.89-95.

[22] Liu, Yanchun. A treebank-based comparative quantitative study: with CCTV programs of Lecture Room and One On One as cases. GUANGXI SHEHUI KEXUE. Vol. 210, No.12, 2012, pp.142-145.

[23] Liu, Yanchun, Hu Fengguo. A Comparative Study of Stylistics between "Reading News" and "Talking News". Language Teaching and Linguistic Studies.Vol. 147, No.1, 2011, pp.97-104.