

Research on the Characteristics of Academic Chinese Vocabulary from the Perspective of International Chinese Education

Qiqing Huang

(Yunnan Chinese Language and Culture College, Yunnan Normal University, Kunming, China 650500)

147627946@qq.com

Keywords: Linguistics and Applied Linguistics; Academic Chinese Vocabulary; Features; International Chinese Teaching

Abstract. Based on the corpus of linguistics and applied linguistics, this paper compares the common vocabulary of modern Chinese (draft), and uses the two methods of word frequency statistics and word frequency order difference for a preliminary analysis of the composition and characteristics of academic Chinese vocabulary in linguistics and applied linguistics. This paper also puts forward some suggestions on how to teach academic Chinese vocabulary in the international Chinese education community.

In recent years, there are more and more undergraduates, postgraduates and even doctoral students studying in China. Basic oral communication and simple written writing can no longer meet the learning and research needs of this group. How to train and improve the professional learning and research ability of this group, especially the academic writing ability is undoubtedly an urgent problem to be solved in the current international Chinese education circle. This is because "the expression of academic Chinese and the enhancement of innovation have profoundly affected the future destiny of China's academic and cultural soft power, and determined the right of 'cultural China' to speak in the international strategic landscape." (Pan Chung, 2018) The international Chinese education community has conducted a lot of research on Chinese and professional Chinese in industry, but it does not pay enough attention to academic Chinese studies for professional study and research. This paper makes a tentative exploration on the characteristics of academic Chinese vocabulary through the use of lexicology and applied linguistic corpus, through word frequency statistics and frequency order difference analysis.

Existing Academic Chinese Studies and Corpus Construction

English as a foreign language teaching is divided into English for General Purposes and English for Specific Purposes. English for special purposes is divided into English for occupational purposes and English for academic purpose. English for academic purpose is often translated as "Academic English". In the international Chinese education community, Hong Zhang(2004) first mentioned the term "Academic Chinese" when talking about the mode of teaching Chinese as a foreign language, and regarded "Academic Chinese" as "a major course for learners of further learning". Later, Chengqing Yang and Jinjun Zhang(2006) talked about the examination of academic Chinese. Zhongwei Wu(2016, 2017) talked about the academic Chinese teaching model and academic Chinese ability. Liu Hong, Xinyi Yang (2017), and Chuanbo Yan (2016) Arrived in academic Chinese thesis writing. Chuanbo Ji(2016), Zengxia Gao, Fuying Liu(2016), and Yong Pan(2018) demonstrated the importance of academic Chinese. Most of the above studies are occasionally talked about and applied, but what is academic Chinese? What are the characteristics of academic Chinese elements? The academic community still has insufficient discussion.

Regarding the nature of academic Chinese, Zengxia Gao, Fuying Liu(2016) and Quan Zhou(2016) all believe that: "The purpose of academic Chinese is to serve professional learning. Chinese for general academic purpose refers to Chinese teaching in all subjects, while Chinese for special academic purpose refers to Chinese teaching in various professional fields, which is what we usually call 'professional Chinese' ." These two studies regard academic Chinese as a Chinese teaching behavior with specific teaching content. Chuanbo Ji(2016) defines academic Chinese as

“mainly referring to Chinese used in professional studies and academic research”. We believe that this definition is convincing because: First, the author understands academic Chinese as a language system for a specific field, rather than a teaching behavior; Second, the author emphasizes the use of the field in the definition of academic Chinese, not the purpose of teaching. We believe that academic Chinese is a social variant of the modern Han Chinese common language in the field of professional learning and research. It is a social dialect with its own characteristics in grammar, vocabulary, pronunciation and rhetoric.

This paper only discusses the characteristics of academic Chinese vocabulary in the field of domestic linguistics and applied linguistics. Corpus is an important basis for lexical measurement research. (Xinchun Su,2010) In order to investigate the composition characteristics of the academic Chinese vocabulary, we selected the four core journals of “Chinese Teaching in the World”, “Language Application”, “Language Teaching and Research” and “Chinese Learning” for nearly three years according to the general principles of constructing corpus. The high-cited papers have established a total of 2.5 million words of academic paper corpus. After text conversion, machine word segmentation, and manual proofreading, 32696 words were extracted.

Descriptive Statistics and Analysis of Chinese Vocabulary in the Field of Linguistics and Applied Linguistics

High-frequency Words in Academic Chinese Vocabulary are Highly Concentrated, Low-frequency Words are Discretely Distributed. According to the statistics of word frequency, 32696 words extracted have the characteristics of low number of high frequency words, high coverage of corpus, high number of low frequency words and low coverage of corpus. It can also be said that the high frequency words are highly concentrated and the low frequency words are discretely distributed. This statistical result is consistent with the statistics of Chinese vocabulary in other fields, such as Su Xinchun, Yang Erhong (2006), Wang Tiejun, Hou Min, Yang Erhong (2007), Ma Xinxin, and Chen Xiaohe (2017). In order to show the distribution characteristics, we sorted 32696 words according to the word frequency and carry out descriptive statistics on the corpus coverage of the first 30,000 words with 1,000 words as the statistical interval. The results are as follows:

Table 1 Statistics of the Distribution of Vocabulary Used in Linguistics and Applied Linguistics

Order on 1,000 words	Coverage of 1,000 words	Total coverage	Order on 1,000 words	Coverage of 1,000 words	Total coverage	Order on 1,000 words	Coverage of 1,000 words	Total coverage
1	74.58	74.58	11	0.37	97.21	21	0.12	99.15
2	9.60	84.18	12	0.31	97.52	22	0.07	99.22
3	4.56	88.74	13	0.29	97.82	23	0.07	99.29
4	2.62	91.36	14	0.22	98.04	24	0.07	99.37
5	1.70	93.07	15	0.22	98.25	25	0.07	99.44
6	1.19	94.26	16	0.19	98.44	26	0.07	99.51
7	0.89	95.14	17	0.15	98.59	27	0.07	99.58
8	0.69	95.84	18	0.15	98.74	28	0.07	99.66
9	0.55	96.39	19	0.15	98.88	29	0.07	99.73
10	0.45	96.84	20	0.15	99.03	30	0.07	99.80

As can be seen from the above table data, the coverage of the most commonly used words is as high as 74.22%. The coverage of the second thousand words dropped sharply, accounting for 9.61%, and the subsequent coverage of 28,000 words was lower, and the cumulative coverage only

accounted for 15.62%. In order to further reflect the high concentration of high-frequency words in the essay of the professional paper, we used Su Xinchun et al. (2005) to compare the statistical methods of news vocabulary and compare the two. The results are as follows:

Table 2 Comparison of Vocabulary Distribution with News Corpus

Total coverage	News corpus words	No. of words in the statistics of this paper
<50%	657	218
<80%	5403	1451
<90%	15136	3421
<95%	34819	6820
<99%	410903	19803

The data in the above table shows that the number of words with a coverage rate of less than 50% in the news corpus is 657, while the number of words with the same coverage in this paper is only 218. The number of words with a coverage rate of less than 80% is 1451, which is much less than the 5403 words in the news corpus. The other comparison intervals are also true. The number of words with the same coverage is quite different. The first reason is related to the difference in the size of the two corpora. The second reason should be related to the lexical features of the two genres, that is, the high-frequency vocabulary in the vocabulary is more concentrated than the news genre. The statistics in this paper are consistent with Coxhead's (2000) statistics on English academic vocabulary (AWL). When the text coverage is 90%, the number of words in this paper is 3421. The statistics of Coxhead are 2570 English words (word Families).

Academic Chinese Vocabulary with Distinctive Features Can be Easily Used in Writing.

Does the high concentration of high-frequency words in academic Chinese mean that the high-frequency vocabulary in the corpus is the representative of academic Chinese vocabulary? Without careful and in-depth comparative analysis, it is obviously impossible to answer the question scientifically. To this end, we conducted word class statistics on the 32,696 words extracted. It is found that there are a large number of domestic and foreign scholars (14.31%), urban local names (2.01%), abbreviations (1.15%) and institutional names (0.14%) in linguistics and applied linguistics. The statistics of other vocabulary categories are as follows:

Table 3 Statistics of the Distribution of Vocabulary Used in Linguistics and Applied Linguistics

Word Category	Sign	No.	Rate	Examples of the most common words (the subscript numbers are word frequency)
Noun	n	9358	28.62	结构 ₅₄₄₈ 、动词 ₅₂₇₅ 、语言 ₅₁₆₇ 、人 ₄₈₅₀ 、语义 ₄₄₉₁
Verb	v	7692	23.53	是 ₂₄₂₃₉ 、有 ₉₄₅₄ 、可以 ₅₁₃₇ 、为 ₄₁₁₉ 、能 ₃₈₈₄
Adjective	a	1930	5.9	不同 ₃₁₅₄ 、大 ₂₀₈₉ 、小 ₁₉₄₇ 、高 ₁₆₆₇ 、多 ₁₅₇₂
Gerund	vn	1697	5.19	研究 ₂₇₇₈ 、教学 ₁₈₆₈ 、影响 ₁₁₅₁ 、分析 ₁₀₉₈ 、认知 ₁₀₉₇
Nominal morpheme □	ng	1260	3.85	语 ₃₅₃₇ 、式 ₂₈₂₀ 、义 ₂₇₃₃ 、体 ₁₆₅₅ 、性 ₁₅₂₃
Adverb	d	1055	3.23	不 ₁₁₂₄₃ 、也 ₆₇₈₁ 、都 ₅₁₂₅ 、就 ₄₄₉₅ 、还 ₃₀₇₈

① The reason why this article refers to these words as “morpheme words” is mainly because these words are mostly ancient monosyllabic words or acronyms of modern Chinese. These words generally exist in the form of morphemes in modern Chinese spoken language. In the written language, the degree of freedom is large, and some can be independent. For example, the “body” and “sex” of grammatical meanings indicate the “present” and “present” of time, and some are used only in some structures, such as “description” and “show” for the “~” structure. Some are often used as affixes or quasi-affixes, such as “non”, “style” and “sex”. The reason why these analyses are taken out is mainly because these ingredients are often used in written language and reflect a strong written language color.

Attribute word	b	448	1.37	一定 ₉₅₂ 、主要 ₅₆₆ 、上述 ₅₃₃ 、单 ₄₆₄ 、所有 ₄₃₅
Verb morpheme	vg	428	1.31	述 ₇₇₁ 、表 ₅₇₁ 、示 ₂₈₅ 、现 ₁₉₇ 、非 ₂₇₆
Numeral	m	421	1.29	一 ₁₇₂₂₅ 、两 ₄₀₉₅ 、三 ₂₆₉₃ 、二 ₂₂₀₂ 、第二 ₁₁₇₆
Time word	t	362	1.11	现代 ₇₇₀ 、现在 ₃₉₄ 、目前 ₃₂₄ 、过去 ₁₆₇ 、以往 ₁₆₄
Quantifier	q	313	0.96	个 ₁₂₁₈₀ 、种 ₅₉₀₈ 、句 ₄₃₂₆ 、类 ₁₈₅₄ 、次 ₁₀₄₅
Proper noun	nz	239	0.73	汉语 ₇₄₆₇ 、汉字 ₁₄₁₁ 、英语 ₁₀₁₆ 、中文 ₂₅₈ 、韩语 ₁₈₇
Adjective and adverb	ad	238	0.73	一般 ₇₂₇ 、完全 ₅₉₁ 、直接 ₅₄₇ 、多 ₃₇₂ 、容易 ₂₇₄
Adjective and noun	an	214	0.65	困难 ₁₃₅ 、矛盾 ₉₄ 、礼貌 ₇₃ 、不足 ₅₀ 、努力 ₄₆
Descriptive morpheme	ag	194	0.59	主 ₂₈₂ 、详 ₁₃₉ 、简 ₄₄ 、歧 ₃₇ 、优 ₂₀
Conjunction	c	169	0.52	和 ₁₄₂₄₃ 、而 ₅₃₉₂ 、与 ₃₉₂₅ 、或 ₃₅₆₅ 、但 ₃₀₉₈
Place word	s	168	0.51	其中 ₁₁₀₆ 、这里 ₄₂₉ 、国内 ₁₂₈ 、家里 ₁₂₀ 、海外 ₁₁₉
Position word	f	164	0.5	中 ₁₃₁₃₈ 、上 ₅₉₀₄ 、时 ₃₅₂₃ 、后 ₂₄₉₁ 、之间 ₁₇₉₉
Demonstrative	rn	128	0.39	这 ₁₀₃₁₇ 、这些 ₁₄₁₆ 、那 ₁₃₀₉ 、该 ₁₁₉₅ 、本文 ₁₁₁₆
Preposition	p	91	0.28	在 ₁₅₇₇₈ 、对 ₇₅₃₄ 、从 ₃₇₄₈ 、与 ₃₆₄₁ 、为 ₂₈₁₃
Onomatopoeia	o	54	0.17	啊 ₁₄ 、哇 ₈ 、嘿 ₆ 、吧嗒 ₆ 、丁当 ₆
Modal	y	51	0.16	了 ₄₀₀₄ 、呢 ₈₅₄ 、吗 ₄₂₄ 、吧 ₂₉₇ 、啊 ₂₇₃
Interrogative pronouns	ry	51	0.16	什么 ₁₄₃₅ 、谁 ₂₈₈ 、为什么 ₂₅₈ 、多少 ₂₀₇ 、哪 ₂₀₀
Personal pronouns	rr	49	0.15	我 ₄₁₄₆ 、我们 ₄₀₂₃ 、其 ₃₉₂₁ 、他 ₂₈₃₁ 、你 ₂₂₀₆
Temporal morpheme	tg	40	0.12	现 ₁₃₃ 、今 ₈₉ 、清 ₅₀ 、汉 ₄₉ 、晚 ₃₂
Particle	u	36	0.11	的 ₉₃₅₇₆ 、了 ₉₀₄₄ 、等 ₃₈₅₀ 、所 ₃₂₆₉ 、着 ₁₇₄₂
Interjection	e	31	0.09	啊 ₇₂ 、嗯 ₄₆ 、哎 ₂₂ 、哦 ₂₂ 、呀 ₁₄
Adverbial morpheme	dg	29	0.09	复 ₈₉ 、甚 ₆₃ 、频 ₂₀ 、俱 ₉
Affix	k	27	0.08	者 ₂₃₅₅ 、式 ₁₈₂₃ 、化 ₁₈₀₀ 、非 ₁₁₇₆ 、型 ₁₀₄₆

From the perspective of word distribution, the vocabulary used in the corpus has a wider distribution of parts of speech, which is no different from other corpora. However, from the specific word class and its most common words, it shows a strong academic and strong written expression.

The most common words in most word classes have strong professional characteristics. Some of the most commonly used words in the above-mentioned word classes are professional terms or research terms related to professional research. In the five most commonly used nouns, “结构”“动词”“语言”“人”“语义”, all four words except “人” are linguistic terms. “结构” mainly refers to the structure of grammar, semantics, vocabulary, speech or Chinese characters in a language. The five most commonly used verbs such as “是”“有”“可以”“为”“能”, and other multi-table judgments, have the ability to function, have a distinct qualitative color. The five most commonly used adjectives “不同”“大”“小”“高”“多” are used for comparison, with distinct comparative and quantitative research colors. Among the five most commonly used words and phrases, “研究” and “教学” reflect the main research content of linguistics and applied linguistics. “影响”“分析” and “认知” reflect the process and perspective of research in this area. Among the five most commonly

used sub-categories, “一般” “完全” “直接”, “多” and “容易” are also used for judgment and reasoning of language laws or phenomena.

A large number of monosyllabic words or morphemes have a strong written style. In order to reflect the simplicity and elegance of the written language, the written genre often uses some monosyllabic words that have been reduced or reduced by two-syllable words, especially in the rigorous and standardized academic written language. These words generally have corresponding modern Chinese two-syllable words, such as “已” and “已经”, “述” and “描述”, “频” and “多次”. The monosyllabic words with such characteristics in the above table succinctly reflect the strong written color of the terminology. Verb morpheme words such as “述” “表” “示” “现” “非”, and adjectives such as “主” “详” “简” “歧” “优”, adverbial morphemes such as “复” “甚” “频” “俱” “顿” and so on. These morphemes are commonly used in expressions such as “如上文所述” “详见下表” “简言之” “最优模式” “复现” and “不甚了解”. The most commonly used noun morpheme words such as “语” “式” “义” “体” “性” are mostly used in professional terms, such as “语际干扰” “后附式话题标记” “常用义” “持续体” “远指性有标话题” and so on.

The other words listed in the table also more or less reflect the color of the term body, such as the demonstrative pronoun “本文”, the preposition “对”, the modal particle “呢” and so on. The preposition “对” is often used in the format of “对.....进行了研究/探讨/讨论/澄清”. The modal particle “呢” is often used in academic questions that emphasize or draw attention. For example, “三种语法形式的语体差异是否可以在时空度上找到印证呢?”. The use of other words will not be listed.

In addition, it should be noted that the frequency of use of the words listed in the table is high or low, and the frequency of use is obviously required to be taken seriously. The use frequency is relatively low, and sometimes it can better reflect the vocabulary use characteristics of the terminology. It should be paid more attention to in the study of terminology.

Characteristics of Academic Chinese Vocabulary in Comparison with “Modern Chinese Vocabulary (Draft)”

Professional Research on 1450 Words not Found in the Vocabulary. As can be seen from Table 1 above, the cumulative coverage to the 10th thousand words, the first 10,000 words, can reach 96.84%. The highest frequency of these 10,000 words is 93,536 times and the lowest frequency is 6 times. That is to say, out of the 32,696 words extracted, 62.42% of the words were used at a frequency of 6 or less. After investigation, it was found that these words with lower frequency of use were mostly used in the example sentences. Therefore, the most representative of the characteristics of academic Chinese vocabulary should be the first 10,000 words or even the first 6,000 words, which is consistent with “The 80/20 Rule” of human language, that is, the top 20% of high-frequency words can reflect 80% of the information. the amount.

Studying the characteristics of academic Chinese vocabulary in comparison with general genre or other genre is undoubtedly a more direct and scientific research method. In this paper, the first 10,000 words are taken out according to the frequency order, and these words are compared with the vocabulary in the “Modern Chinese Common Vocabulary (Draft)” (hereinafter referred to as “Word List”) which can represent the general Chinese vocabulary. Since the entries in the “Modern Chinese Common Vocabulary (Draft)” are arranged in ascending order of frequency, and the same frequency is arranged in the order of Chinese Pinyin, there is no specific word frequency and part of speech, and the frequency cannot be subdivided. To this end, we use word types as statistical analysis objects, and merge two or more homographs of 10,000 words or meaning types into one word type, and the word frequency is the sum of word frequencies of different types of words. A total of 8389 word forms were obtained. The 56008 words in the Vocabulary are merged into 55,581 word forms, and the word order of the word form is the average value of the word order. A total of 8389 word forms were combined with 55,681 word types. The results showed that there were 6,939 identical words in the two tables, and 1,450 words were not found in the vocabulary,

accounting for 17.28%. Such as the most commonly used top 50 word types: “与”“语素”“用例”“看到”“吕叔湘”“不同”“域”“来说”“看出”“量级”“学习者”“们”“谓词”“区别词”“提到”“第二”“分为”“只能”“正确率”“韩语”“语用”“语言学”“两者”“而是”“主谓”“句式”“补语”“一点”“做出”“来看”“宾”“很多”“施事”“宾格”“疑问句”“述”“复合词”“带来”“复数”“条件句”“第”“第三”“示”“词素”“放在”“提出”“某种”“可能性”“研究者”“中国式”, etc.

There are many reasons for not seeing the vocabulary. First, the vocabulary has received mistakes, such as the common conjunction “与” lost. □ In addition, the main basis of the vocabulary collection is the word frequency, which does not consider the systematic nature of the vocabulary, such as the “语音”, “语法” and “语义”, but there is no the “语用”. Second, the words not included in the vocabulary are mostly terminology. In addition to “语用”, there are also “句式”“语素”“语言学” and “补语”. There are also some professional researchers' names. The frequency of use of these words in general corpus is clearly not too high. The third is related to the machine word segmentation method. The word segmentation software used in this paper divides some components between morphemes and words, words and phrases into independent language components. The former is prefixed with “第”, suffix “们”, professional suffix “域”, “述” “示” for “所 V” structure, “宾” for parallel abbreviation, etc., the latter is “不同” “学习者” “第二” and so on. In general terms, the 1450 words that are not found in the vocabulary are mostly terminology, researcher names, and used for quantitative statistics (such as “自变量”“因变量”“相关性”“大部分”“绝大多数”“绝大部分”“大规模”, etc.), qualitative analysis (such as “不足以”“不可避免”“毫无疑问”“必不可少”“并未”“有所不同”“截然不同”, etc.), the table refers to (such as“此类”“该类”“本次”“两者”, etc.), the number of tables (such as “第四”“其一”), the insertion (“事实上”“一般来说/说来”“相比之下”“相对而言”“整体而言”). Obviously, these types of words are mostly related to professional research, and have a clear color of the term.

A Study of the Difference of 6939 Word Forms Shared with the Vocabulary. American scholar Zipf (1932, 1936) found that in the real text of human language, the product of the frequency and frequency of words in any article is always a constant, and the two are inversely proportional. In other words, the distribution of words in real text is not a loose sand, but a regular one. Domestic scholars such as Ma Chuangxin, Chen Xiaohe (2017, 2018), Liu Rui (2017), Chen Haibo (2010) and others have proved this. The difference or grade difference of common words in different styles has become one of the main research methods of academic lexical distribution, especially high frequency feature words. The two tables have a total of 6939 word forms, accounting for 82.72%. Subtracting the frequency sequence of the same word type in the vocabulary from the frequency sequence of the vocabulary in this paper, the order difference of each word is obtained, and the order difference is arranged in ascending order, and the sequence number is sequentially numbered from 1 to the order difference. (r). Then each word type has a fixed coordinate (r, D), and 6939 words are placed in the coordinate system to obtain a scatter plot of all word type differences (Figure 1). The scatter plot shows that the order difference between the two texts has a distinct regularity, the overall trend is smooth, the upper tail is shorter, and the lower tail is longer. This shows that many words in the body of the term are more commonly used than the general style. The lower tail section is obviously the part of this paper, so which words should be used to focus on? In order to further determine the focus on the word-type object, we used SPSS to calculate the box plot of the sequence difference (Figure 2).

② In addition to the conjunction “and” in the table, the “vocabulary” also does not include the words "strong", "director", "microblogging", "applause", "red wine", "weak" and "precision" in the 1450 word types.

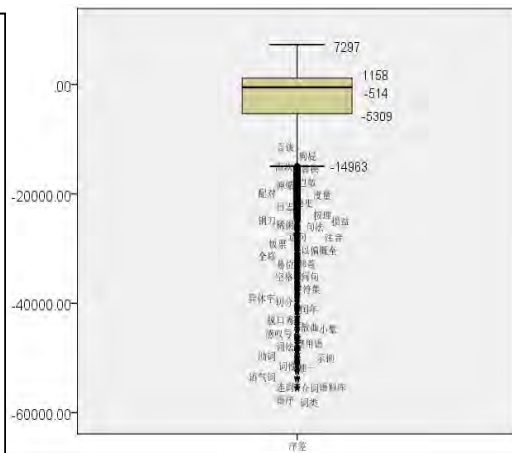
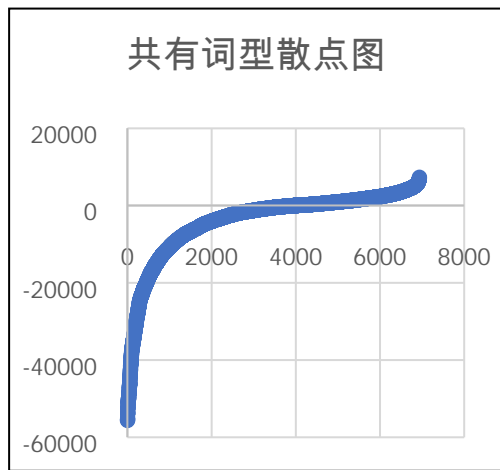


Figure 1. Scatter Chart of Common Words

Figure 2. Differential Box Plot

According to the research conclusion of Liu Rui (2017), the word type between 1158 and 7292 is a common genre of common genre, corresponding to the upper part; The word type between -5309 and 1158 is a common vocabulary of two styles, corresponding to the middle word; The 1040 word types with a sequence difference between -14963 and -5309 are commonly used words in the terminology, corresponding to the lower tail segment; the 695 word types less than -14963 are abnormal words and extreme words, which are more commonly used words. Due to space limitations, this article focuses on 695 anomalies and extreme word types. The 1040 words in the lower end will be left for discussion.

There are many professional terms in linguistics and applied linguistics. The analysis found that nearly 5% of the 695 abnormal and extreme word types were linguistic terms or fixed-expression terms related to the profession. For example, the 30 most commonly used words are “动词”“语义”“句法”“宾语”“母语”“语境”“副词”“习得”“短语”“形容词”“量词”“字句”“语料”“主语”“音节”“介词”“情态”“语料库”“代词”“谓语”“主观性”“教学法”“韵律”“词义”“受事”“词类”“虚词”“状语”“语序”“义项”and so on.

More Quantitative Research Vocabulary. In addition to professional terminology, 90% of the 90 word types are used to express quantitative research methods of academic papers, with a strong empirical research color. This is in line with the development trend of linguistics and applied linguistics, that is, more and more empirical research based on quantitative. The 30 most commonly used words are as follows: “评分”“序数”“替换”“交互”“均值”“切分”“百分比”“标注”“推导”“实证”“互换”“线性”“选项”“度量”“级差”“异同”“小量”“归类”“类同”“合计”“排序”“遗漏”“导出”“介于”“无误”“个数”“频度”“换位”“低频”“交集”and so on.

A Certain Proportion of Vocabulary for Qualitative Analysis. Qualitative is one of the fundamental purposes of academic research. In statistics, we find that a certain proportion of word forms are often used in the conclusion part of academic papers, that is, the commonly used word types when defining the essential attributes of things, such as “显性”“隐性”“特指”“突显”“唯一”“申明”“存疑”“推论”“彰显”“明示”“泛指”“契合”“鲜有”“截然”“贴切”“相悖”“己见”“略为”“悖论”, etc.

Word Forms that Can Reflect the Structure and Introduction of Academic Papers. The structure of modern academic papers mainly includes the macro structure and microstructure of the paper. The former mainly refers to the structure of the journal requirements and the structure of the content of the paper, while the latter mainly refers to the logical expression structure of the language of the paper. The word forms that embody these structures also appear in the outliers and extreme values counted in this paper. The word types that reflect the structure of the journal requirements are “关键词”“引言”“结语”“提要”“附录”“小结”“索引”and so on. The words that embody the logical expression structure of academic language are mainly interjections with cohesive function, general

function and confidence function, such as “综上所述”“如前所述”“毋庸置疑”“有鉴于此”“换言之”“说白了”“简言之”“同理” and so on. Introducing the academic views of others is one of the commonly used arguments for academic papers. Some abnormal types or extreme values also appear in the form of quotations, such as “参见”“引述”“转引”“引语”, etc.

In addition to the above four types of words related to professional research, there are about 22% of the 695 abnormal and extreme types of words used for examples or research objects. These word types are not used frequently, professionalism is not strong, and the written colors are not obvious, such as “综艺”“惨败”“完胜”“口诀”“乘法”“口香糖”“性行为”“再婚” and so on.

Conclusion and Teaching Suggestions

This paper takes linguistic and applied linguistic corpus as the material for academic Chinese vocabulary research. The word frequency statistics show that the academic Chinese vocabulary is the same as the Chinese vocabulary in other fields, showing the high concentration of high frequency words and the overall characteristics of low frequency words. According to further statistics, in addition to the names, organization names and place names of a large number of scholars, the high-frequency words in most word classes have distinct professional colors and strong written colors. Comparing the first 10,000 words selected according to the word frequency with the "Modern Chinese Common Vocabulary (Draft)" and the difference analysis, the academic Chinese vocabulary mainly includes professional terms, written vocabulary, research vocabulary and structural vocabulary. The professional terminology is a professional vocabulary often discussed in the academic world, and the latter three are general academic Chinese vocabulary.

It can be seen from the above conclusions that academic Chinese vocabulary has both common word frequency statistics and distinct stylistic features compared with other Chinese vocabulary. This inspires us that academic Chinese studies can not only enrich the theoretical study of Chinese or Chinese language, but also improve the practical level of international Chinese education. For Chinese learners who are mainly engaged in professional study and research, it is necessary to set up a special academic Chinese course in the advanced stage, preparatory stage or professional study stage. When setting up an academic Chinese course, if the interlacing is like a mountain, the professional terminology should be taught by a professional teacher, and the general academic Chinese vocabulary, structure or rhetorical means should be studied by a Chinese teacher with rich experience through specialized Chinese textbooks. General academic Chinese can be studied through specialized courses or in advanced Chinese or essay writing courses.

References

- [1] Coxhead, A. "A new academic word list". *TESOL Quarterly*,2000(34).
- [2] Paul Nation, A study of the most frequent word families in the British National Corpus. P. Bogaards & B. Laufer. *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Amsterdam: John Benjamins Publishing Company,2004.
- [3] Zipf, G. K. *The Psycho-biology of language: An Intro-duction to dynamic philology*. London: George Routledge & Sons Ltd.,1936
- [4] Zipf, G. K. *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge: Harvard University Press,1932.
- [5] Haibo Chen. *Application of Order Difference in the Study of Text Distinguishing Features*. Yangtze River Academic,2010,(4).
- [6] Yunming Shan. *The Compilation of Special Purpose Chinese Textbooks*. *Journal of College of Chinese Language and Culture of Jinan University*,2008,(2).
- [7] Zengxia Gao, Fuying Liu. *On the Importance of Academic Chinese in Teaching Chinese as a Foreign Language*. *Journal of Yunnan Normal University(Teaching and Research on Chinese as A Foreign Language)*,2016,(2).
- [8] Chuanbo Ji. *A Preliminary Study on the Use of Chinese Language Structures in Korean*

- Students' Chinese Academic Papers. *Chinese Language Learning*,2016,(2).
- [9] Hong Liu, Xinyi Yang. An Investigation of the Introduction Structure of Undergraduate Teaching Research in Chinese International Education Based on the Step-by-Step Method. *Journal of Yunnan Normal University(Teaching and Research on Chinese as A Foreign Language)*,2017,(2).
- [10]Rui Liu, Bize Sun, Yunfei Long, Shan Wang. The Distribution Characteristics of Words and Their Differences in Vocabulary Between Words. *Journal of Chinese Information Processing*,2017,(5).
- [11]Chuanxin Ma, Xiaohe Chen. The Lexical Division Law and the Discovery of High-Frequency Feature Words in the Literature . *Applied Linguistics*, 2018, (3).
- [12] Chuangxin Ma, Xiaohe Chen. The Distribution of Words, Lexical Grades and Styles in the Literature. *Journal of Chinese Information Processing*,2017,(4).
- [13]Yong Pan. Positive Pragmatics: An Important Way to Improve the Soft Power of Academic Culture in China. *Journal of Capital Normal University(Social Sciences Edition)*, 2018, (1).
- [14] Xinchun Su, Erhong Yang. Analysis and Reflection on Chinese Vocabulary Statistics in 2005. *Journal of Xiamen University(Arts & Social Sciences)*,2006,(6).
- [15]Xinchun Su. *Vocabulary Measurement and Implementation*. Beijing: The Commercial Press, 2010.
- [16]Tiekun Wang, Min Hou, Erhong Yang. Investigation on the Use of Words in Newspapers, Radio and Television, and the Internet. *Applied Linguistics*, 2007, (1).
- [17]Zhongwei Wu, Zhi Geng, Tingting Xu. The Development Trend of Chinese Textbook Construction and Related Theoretical Issues. *I International Chinese Language Education*,2017,(1).
- [18]Zhongwei Wu. The Integration, Innovation and Optimization of Chinese Teaching Model. *TCSOL Studies*,2016,(1).
- [19] “Modern Chinese Common Vocabulary” Group. *Common Chinese Vocabulary (Draft) [M]*. Beijing: The Commercial Press, 2008.
- [20]Hong Zhang. Using Service to Improve the Competitiveness of Teaching Chinese as a Foreign Language. *Liaoning Education Research*,2004,(8).
- [21]Quan Zhou, *Academic Chinese: New Development of Chinese International Education in the Background of “One Belt, One Road”*. *Chinese Vocational and Technical Education*, 2017, (26).