A corpus-based study on Chinese modification patterns of nouns across registers

Dan Zhang[1, 2], Minglu Xu[1], Yunhua Qu[1*]

[1] School of International Studies, Zhejiang University, Hangzhou, China.
[2] School of Foreign Languages, Yantai University, Yantai, China.
[*] Corresponding author's email: qu163hua@163.com

DOI: https://doi.org/10.53482/2021_51_392

ABSTRACT

Nominal modification works to describe and restrict noun phrases, making the information delivery more vivid and precise. In English, the communicative functions of different modification patterns of head-nouns have been studied in a lot of corpus-based investigations of the written and the spoken registers, but few corpus-based register studies have been ever conducted in Chinese. This research takes the initiative attempt to conduct a corpus-based study on Chinese modification patterns across registers. A one-million-word corpus including both written and spoken Chinese is first built and all the modification patterns of noun phrases are extracted in Chunker, a self-developed colligation query and analysis tool. Through classification of modification patterns and statistical processing, the study displays the distributions of simple and complex modification patterns and the relationship between the frequency of modification patterns and the information density across registers and discusses the functional implication of such distributions and relationship under the guidance of Biber's register theory.

Keywords: corpus-based study, Chinese modification patterns of nouns, registers, communicative functions.

1 Introduction

Nominal modification serves to describe and restrict noun phrases, making the information delivery more vivid and precise. In Chinese, modification is located before the head noun in the noun phrase. Modification patterns of nouns are all the words and structures which are regularly associated with head nouns.

The studies on Chinese modification of nouns began in the late 19th century, but most of the theories and ideas mainly rely on traditional researchers' language intuition and introspective thinking. These studies range across several main research perspectives, such as the grammatical and structural studies (Zhang and Han, 1997; Huang and Liao, 2007), the comparative studies between English and Chinese

(Xiong, 1996), and the translation studies (Zhang and Zhao, 2011). The components in modification (Zhang and Han, 1997) were classified into prototypical and non-prototypical categories. The semantic relation between the modification and the head word (Huang and Liao, 2007) was generalized into restrictive and descriptive types. In Xiong (1996), the sequence and the structures of the multiple modifiers of nouns were compared with those in English, indicating that the sequence of the modifiers in English was more fixed while in Chinese that was much flexible. In Zhang and Zhao (2011), from the perspective of the English−Chinese translation, the modifiers for the head-nouns were analyzed through their arrangement in the Chinese used by translators.

With the wide use of electronic and authentic texts, more and more investigations on Chinese noun phrases have adopted a corpus-based approach. These studies based on corpus are often conducted from the perspectives of the semantic relations between the modification and the head nouns (Hu, 2003), the order of the attributives (Cheng, 2009), and the modification characteristics of the head-nouns in Chinese translation compared with those in English or original Chinese (Hu and Zeng, 2009). In the study by Hu (2003), based on an annotated corpus with about 6,000 noun phrases, the semantic relations between the components in the noun phrases were discussed and the operational processes to distinguish the semantic relations were designed, which provided a structural resource for Chinese information processing. In Cheng (2009), with the aid of corpus, the semantic types and the order of multiple attributives were investigated on the basis of the corpus of 100,000 words. In Hu and Zeng (2009), based on the comparable corpus, the study indicated that the unusual sequence and frequency of modifiers were essential features of translational Chinese and showed that the modifiers were a key factor in describing language.

Among the studies on Chinese noun phrases, modification patterns have been rarely investigated based on corpus. Compared with the traditional researchers' intuition and experience concerning language features, the corpus-based approach is more reliable to justify that one element is more frequent than another and to discover the unusual features that are less easily noticed than the ordinary ones (Tony and Andrew, 2012). The large amount of authentic texts stored in the electronic form enable the scholars to conduct the research more accurately by extracting the particular words, syntactic constructions, and collocations by various programmes. Moreover, the corpus approach is helpful to investigate the language patterns that tend to be unnoticed, which offers novel perspectives for linguistic studies. Therefore, in this paper, the corpus-based approach will be employed to examine the modification patterns of nouns in Chinese.

For noun phrases in English, Biber (1999) conducted a full investigation across four different registers (conversation, fiction, newspaper writing, and academic prose). Biber made detailed corpus-based research on the distributional features of the head nouns, the elements in the simple noun phrases and the pre-modification & the post-modification in the complex noun phrases. He pointed out that noun phrases are one of the essential linguistic features in register variation, including the semantic category of

nouns, the determiner or the article, the nominal pre-modifiers, the nominal post-modifiers, and the noun complement clauses.

Besides Biber, the linguistic differences in various registers have been investigated by a number of scholars, such as O'Donnell (1974), Olson (1977), and Chafe (1982). "A register is a variety associated with a particular situation of use" (Biber, 2009). The distinction between spoken and written registers is one of the most important situational parameters for the linguistic description on registers (Biber, 2009). Spoken registers are usually interactive and concerned about conveying speakers' own feelings and attitudes (Biber, 2009). Written registers allow time for planning and revising, and their major situational characteristic is a primary focus on communicating information (Biber, 2009). A register makes frequent use of a linguistic feature because that feature is well suited to the communicative purpose and situational context of the register. Similarly, linguistic co-occurrence patterns are functional: linguistic features occur together in texts because they serve related communicative functions (Biber, 2009). For a register in a specific situation or with some special purpose, its situational features can be investigated by its lexical and grammatical patterns. Words, collocations, and syntactic constructions can be examined to distinguish one register from another.

In the previous studies on Chinese modification patterns, researchers rarely give explanations from the perspective of communicative functions (Zhang and Han, 1997). The structural and distributive features among different registers are seldom compared, and most of the time, these analysis are limited to only one register − the written one (Xiong, 1996; Zhang and Zhao, 2011). In addition, compared with cross-register English studies based on corpus, most previous results of Chinese noun phrases lack accuracy and generality due to their inaccessibility to the comprehensive corpus data.

Therefore, it is worth studying how modification patterns of nouns are used in various registers and how communicative goals are achieved by linguistic structures. Since each modification pattern in different registers will display their corresponding distribution features with their respective communicative functions, it is possible to obtain an overview on how the distributions of the modification patterns contribute to the registers' communicative functions. The two main research questions in this study are:

1. What are the distributions of the major modification patterns of nouns in Chinese across registers?
2. How do these distributions of modification patterns serve to realize the communicative functions of different registers?

The paper conducts a corpus-based research on modification patterns of nouns in Chinese across the written and the spoken registers. It applies the corpus-based approach to find out the Chinese modification patterns by, first, building a one-million-word corpus with both written and spoken Chinese and then, extracting all the modification patterns of the noun phrases through the programme Chunker. After data processing and re-classification of Chinese modification patterns, the study not only analyzes the

quantitative data of the frequency of each pattern, the pattern distribution and complexity across registers, but also discusses their communicative functions in different registers under the guidance of Biber's theory.

2 Method

To investigate modification patterns of nouns in Chinese across registers, the main procedure consists of three phases. First, a comprehensive Chinese corpus is built incorporating both written and spoken registers. Second, all the noun phrases are extracted and tagged through Chunker, which is a self-developed colligation query and analysis tool for Chinese noun phrases and the modification patterns are sorted into categories according to the number of lexical modifiers before the head nouns. At last, the distributional results of the modifications patterns across registers are explained based on Biber's finding (2009).

2.1 Zhejiang University Corpus of Spoken and Written Mandarin Chinese

In Zhejiang University Corpus of Spoken and Written Mandarin Chinese, the sub-corpus of written Mandarin Chinese is built with reference to the Lancaster Corpus of Mandarin Chinese (LCMC), and the sub-corpus of spoken Chinese with reference to the Lancaster Los Angeles Spoken Chinese Corpus, in both of which the register classification, the distribution of the texts in each register, the size of the corpus (960,000 words in LOB and 1,000,000 in ZJUCSWMC), and other crucial criteria in the corpus are built according to the Lancaster-Oslo//Bergen Corpus (LOB). As LCMC is constructed with only written Mandarin Chinese texts published in Mainland China, the spoken Chinese texts have been included in this study in order to conduct comparative analyses.

The composition of the self-built corpus is displayed in Table 1. The written Chinese sub-corpus consists of 500,000 words, covering press, editorials, academic prose, official documents, magazines, and fiction; the other 500,000 words in spoken Chinese sub-corpus include TV drama, talk show, Internet speech, debate, and court trial. Unlike the written texts, the spoken texts have to be transcribed from the oral form to the written one. The texts of TV drama, debate, court trial, and Internet speech collected on line and talk shows downloaded are manually transcribed.

Table 1: Composition of Zhejiang University Corpus of Spoken and Written Mandarin Chinese.

Categories	Number of words	Percentages
News	100,000	10%
Academic papers	100,000	10%
Official documents	100,000	10%
Magazine	100,000	10%
Fiction	100,000	10%
Total for Written	**500,000**	**50%**
Natural conversation	54,000	5%
Beijing dialect	38,000	4%
Debate	82,000	8%
Court trial	82,000	8%
TV drama	82,000	8%
Talk show	82,000	8%
Internet speech	80,000	8%
Total for Spoken	**500,000**	**50%**
Total for Corpus	**1,000,000**	**100%**

2.2 Extraction of noun phrases

From the self-built corpus, all the noun phrases have been extracted by Chunker to create a noun phrases corpus for this study. Within the smaller noun phrases corpus, it is possible to identify the major modification patterns according to their frequencies and to figure out their distributions across registers. When noun phrases have been extracted, the elements in the noun phrases have been tagged at the same time. Table 2 shows the explanations for tags in Chunker.

Table 2: Tags explanation in Chunker.

Tags	Explanation
AD	Adverbs
AS	Aspect marker
BA	“把(ba) ” in ba-construction
CC	Coordinating conjunction
CD	Cardinal numbers
CS	Subordinating conjunction
DEC	“的(de) ” for relative-clause
DEG	Associative “的(de)”
DER	“得(de)” in V-de construction
DEV	“地(de)” before VP
DT	Determiner
ETC	Tag for words “等(deng)，等等(dengdeng)” in coordination phrase
FW	Foreign words
IJ	Interjection
JJ	Noun-modifier other than nouns
LB	“被(bei)” in long bei-construction
LC	Localizer
M	Measure word (including classifiers)
MSP	Some particles
NN	Common nouns
NR	Proper nouns
NT	Temporal nouns
OD	Ordinal numbers
ON	Onomatopoeia
P	Prepositions (excluding “把(ba)” and “被(bei)”)
PN	Pronouns
PU	Punctuations
SB	“被(bei)” in long bei-construction
SP	Sentence-final particle
VA	Predicative adjective
VC	Copula “是(shi)”
VE	“有(you)” as the main verb
VV	Other verbs

2.3 Statistical analysis

To describe the distributions of modification patterns, a couple of statistical means are exploited in the following figures and tables, including the frequency, the normalized frequency and the $\chi 2$ test.

$\chi 2$ test in SPSS has been used to test whether the differences in the distributions of modification patterns between the written and the spoken corpus are significant. In SPSS, “Sig.” is the *P* value. Generally

speaking, if the *P* value is less than 0.05, there is a significant difference; if the *P* value is greater than 0.05, there is no significant difference.

3 Modification patterns of nouns across registers

Through Chunker, 112,297 noun phrases are extracted from the self-built Zhejiang University Corpus of Spoken and Written Mandarin Chinese, and 1,226 types of modification patterns with the frequency of more than one are identified.

3.1 Classification of modification patterns

To identify the main modification patterns, the top 57 patterns with more than 200 occurrences in the whole corpus are chosen to be identified and examined; the other patterns with lower frequencies are not worth investigation, given their minor occurrences in each of the specific registers. Among the top 57 patterns, some are eliminated as they are not nominal patterns with modification, such as the coordination pattern NN CC NN. Finally, 37 patterns are selected for this study. Each frequently used pattern has been explained under the pattern and their frequencies in the corpus are showed in Table 3. For example, in the noun phrases "CD M NN NN" and "VA DEC NN NN", "CD M NN (cardinal number+measure word+noun)" is the modification pattern of the head noun "NN", and "VA DEC NN (adjective+"*de"*+noun)" is the modification pattern of the head noun "NN".

Table 3: 37 frequently used modification patterns of nouns.

	Patterns	Freq.		Patterns	Freq.
1	NN NN (noun+noun)	11,735	20	NR DEG NN (proper noun+ "*de"*+noun)	483
2	JJ NN (noun-modifier+noun)	4,800	21	CD M NN NN (cardinal number+measure word+noun+noun)	464
3	DT NN (determiner+noun)	4,168	22	OD NN (ordinal number+noun)	442
4	NN NN NN (noun+noun+noun)	2,819	23	DT NN NN (determiner+noun+noun)	432
5	CD NN (cardinal number+noun)	2,713	24	VV NN DEC NN (verb+noun+"*de"*+noun)	391
6	CD M NN (cardinal number+measure word+noun)	2,637	25	CD NN NN (cardinal number+noun+noun)	361
7	NN DEG NN (noun+ "*de"*+noun)	2,177	26	NN JJ NN (noun+noun-modifie+noun)	354
8	NR NN (proper noun+noun)	2,049	27	PN NN NN (pronoun+noun+noun)	337
9	PN DEG NN (pronoun+"*de"*+noun)	1,869	28	NN DEG NN NN (noun+ "*de"*+noun+noun)	332
10	JJ NN NN (noun-modifier+noun+noun)	978	29	DT CD M NN (determiner+cardinal number+measure word+noun)	320
11	PN NN (pronoun+noun)	923	30	JJ NN DEG NN (noun-modifier+noun+"*de"*+noun)	298
12	DT M NN (determiner+measure word+noun)	897	31	CD M JJ NN (cardinal number+measure word+noun-modifier+noun)	288
13	VA DEC NN (adjective+"*de"*+noun)	778	32	PN DEG NN NN (pronoun+ "*de"*+noun+noun)	276
14	NN NN DEG NN (noun+noun+ "*de"*+noun)	775	33	NT DEG NN *(temporal noun+"de"+noun)*	272
15	AD VA DEC NN (adverb+adjective+"de"+noun)	640	34	AD VV DEC NN *(adverb+verb+"de"+noun)*	246
16	NN NN NN NN (noun+noun+noun+noun)	624	35	NR JJ NN (proper noun+noun-modifier+noun)	246
17	NR NN NN (proper noun+noun+noun)	562	36	VA DEC NN NN (adjective+"*de"*+noun+noun)	227
18	VV DEC NN (verb+"*de"*+noun)	541	37	DT NN DEG NN (determiner+noun+ "*de"*+noun)	226
19	JJ DEG NN (noun-modifier+"*de"*+noun)	485			

The 37 modification patterns of nouns are classified into two types according to the number of lexical modifiers before the head nouns. Since "DEG" and "M" in the noun phrases are functional words, they are not taken into account as lexical modifiers. One type is a simple modification pattern of nouns with a single modifier (see Table 4). The other type is a complex modification pattern of nouns with two or more modifiers (see Table 5). Furthermore, the modification patterns of both types are reclassified into different categories according to the tags of the core lexical modifier in the modification pattern. For example, NN NN, NN DEG NN, NR NN, NR DEG NN, NT DEG NN all belong to the NN simple modification pattern, because NN is generally the core modifier, with the tags NR and NT subordinated to NN.

Table 4: Simple modification patterns of nouns.

Category	Pattern	Example
NN simple modification pattern	NN NN	教育/NN 理念/NN Jiaoyu linian Education concept
	NN DEG NN	妻子/NN 的/DEG 名字/NN Qizi de mingzi Wife's name
	NR DEG NN	青/NR 的/DEG 脸庞/NN Qing de liangpang Qing's face
	NR NN	洛云/NR 眼睛/NN Luoyun yanjing Luoyun's eys
	NT DEG NN	今天/NT 的/DEG 兴趣/NN Jintian de xingqu Interest today
PN simple modification pattern	PN NN	你们/PN 家/NN Nimen jia Your home
	PN DEG NN	我们/PN 的/DEG 目的/NN Women de mudi Our purpose
JJ simple modification pattern	JJ NN	小/JJ 夫妻/NN Xiao fuqi Young couple
	JJ DEG NN	最后/JJ 的/DEG 审判/NN Zuihou de shenpan The last trial
DT simple modification pattern	DT NN	这/DT 孩子/NN Zhe haizi This child
	DT M NN	这/DT 段/M 证言/NN Zhe duan zhengyan Part of testimony
CD simple modification pattern	CD M NN	两/CD 类/M 人/NN Lianglei ren Two types of people
	OD NN	第二/OD 阶段/NN Di'er jieduan The second stage
	CD NN	二/CD 爷/NN Er ye Father's second elder brother
VA simple modification pattern	VA DEC NN	痛苦/VA 的/DEC 根源/NN Tongku de genyuan
VV simple modification pattern	VV DEC NN	买来/VV 的/DEC 幸福/NN Mailai de xingfu Happiness bought

Table 5: Complex modification patterns of nouns.

Category	Patterns	Example
NN complex modification pattern	NN NN NN NN NN NN NN NN DEG NN NN NR NN NN NN NN DEG NN	我国/NN 民事/NN 诉讼法/NN Woguo minshi susongfa The civil procedural law of our country 旅客/NN 运输/NN 合同/NN 纠纷/NN Lvke yunshu hetong jiufen The conflict caused by the contrast of transporting passengers 全国/NN 的/DEC 网络/NN 故障/NN Quanguo de wangluo guzhang Nationwide internet breakdown 中国/NR 外交/NN 政策/NN Zhongguo waijiao zhengce Chinese diplomatic policy 种族/NN 屠杀/NN 的/DEG 后果/NN Zhongzu tusha de houguo Consequence of genocide
PN complex modification pattern	PN NN NN PN DEG NN NN	我们/PN 杯子/NN 表面/NN Women beizi biaomian Surface of our cup 你们/PN 的/DEG 诚信/NN 问题/NN Nimen de chengxin wenti Your integrity problem
JJ complex modification pattern	VA DEC NN NN AD VA DEC NN	简单/VA 的/DEC 胜负/NN 关系/NN Jiandan de shengfu guanxi Simple relationship between loser and winner 极度/ VA 危险/VA 的/DEC 境地/NN Jidu weixian de jingdi Extremely dangerous condition
DT complex modification pattern	JJ NN DEG NN JJ NN NN NR JJ NN NN JJ NN	原来/JJ 村口/NN 的/DEG 牌坊/NN Yuanlai cunkou de paifang The original arch beside the entrance to a village 当代/JJ 雷锋/NN 人物/NN Dangdai Leifeng renwu Contemporary heroes like Leifeng 中华民族/NR 优秀/JJ 品质/NN Zhonghuaminzu youxiu pinzhi Chinese excellent quality 职工/NN 先进/JJ 事迹/NN Zhigong xianjin shiji Staff's outstanding achievement
CD complex modification pattern	DT NN DEG NN DT NN NN DT CD M NN	这个/DT 问题/NN 的/DEG 重点/NN Zhege wenti de zhongdian key point of this issue 全/DT 社会/NN 组织/NN Quan shehui zuzhi All social organizations 那/DT 两百/CD 吨/M 物资/NN Na liangbai dun wuzi That two hundred tons of supplies
VA complex modification pattern	CD M NN NN CD NN NN CD M JJ NN	几/CD 个/M 幸福/NN 夜晚/NN Jige xingfu yewan Several happy nights 这种/CD 精神/NN 力量/NN Zhezhong jingshen lilang This spiritual strength 千万/CD 种/M 不同/JJ 职业/NN

		Qianwan zhong butong zhiye Thousands of different occupations
VV complex modification pattern	VV NN DEC NN AD VV DEC NN	被迫/AD 撤离/VV 的/DEC 战士/NN *Beipo chetui de zhanshi* Soldiers forced to retreat

3.2 Distribution of simple modification patterns in written and spoken registers

Simple modification patterns are the modifying parts of head-nouns with a single modifier. The distributions of simple modification patterns of nouns across registers are investigated in detail and displayed in the following.

3.2.1 Distribution of NN simple modification patterns

NN simple modification pattern is most favored among the modification patterns in both written and spoken registers. According to the $\chi 2$ test (Table 6), the frequency of NN simple modification in the written registers is significantly higher than the one in the spoken registers. The distribution of NN simple modification patterns across registers is shown in Figure 1 and 2. Among the written registers, the proportions of NN simple modification fluctuate smoothly, and especially in official document, NN simple modification occurs more common than in the other written registers. For the spoken registers, the frequency of NN simple modification displays a zigzag tendency, where court is at the top preferring NN simple modification, and in contrast, Beijing dialects has the lowest proportion of NN simple modification.

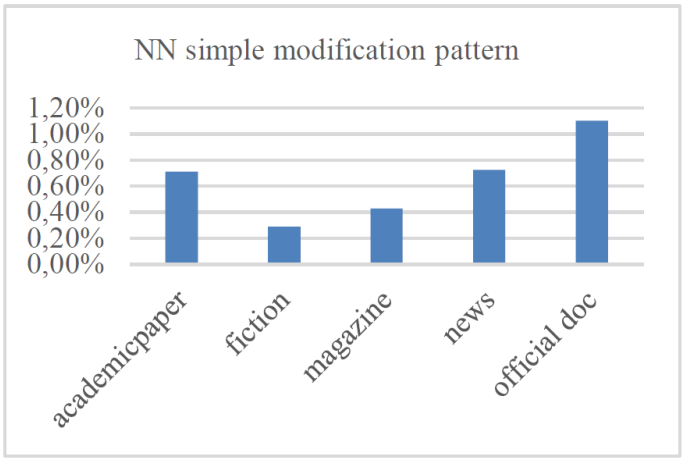

Figure 1: Distribution of NN simple modification patterns in written registers.

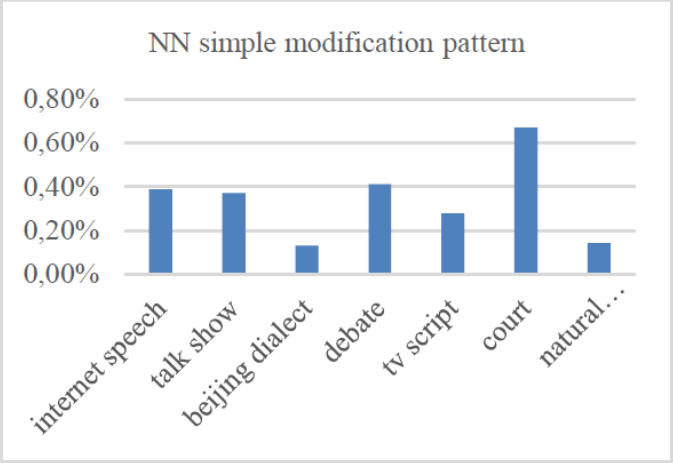

Figure 2: Distribution of NN simple modification patterns in spoken registers.

Table 6: χ2 of NN simple modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	χ2	Sig
NN simple modification patterns	8,760	8,153	21.785	<0.001

3.2.2 Distribution of PN simple modification patterns

According to the χ2 test (Table 7), there are important differences in the distribution of PN simple modification patterns between the written and the spoken corpus. PN simple modification patterns in the written corpus are much less common compared with those in the spoken corpus. PN simple modification pattern is generally more preferred in each register of the spoken corpus than in the written registers (see Figure 3 and 4). Exceptionally, fiction in the written corpus has the highest frequency of PN simple modification patterns among the written registers (see Figure 3).

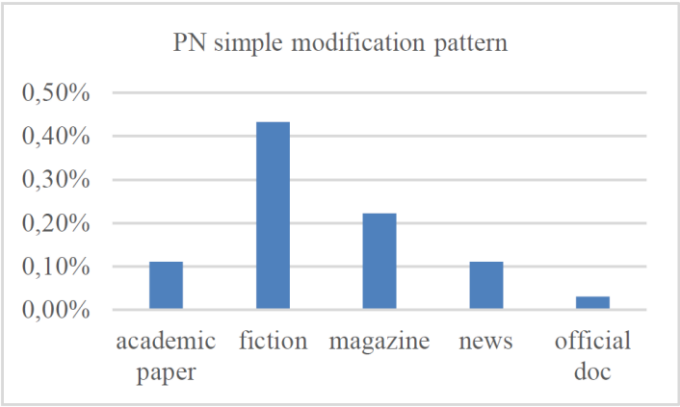

Figure 3: Distribution of PN simple modification patterns in written registers.

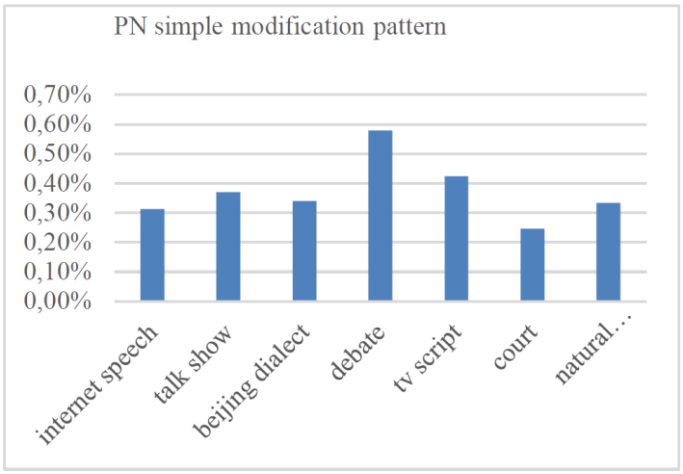

Figure 4: Distribution of PN simple modification patterns in spoken registers.

Table 7: $\chi 2$ of PN simple modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig
PN simple modification patterns	907	1,885	342.580	<0.001

3.2.3 Distribution of JJ simple modification patterns

The distribution of JJ simple modification patterns differs significantly between the written and the spoken registers. The written registers show a stronger preference for JJ simple modification patterns (see Figure 5, Figure 6 and Table 8). Conversely, JJ simple modification patterns in the spoken registers is less common than in the written registers, and Beijing dialect and natural conversation have the least number of JJ simple modification patterns.

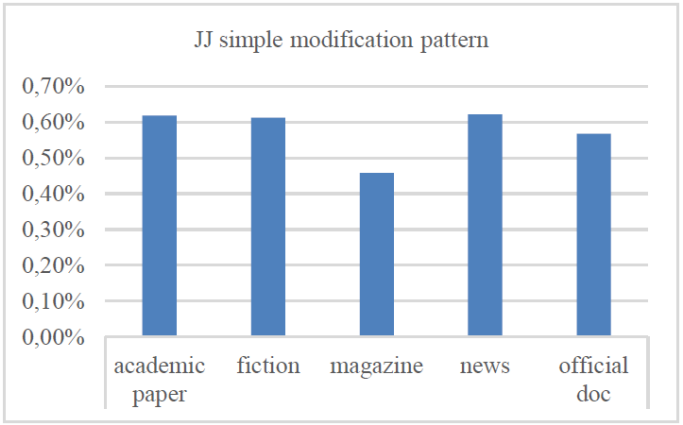

Figure 5: Distribution of JJ simple modification patterns in written registers.

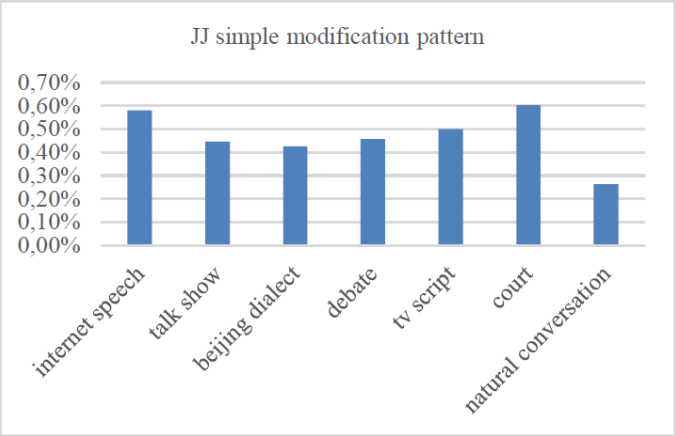

Figure 6: Distribution of JJ simple modification patterns in spoken registers.

Table 8: χ2 of JJ simple modification patterns

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig
JJ simple modification patterns	2,874	2,411	40.562	<0.001

3.2.4 Distribution of DT simple modification patterns

There is a marked difference across registers in DT simple modification patterns (see Table 9). DT simple modification patterns in the spoken registers are more frequent than those in the written registers (see Figure 7, Figure 8 and Table 9). Among the spoken registers, DT simple modification pattern is proportionally most common in Beijing dialect, and in the written registers, fiction has the relatively highest frequency of DT simple modification patterns (see Figure 7).

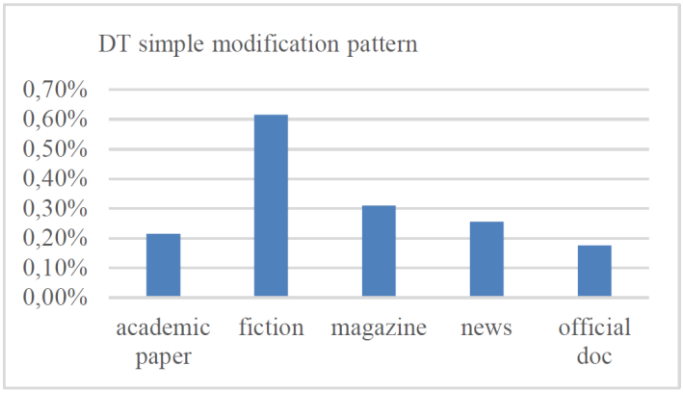

Figure 7: Distribution of DT simple modification patterns in written registers.

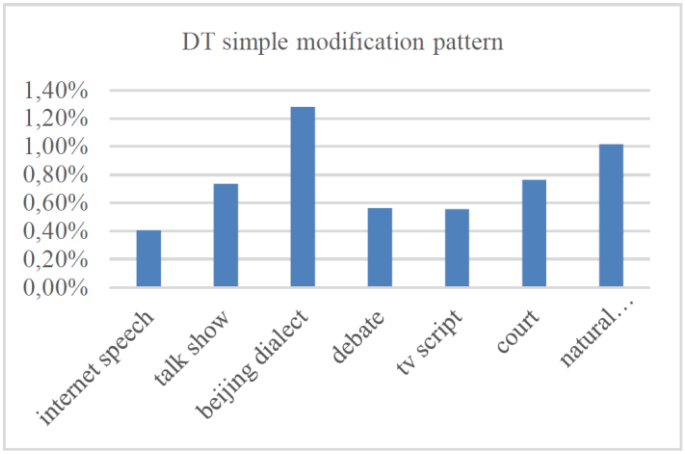

Figure 8: Distribution of DT simple modification patterns in spoken registers.

Table 9: $\chi 2$ of DT simple modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig
DT simple modification patterns	1,760	3,901	809.730	<0.001

3.2.5 Distribution of CD simple modification patterns

The distribution of CD simple modification patterns is significantly different between the written registers and the spoken registers. CD simple modification patterns in the written registers are less common than those in the spoken registers (see Table 10). Among the spoken registers, CD simple modification patterns is by far most common in the court, and among the written registers, the frequency of CD simple modification pattern in fiction outweighs other registers (see Figure 9 and 10).

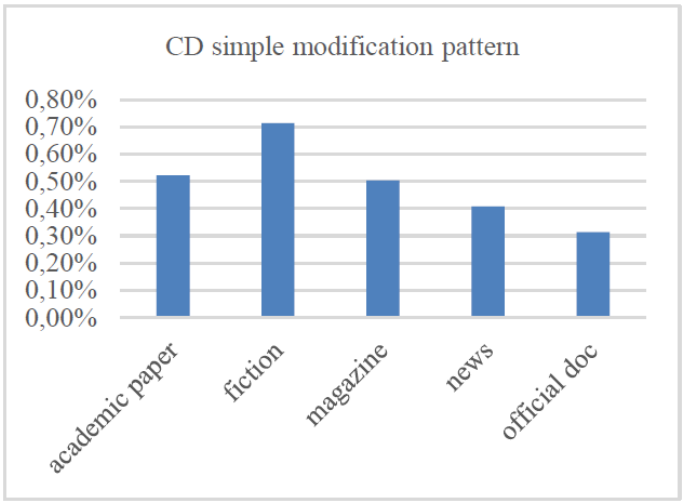

Figure 9: Distribution of CD simple modification patterns in written registers.

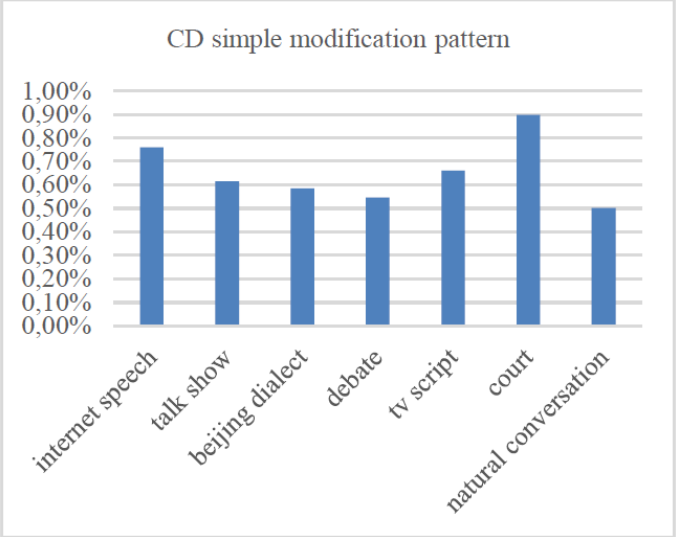

Figure 10: Distribution of CD simple modification patterns in spoken registers.

Table 10: $\chi 2$ of CD simple modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig
CD simple modification patterns	2,460	3,332	131.282	<0.001

3.2.6 Distribution of VA simple modification patterns

In general, there is nearly no difference in the frequencies of VA simple modification patterns between the written and the spoken registers (see Table 11), but a couple of registers from the written and spoken corpus show significant difference from other registers, including official document and debate (see Figure 11 and 12). From Figure 6, it can be seen that official document uses a much lower number of VA simple modification patterns than the other registers from the written corpus, and debate has the highest frequencies of VA simple modification patterns in the spoken corpus.

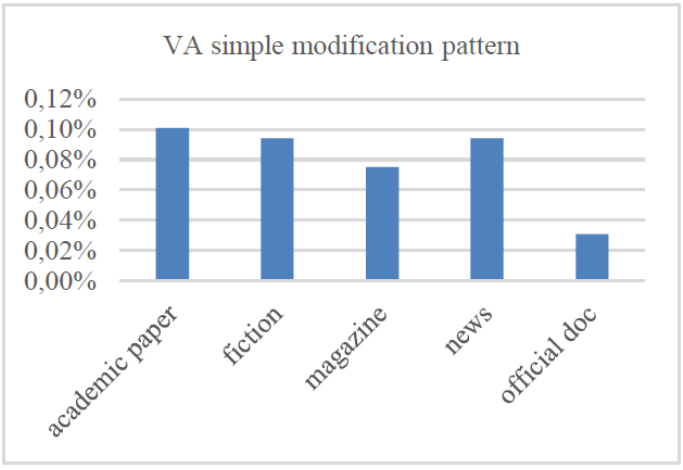

Figure 11: Distribution of VA simple modification patterns in written registers.

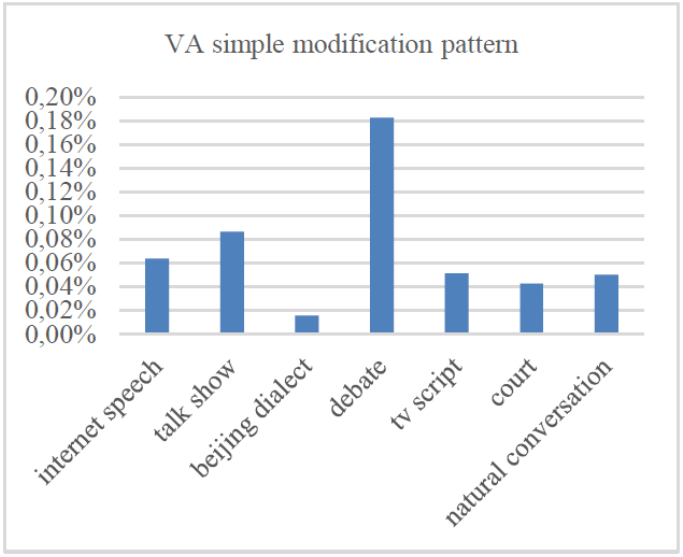

Figure 12: Distribution of VA simple modification patterns in spoken registers.

Table 11: $\chi 2$ of VA simple modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig
VA simple modification patterns	395	382	0.218	0.641

3.2.7 Distribution of VV simple modification patterns

VV simple modification patterns are significantly more frequent in the spoken registers than in the written registers (see Table 12). In debate, verb modifiers are used with the highest frequency, while in official document, such patterns are very rare (see Figure 13 and 14).

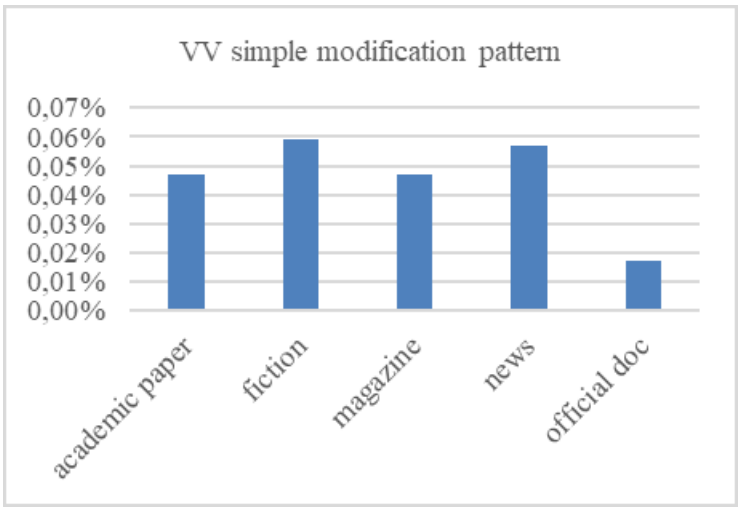

Figure 13: Distribution of VV simple modification patterns in written registers.

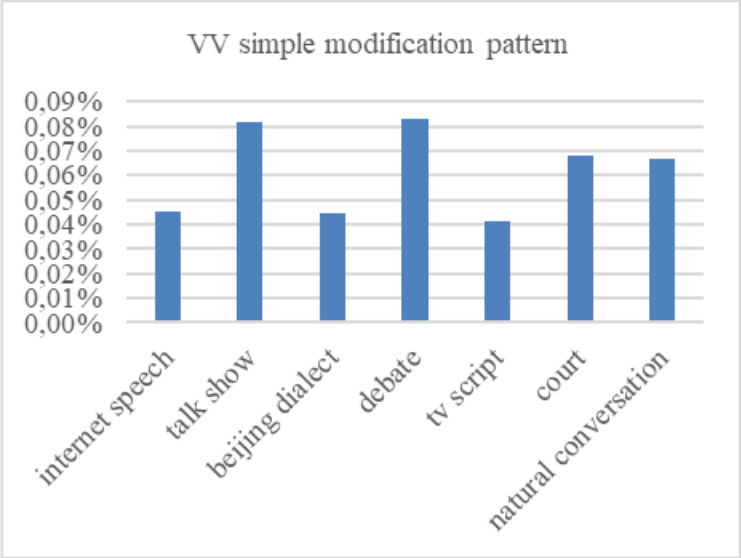

Figure 14: Distribution of VV simple modification patterns in spoken registers.

Table 12: $\chi 2$ of VV simple modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig
VV simple modification patterns	227	314	13.991	<0.001

3.3 Complex modification patterns across registers

The complex modification patterns are the modifying structures of head-nouns with longer sequences of premodifiers − two- or three-word premodifications.

3.3.1 Distribution of complex modification patterns across registers

The complex modification patterns distribute significantly differently from the simple modification patterns, and the simple modification patterns occur three times more than the complex modification patterns (see Table 13). The frequency of complex modification patterns is significantly higher in the written registers than in the spoken registers (see Table 14). The proportion of complex modifiers is much higher in official document and academic paper than in the other registers. In the spoken registers, talk show, debate, and court have relatively more common complex patterns than the other spoken registers (see Figure 15 and 16).

Table 13: χ2 test result of two modification types.

Pattern type	simple modification patterns	complex modification patterns
frequency in corpus	37,165	10,828
χ2		14452.890
Sig.		<0.001

Table 14: χ2 test result of complex modification patterns.

Pattern type	Freq. in Written Corpus	Freq. in Spoken Corpus	χ2	Sig.
complex modification patterns	7,457	6,066	143.081	<0.001

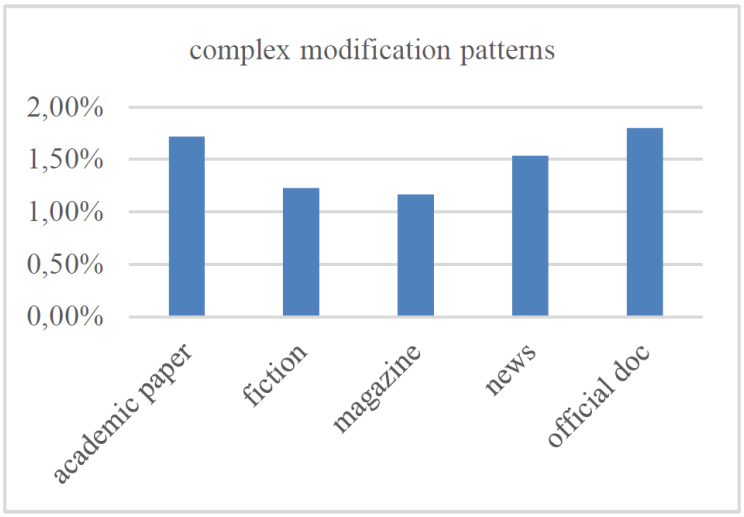

Figure 15: Distribution of complex modification patterns in written registers.

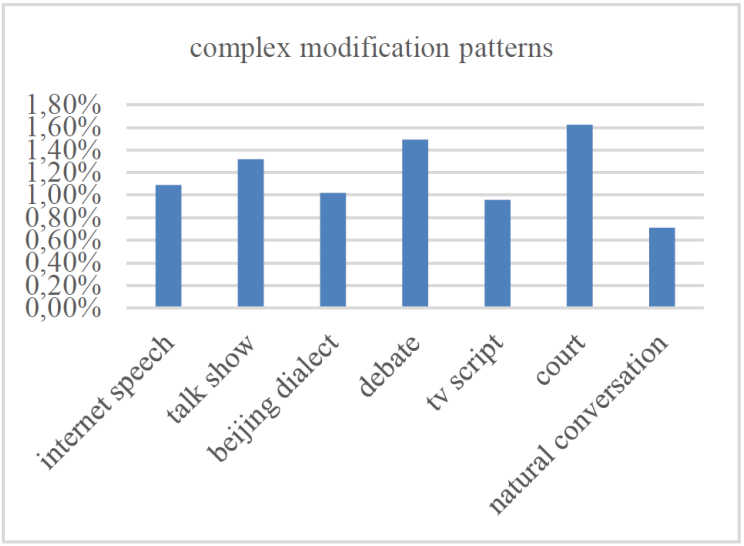

Figure 16: Distribution of complex modification patterns in spoken registers.

3.3.2 Complex modification patterns across registers

The distribution of complex modification patterns are in line with their simple forms. Among the seven kinds of complex modification patterns, NN complex pattern, JJ complex pattern, and CD complex pattern are significantly more frequent in the written corpus, and PN complex pattern, DT complex pattern, and VV complex pattern are relatively more common in the spoken corpus. The frequency of VA complex pattern in both corpora is roughly equal (see Table 15).

Table 15: $\chi 2$ test result of complex modification patterns.

Pattern	Freq. in Written Corpus	Freq. in Spoken Corpus	$\chi 2$	Sig.
NN complex pattern	3,256	1,856	384.831	<0.001
PN complex pattern	170	443	120.766	<0.001
JJ complex pattern	978	563	111.396	<0.001
VA complex pattern	433	434	0	1
DT complex pattern	364	614	63.458	<0.001
CD complex pattern	699	414	72.584	<0.001
VV complex pattern	247	357	19.682	<0.001
All	6,147	4,681	200.380	<0.001

4 Major Findings

In this chapter, the major modification patterns of nouns identified in the corpus and their distributions across the registers according to their communication functions will be discussed.

4.1 More noun modification patterns found based on corpus

With the aid of the corpus, the general descriptions on modification patterns in the previous studies can be transformed into specific and detailed ones in this study. For instance, noun modifiers are classified into NN (common noun), NR (proper noun) and NT (temporal noun). Besides, some new patterns are discovered with the computer-aided method, such as DT NN and JJ NN. As for the complex modification patterns, their internal elements are displayed by the part of speech, which is specific and transparent, while in the previous studies, these patterns are mainly investigated from the more abstract perspective of syntactic functions like “subject-predicate construction”. Since the part of speech is only a limited set of elements that relates only to surface manifestations instead of syntactic abstractions, the metalanguage is more helpful for the learners or the machines of natural language processing to understand the structures of noun phrases quickly (Hunston, 2000). The comparison between the main modification structures in the previous studies and the modification patterns in the current study is displayed in Table 16.

Table 16: Modification patterns in the previous studies and this study

Previous studies	Current study	
Structures of noun phrases	**simple modification patterns of noun**	**complex modification patterns of noun**
noun or noun phrase + (de) + noun	NN NN	NN NN NN
	NN DEG NN	NN NN DEG NN
	NR NN	NR NN NN
	NR DEG NN	NN DEG NN NN
	NT DEG NN	NN NN NN NN
pronoun or pronoun phrase+ (de) + noun	PN DEG NN	PN NN NN
	PN NN	PN DEG NN NN
adjective or adjective phrase + (de) + noun	JJ NN	JJ NN NN
	JJ DEG NN	NN JJ NN
		NR JJ NN
		JJ NN DEG NN
determiner or determiner phrase + (de) + noun	DT NN	DT NN NN
	DT M NN	DT NN DEG NN
		DT CD M NN
quantifier or quantifier phrase+ (measure words) + noun	CD NN	CD NN NN
	CD M NN	CD M NN NN
	OD NN	CD M JJ NN
adjective or adjective phrase + (de) + noun phrase	VA DEC NN	VA DEC NN NN
		AD VA DEC NN
verb or verb phrase + (de) + noun	VV DEC NN	VV NN DEC NN
		AD VV DEC NN

4.2 Communication functions of simple modification patterns

NN modification pattern is the most favored modification pattern in both written and spoken registers, but the written language prefers much more noun modifications than the spoken register. NN modification pattern conveys an extremely dense informational package and shows the logical relations between the nominal modifier and the head noun (Biber, 2009). Therefore, in the written registers, which are elaborate and concise, NN modification pattern is highly needed to imply a complicated meaning with high informational density. In particular, in official document of the written registers and court of the spoken registers, the higher ratio of NN modification patterns corresponds to its crucial function of transmitting exact, formal, and abstract information; NN modification patterns are by far least common in Beijing dialect, a typically local conversation which requires lower density of information.

PN modification patterns in the spoken corpus are much more common compared with those in the written corpus. Pronouns generally refer to things that are present in the communication situation: oneself, the listener, other people, or objects (Biber, 2009). The denser use of PN modification patterns reflects the interactive situation and personal participation of the spoken registers. In contrast, the informational purposes of the written registers need a lower proportion of PN modification patterns. Fiction has a relatively higher frequency of PN modification patterns, as it is similar to the spoken registers because the fictional characters interact with one another revealing their personal thoughts and attitudes in the fictional world (Biber, 2009).

DT modification patterns in the spoken registers are generally more frequent than those in the written registers. Determiner has the function of indicating the subjects related to human beings, and it semantically indicates the knowledge of the referent between the speaker and the addressee, the proximity of the reference to the speaker and the addressee, and the connection between the participants (Biber, 1999). Therefore, the spoken registers, where the human-centred topics are abundant, have much higher frequencies of DT modification patterns. The density of DT modification patterns is lower in the written registers, as they are generally concerned with kinds of entities and concentrate less on human relations. As Beijing dialect is particularly centred on the interaction between human beings, it has the highest frequency of DT modification patterns among the spoken registers. Fiction in the written registers prominently deals with the relationship of the characters, and therefore, DT modification patterns in it are also far more common than in other written registers.

CD modification patterns are more numerous in the spoken registers than in the written registers. Cardinal and ordinal numbers are used in counting to indicate quantity; CD modification patterns specifies the number or amount of the entities referred to. In the interactive speech, the speakers commonly use CD modification patterns to offer the quantitative information about the entities referred to. Especially in court, the information concerned with articles of law, time of events, amount of money, number of participants, and so on are extremely frequently mentioned. In contrast, the specific quantitative information is not so much needed in the written registers which involve more abstract topics and concepts. Fiction in the written registers uses higher frequency of CD modification patterns due to its great number of dialogue passages and entity descriptions.

JJ modification patterns are more preferred by the written registers than the spoken registers.Most JJ modifiers are attributive adjectives, preceding head nouns and modifying common nouns. Attributive adjective is one of the essential methods to pack additional information into noun phrases (Biber, 1999). The greater frequency of JJ modification patterns in the written registers reflects their heavy reliance on the denser presentation of the packed information. Conversely, in the temporary communicative situations like Beijing dialect and natural conversation, the lower information capacity requires fewer additional modifying adjectives.

VA modification patterns distribute likewise in the written and the spoken registers. This sort of patterns occurs much less in official document than the other written registers, and more common in debate than the other spoken registers. Predicative adjectives characterize the qualities of people, things, and the states of affairs. In the formal communicative situation of official document, with large amounts of highly condensed and professional noun phrases, VA modification patterns are rather slightly used, in order to avoid the overloaded information. On the contrary, in the impromptu debate, debaters require relatively dense use of VA modification patterns to add judgment information or to achieve the accurate expressions.

VV modification patterns are significantly more frequent in the spoken registers than in the written registers. Verbs denote actions, processes, or states (Biber, 1999). In the spoken registers, a larger number of actions and events are frequently referred to by conversational participants; in contrast, in the written registers, actions and events are less concentrated on than entities.

4.3 Communication functions of complex modification patterns

Complex modification patterns occur significantly much less than simple modification patterns. Complex modification patterns efficiently packs dense informational content into as few words as possible (Biber, 1999); they usually have embedded or ambiguous logical relations among constituents, as some words in modification patterns modify other modifiers instead of the head noun. Complex modification patterns place a heavy burden on the participants' memory and comprehension, and it takes more time for the readers to understand the meaning of them. Therefore, compared with complex modification patterns, simple modification patterns are generally more preferred in both written and spoken registers.

The proportion of complex modifiers is much higher in the written registers than in the spoken registers. The greater frequency of modification patterns in the written registers contributes to the higher lexical density in written discourse. More complex modification patterns and a higher lexical density are well-adapted to serve the communicative function of the written registers, especially in official document and academic paper, which typically involve complicated subject matters and have a high information capacity.

Conversely, in the concrete and context-dependent spoken registers, with more simple modification patterns, information is much less tightly packed, which simplifies the process of the speakers' encoding and the hearers' decoding. However, as talk show, debate, and court allow, to some extent, a pre-edition, and the information density in these registers is much higher than other spoken registers, the higher frequency of complex modification patterns contributes to the more formal communicative situations of these spoken registers.

4.4 Information density across registers

NN modification pattern is the most favored one in both written and spoken registers, and nominal features are one of the most obvious ways in which the written registers differ from the spoken registers (Biber, 2009). The frequency of nouns in a register is closely connected with its information density (Biber, 1999). In this study, it can be inferred that the more NN modification patterns a register has, the denser information it expresses. Since NN modification patterns have one to three or more sequential noun modifiers, the multiple nouns are embedded in the sophisticated semantic and structural relations, in which one noun modifies the other nouns or the rest of nouns in the sequence. NN modification pattern delivers a wide range of meaning relationships in a succinct form. Therefore, the higher frequency of NN modification patterns contributes to the higher informational density in registers. The

distribution of NN modification patterns can be regarded as the manifestation of the information density across registers. In Figure 17, the information density is arranged in the descending order.

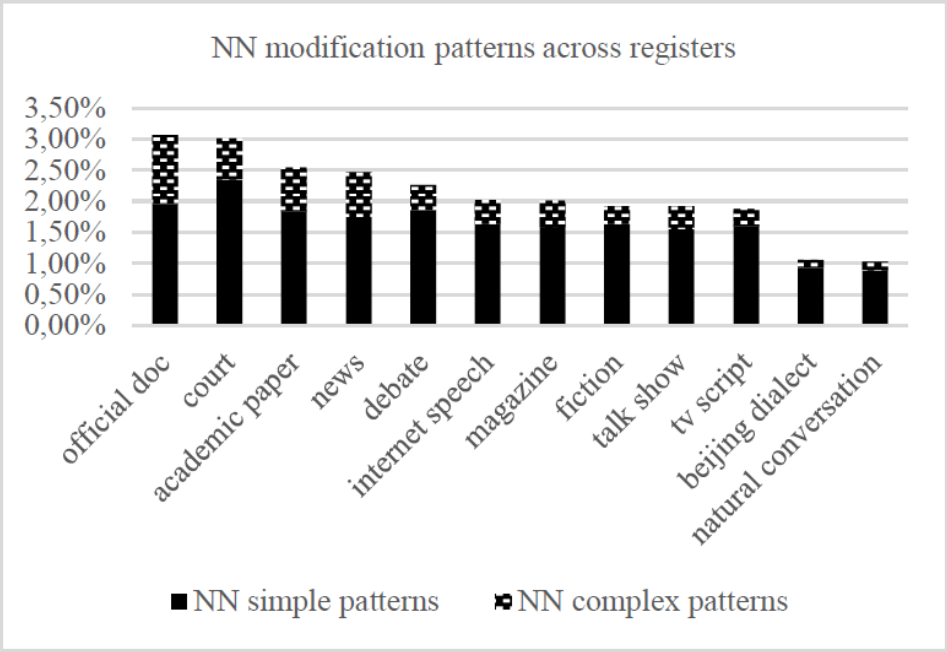

Figure 17: Distribution of NN modification patterns across registers

Official document and court have greater information density than other registers. Official document focuses on conveying the actual intention of official authority and affairs accurately (Peng and Zhao, 2014). The greater information density of official document makes it more logical, concise, and accurate. Court belongs to the professional written registers. It requires the precise, formal, and solemn representation of the legal meaning (Li, 1994). Legal terminologies and statements require a large number of NN modification patterns to redefine the head nouns, so the information density is rather high.

On the contrary, Beijing dialect and natural conversation hold the slightest information density compared with other registers. They are produced and processed in real time, by people who are face-to-face, sharing personal information and developing a personal relationship (Biber, 2009). The speakers in both registers are planning what to say while they are speaking, so they frequently use short sentences, with many utterances not being structurally complete sentences at all. The speakers in conversations do not have enough time to formulate the dense noun phrases which demand more time to process. The slight information density results from the communicative focus on "you and I" and the fact that the participants are together at the same place and time (Biber, 2009).

In general, the information density decreases gradually from official document to natural conversation (see Figure 17). On the grounds of the information density of register, it is of practical value to

distinguish the type and the difficulty of texts. Firstly, as for the selection of language-teaching materials, the degree of information density should be taken into account in order to be adapted for learners' language acquisition levels. Secondly, in natural language processing, inserting the feature of information density is beneficial for improving the accuracy of text analysis and language generation.

5 Conclusion

According to the above analysis, the major findings can be concluded: a. based on corpus, more modification patterns of head-nouns are found than in the previous theoretical research; b. in general, simple modification patterns and complex modification patterns have the similar distributional tendency across registers. Among the seven kinds of modification patterns, NN simple/complex pattern, JJ simple/complex pattern and CD simple/complex pattern are significantly more frequent in the written corpus, and PN simple/complex pattern, DT simple/complex pattern and VV simple/complex pattern are relatively more common in the spoken corpus. The frequency of VA simple and complex pattern is roughly equal in both written and spoken registers; c. complex modification patterns distribute significantly differently from simple modification patterns, and the frequency of complex modification patterns is one-third of the frequency of simple modification patterns; d. the more NN modification patterns a register has, the denser information it expresses and the information density across registers decreases gradually from official document to natural conversation; e. the distinctive distributions of the modification patterns of head nouns reflect their differences in the communicative purposes, the situational circumstances, and the physical settings of the different registers.

This study provides a comprehensive description on the modification patterns for nouns in Chinese with quantitative evidence based on corpora. It analyzes the usage of the modification patterns in the authentic communicative contexts. Through the examination on the distributions of the modification patterns in the written and the spoken registers, the communicative functions of different modification patterns are discussed across registers, which is rare in the previous studies. This study shows the close relationship between the functions of modification patterns and the registers' communicative functions.

Taking the communicative functions into account will be beneficial for a range of theoretical and practical studies in Chinese, such as translation studies between Chinese and other languages, Chinese natural language processing, and Chinese teaching methods. For translation studies, the functions of the structural patterns across registers are valuable references for translators so that they could select the proper language style in the translation process. In the natural language processing, the functional information of language patterns can be tagged as the language features to improve the analytical accuracy. What is more, the understanding of the communicative functions of Chinese patterns is very helpful for learners of Chinese – they can choose the appropriate patterns in the specific communicating circumstances.

Acknowledgements

This project is supported by the National Social Science Foundation of China (17BYY002).

References

Biber, D., Conrad, S. (2009). *Register, genre, and style*. Cambridge and New York: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman grammar of spoken and written English* (Vol. 2). London: Longman.

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral Literature. In: Tannen, D. (ed.). *Spoken and written language: Exploring Orality and Literacy*. Norwood, NJ: Ablex.

Cheng, S. (2009). *A Study on the Priority Sequences of Multiple-attributive Phrases in Modern Chinese*. Central China Normal University. (Ph.D. dissertation)

Hu, G. (2003). *The Development of the Information Database for the Semantic Structures of the Chinese Noun Phrases with Words of Events*. Graduate School of Chinese Academy of Social Sciences. (MA thesis)

Hu, X., Zeng, J. (2009). A corpus-based study of explicitation of grammatical markers in Chinese translated fiction. *Foreign Languages Research*, 5, pp. 72–79.

Huang, B., Liao, X. (2007). *Contemporary Chinese*. Beijing: Higher Education Press.

Hunston, S., Francis, G. (2000). *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.

Li, W. (1994). On the characteristics and translation of legal English. *Chinese Translation*, 6, pp. 15–18.

O'Donnell, R.C. (1974). Syntactic differences between speech and writing. *American Speech*, 49, pp. 102–110.

Olson, D. (1977). From utterance to text: the bias of language in speech and writing. *Harvard Educational Review*, 47(3), pp. 257–281.

Peng, H., Zhao, L. (2014). *Official document writing*. Guangzhou: Jinan University Press.

Tony, M., Andrew, H. (2012). *Corpus linguistic: method, theory and practice*. Cambridge: Cambridge University Press.

Xiong, W. (1996). Position of attributive, adverbial and object in English and Chinese. *Chinese Teaching in the World*, 4, pp. 71–75.

Zhang, A., Han, L. (1997). Grammatical function of the attributives. *Journal of Jiangsu Normal University (Philosophy and Social Sciences Edition)*, 1, pp. 69–73.

Zhang, L., Zhao, F. (2011). Splitting Strategy of the attributives in English-Chinese translation. *Modern Chinese*, 7, pp. 97–99.