

Quantitative Analysis of Function Words in Chinese Legal Texts*

Ivan Straňák

Kvantitatívna analýza sysémantických slov v právnom texte čínskeho jazyka

Resumé V článku sa zaoberáme kvantitatívnou analýzou právnych textov z ČĽR, na základe ktorej sme určili významnosť niektorých slov nami skúmaných sysémantických slovných druhov a pasíva. Spomínané štylisticky príznakové prvky sme následne porovnali s ich výskytom v registri písaného jazyka *shūmiànyǔ* 書面語, pričom sme v oboch prípadoch využívali korpusy čínskeho jazyka dostupné na internete.

Abstract In this paper we analyzed the function words as the specific language items of the register of Chinese legal texts. The aim of our research was the quantitative analysis of the function words and the passive markers in the selected Chinese legal texts, which we compared with their occurrence in the register of written language *shūmiànyǔ* 書面語. Throughout the research we used online Chinese corpora.

Key words Slovak, Chinese, Language · Corpus Linguistics · Legal Texts, Part of Speech, Functions Words

I Introduction

With the social development people start to use different ways of expressing their thoughts and ideas, start to use specific vocabulary both in spoken and

* Contents of this article first appeared as parts of the author's M. A. Thesis »Štylisticky príznakové prvky právnych textov čínskeho jazyka« [Stylistic Markers in the Chinese Legislative Texts] (Bratislava: Univerzita Komenského, 2015).

written language, which results in creation of new language registers as well. By creating rules and law throughout the past centuries, we have also managed to create a distinct register based on legal texts. Since each language register consists of specific vocabulary, it is also regarded as unique variety of language. Based on specific elements of each register we can determine the differences among registers, such as specific vocabulary, or different word usage, which might prove their uniqueness.

2 *Language Register*

Firstly, let us briefly summarize what language register is. Language register is a term used for language typical for specific domain, e.g. register of legal language, etc.¹ It can also be characterised as a functional variety of language, which is appropriate to different situations of use.²

It is also important to note, that the language register differs from the language dialects, which are defined by their usage by specific groups living in the specific area, whereas language register is defined by their situation of use.³

Based on the information above, we tried to determine the register of the Chinese legal texts as a unique variety of language used in specific situations. Mastering the language register can be considered as a crucial part of learning a foreign language, due to its usage of unique language items and other specific elements, which make it different from other language varieties. Therefore, it is important to analyse and understand the main characteristics of each register, which make registers either similar, or different from each other.

- 1 Douglas Biber, *University Language: A Corpus-Based Study of Spoken and Written Registers* (Philadelphia, PA: John Benjamins, 2006), 11.
- 2 Deborah Cao, *Translating Law* (Clevedon: Multilingual Matters, 2007), 18.
- 3 Douglas Biber, Susan Conrad and Randi Reppen, *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge, England: Cambridge University Press, 1998), 135.

3 *Language Register of Legal Texts*

When analysing the legal texts, we suppose that the legal texts are a subset of the register of written language in the dichotomous model of written and spoken language registers, therefore it is necessary to differentiate between spoken and written register in Chinese language. For instance, Gajdoš considers the spoken language *kǒuyǔ* 口語 as spoken variety of Chinese language, which is used mainly in dialogues, is more redundant, spontaneous, consists of short sentences etc., whereas long sentences, more complicated sentential constructions, monological nature and exactness are considered to be the predominant characteristics of the register of written language.⁴ Through many changes and developments, it has developed into a specific new functional variety different from other language varieties.⁵

Legal texts analysed in this paper are therefore considered to be the subvariety of the written language and we also suppose that the language of the register of legal texts should be partially different from the register of the written language.

The language of the legal texts is considered to be the technical language and she also emphasises the need to consider it as a language register, since the legal language is not just made of the specific vocabulary used by authors.⁶ It also

4 Euboš Gajdoš, »Synsémantické slová v rámci stratifikácie čínskeho jazyka« [Synsemantic Words and Registers Variation in Chinese Language], in *Miscellanea Asiae Orientalis Slovaca*, ed. by Martin Slobodník and Euboš Gajdoš (Bratislava: Univerzita Komenského, 2014), 125.

5 Pan Qingyun 潘慶雲, »Falü yuyan shi yi zhong you bie yu ziran yuyan de jishu yuyan 法律語言是一種有別於自然語言的技術語言« [Language of Law is a Type of Technical Language Different from the Natural Language], *Jianghan daxue xuebao* 江漢大學學報 23 (2004), 18–22.

6 Cao, *Translating Law*, 15–17.

uses the vocabulary different from the common language, with its specific expressions and lexical items.⁷

Based on the information above we can deduce that the language register of the legal texts might be considerably different from other language registers, which was also proved by Gotti, who considers the main features of the register of legal texts to be the precision of interpretation, the avoidance of ambiguity and complexity.⁸

4 Methodology of Research

Legal texts have the specific features which make them unique among other language registers. The presence of the specific features in the legal texts help us further specify the language register of the legal texts and its characteristics. Legal texts, as well as other varieties of written language, have typical lexical items, which make them different from the other language varieties. Therefore we decided to analyse function words and markers of passive voice in the Chinese legal texts as significant features of the language register of legal texts and we supposed that their frequency of presence in legal texts might be similar with their frequency of presence in the register of written language, yet we also supposed there might exist some differences in their presence in both registers as well, which we analysed by quantitative analysis using Chinese language corpus.

Throughout the research and analysis of the function words and the passive markers in the Chinese legal texts, we used the legal texts available in the Slovak-Chinese parallel corpus (SiHanKu), which were available in subcorpus *falv*. Since we also wanted to compare the values gained from the analysis of the legal texts with their presence in the register of the written language, in order to get the

7 Dennis Kurzon, »Legal Language: Varieties, Genres, Registers«, *International Journal of Applied Linguistics* 7 (1997), 124.

8 Maurizio Gotti, »Linguistic Features of Legal Texts: Translation Issues«, *Statute Law Review*, first published online 23 Aug 2014 <doi:10.1093/slr/hmu027>, 12 pp.

frequency of occurrence of the equivalents of analysed function words and markers in the register of written language, we used the online corpus Araneum Sinicum (Zh Web 15.03) 940 M (AranSini), where we searched for the specific units we were looking for and the frequency of their occurrence in the corpus. The measured frequency served us to determine the occurrence of function words and passive markers in SMY. The i.p.m. values⁹ of frequencies gained from each corpus were later compared in order to see the differences between the two registers and the specific usage of analysed language items in the register of Chinese legal texts.

Throughout the research we used two available corpora of Chinese language, namely the Slovak-Chinese parallel corpus (SiHanKu)¹⁰ and the Chinese web corpus Araneum Sinicum (ZH Web 15.03) 940 M (AranSini).¹¹ SiHanKu was the corpus, which contained the Chinese legal texts,¹² where we were looking for the language items, which were the object of our research. AranSini served us as a corpus containing texts representing the register of written language. During our research the corpus SiHanKu consisted of 505,859 tokens, while the corpus AranSini, on the other hand, consisted of 940,398,466 tokens.¹³ Both corpora

- 9 I.p.m. (Instances per million) represents the base unit of frequency in corpus linguistics, which determines the average occurrence of a word in a corpus with a hypothetical range of one million words.
- 10 The Slovak-Chinese parallel corpus. Available on: <sihanku.fphil.uniba.sk/> (last retrieval 30 March 2015).
- 11 Chinese web corpus. Available on: <uniba.sk/bonito/run.cgi/first?> (last retrieval 30 Apr 2015).
- 12 The texts we used and which belonged to subcorpus *falv* contained only the legal texts of People's Republic of China written in simplified characters.
- 13 Number of tokens during the research, which took place on 10 Apr 2015. It is important to note, that the corpus AranaSini does not consist of texts belonging to the register of written language SMY, but of other texts as well, therefore it cannot be considered as a perfect source of texts belonging to the register of written language SMY. Yet with its huge amount of tokens and database of texts, it can be considered as a reliable source of written language, even if there might be some deflections in the calculation, regarding the non-uniform character of the corpus.

use the Non Sketch engine software interface, which offers the services aimed at creating language corpora.

Even though both SiHanKu and AranSini are reliable sources for our research, there are some deficiencies in using both corpora as well, which can be seen in the wrong morphological annotation. These mistakes are regarded as common phenomena, which happens in the morphological annotation process. From a statistical point of view, we decided to use corpora, which were both made by Sketch Engine corpus query system. The advantage of using the above mentioned corpora is the fact that they both use the same morphological annotation, which we worked with throughout the whole research.¹⁴

5 *Parts of Speech in the Chinese Legal Texts*

Firstly, we wanted to determine the i.p.m. value of the specific parts of speech in the legal texts of SiHanKu.¹⁵ We noticed that the nouns are the most frequent part of speech in the Chinese legal texts, followed by verbs. Dominance of content words might indicate that the Chinese legal texts are also the texts with informational value providing static information.

Part of Speech	i.p.m.
Nouns	242 947
Verbs	180 449
Transitive verbs	67 819
Particles	61 256

14 Morphological annotation (tagset) of Slovak-Chinese parallel corpus available on <fphil.uniba.sk/katedry-a-odborne-pracoviska/katedra-vychodoazijskych-studii/slovensko-cinsky-paralelny-korpus/sihanku/>. Morphological annotation (tagset) available on: <lancaster.ac.uk/staff/xiaoz/lcmc/lcmc_tagset.htm> (last retrieval 3 March 2015).

15 In order to determine the frequency of occurrence of specific parts of speech in the legal texts of SiHanKu, we used the command [tag=>x<] for querying the parts of speech, where x represented the part of speech we looked for in the corpus.

Prepositions	40 274
Numerals	36 787
Conjunctions	33 563
Classifiers	28 130
Adverbs	21 856
Pronouns	19 023
Location nouns	16 001
Adjectives	13 292

Table 1
Values of i.p.m. of Parts of Speech in Legal Texts of SiHanKu

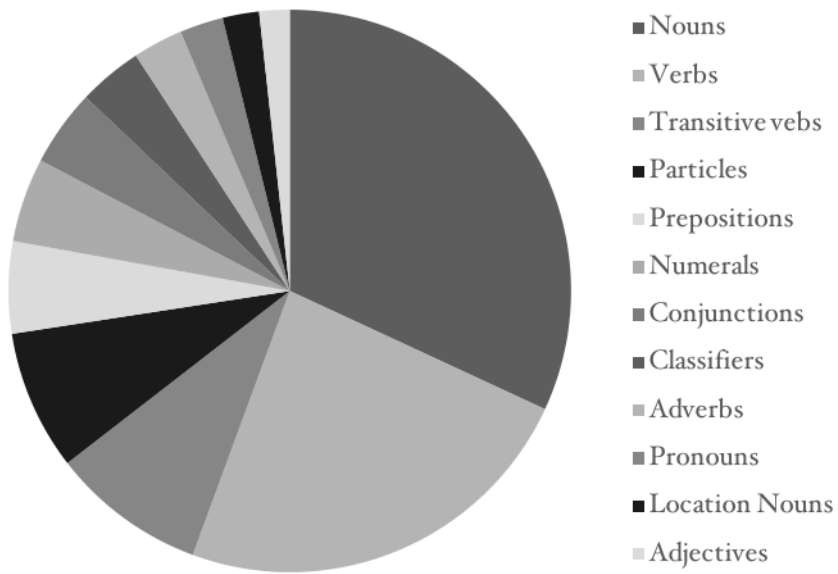


Chart 1
Parts of Speech in Legal Texts of SiHanKu

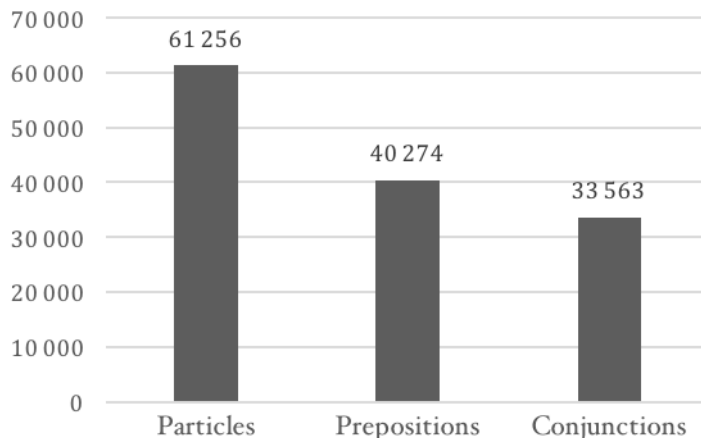


Chart 2

Value of i.p.m. of function words in the legal texts of SiHanKu

During our research we also found out that the level of occurrence of function words in the Chinese legal texts of SiHanKu was also very high, with particles being the most frequent function words, followed by prepositions and conjunctions.

5.1 *Conjunctions in the Register of Legal Texts*

Conjunctions in Chinese language are regarded as function words, which connect words, phrases or parts of a sentence. They express syntactic relationships between the constituents of a clause, or between the clauses of a sentence, etc.¹⁶

In Chinese, we distinguish between the conjunctions that indicate additive relationship, e.g. *hé* 和 ('and'), *gēn* 跟 ('and', 'with'), *yǔ* 与 ('and'), *tóng* 同 ('and', 'with'), and disjunctive relationship, e.g. *háishì* 还是 ('or') and *huòzhě* 或者 ('or'). We use the conjunction *háishì* 还是 in question sentence, when we need to choose between two alternatives, only one of which can be true or possible. The

¹⁶ Jozef Mistrík & al., *Encyklopédia jazykovedy* [Encyclopedia of Linguistics] (Bratislava: Obzor, 1993), 417.

conjunction *huòzhě* 或者 is used to present two alternatives, both of which can be possible.¹⁷

Comparing the most frequent conjunctions in both corpora we found out that in the case of the Chinese legal texts, there were some differences in usage of conjunctions, compared to their usage and frequency in AranSini corpus. For instance, the contrast between the usage of conjunction *huòzhě* 或者 in legal texts of SiHanku and the texts available in AranSini was obvious, since it was not present among the top 10 most frequent conjunctions in the AranSini corpus, whereas, according to our research, it was the most frequent conjunction with the highest level of i.p.m. value in the legal texts of SiHanKu corpus.

SiHanKu		AranSini	
Conjunctions	i.p.m.	Conjunctions	i.p.m.
<i>huòzhě</i> 或者 'or'	13 011,53	<i>bé</i> 和 'and'	2 698,45
<i>bé</i> 和 'and'	11 228,43	<i>ér</i> 而 'and', 'yet'	1 014,26
<i>bìng</i> 并 'and'	3 694,71	<i>dàn</i> 但 'and'	669,22
<i>yǐjǐ</i> 以及 'and'	1 043,77	<i>bìng</i> 并 'and'	542,64
<i>jíqí</i> 及其 'and'	735,38	<i>yǔ</i> 与 'and'	466,4
<i>yǔ</i> 与 'and'	640,49	<i>huò</i> 或 'or'	440,16
<i>ér</i> 而 'and', 'yet'	579,21	<i>rúguǒ</i> 如果 'if'	314,89
<i>dàn</i> 但 'but'	488,28	<i>jí</i> 及 'and'	288,56
<i>dànshì</i> 但是 'but'	413,16	<i>suǒyǐ</i> 所以 'so'	265,81
<i>jí</i> 及 'and'	381,53	<i>dànshì</i> 但是 'but'	244,35

Table 2

List of conjunctions with the highest level of i.p.m. in SiHanKu and AranSini

17 Claudia Ross and Jing-heng Sheng Ma, *Modern Mandarin Chinese Grammar: A Practical Guide* (Abingdon: Routledge, 2006), 97.

Based on our research we can conclude that the occurrence of conjunctions in the register of the Chinese legal texts was different from their occurrence in the register of written language SMY, which we managed to notice on the occurrence of 10 most frequent conjunctions in both corpora.

Comparing the register of the legal texts to the register of written language, we found out that some conjunctions in the register of the legal texts had higher level of i.p.m. than their equivalents in the register of written language SMY.

Sihanku		AranSini	
Conjunctions	i.p.m.	Conjunctions	i.p.m.
huòzhě 或者 'or'	13 011,53	huòzhě 或者 'or'	146,13
hé 和 'and'	11 228,43	hé 和 'and'	2 698,45
bìng 并 'and'	3 694,71	bìng 并 'and'	542,64
yǐjí 以及 'and'	1 043,77	yǐjí 以及 'and'	167,57
jíqí 及其 'and'	735,38	jíqí 及其 'and'	33,41
yǔ 与 'and'	640,49	yǔ 与 'and'	466,4
ér 而 'and', 'yet'	579,21	ér 而 'and', 'yet'	1 014,26
dàn 但 'but'	488,28	dàn 但 'but'	669,22
dànshì 但是 'but'	413,16	dànshì 但是 'but'	244,35
jí 及 'and'	381,53	jí 及 'and'	288,56

Table 3

Comparison of i.p.m. of the Most Frequent Conjunctions in SiHanKu and Their Equivalents in AranSini

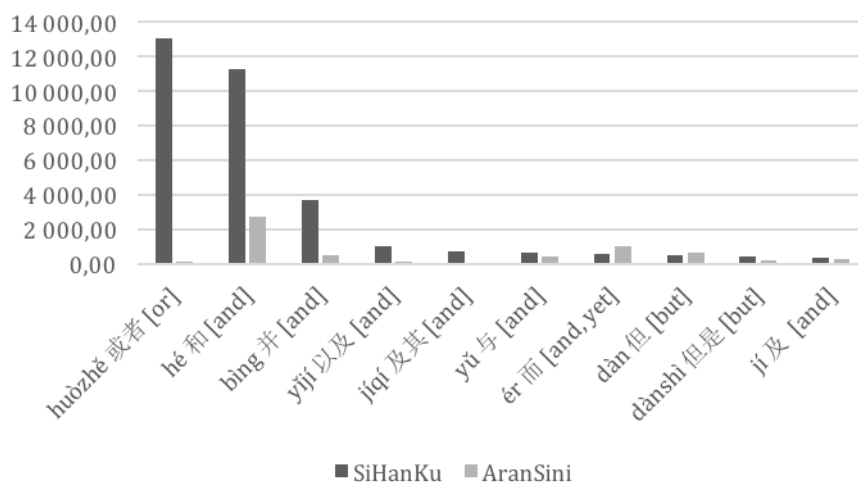


Chart 3

Conjunctions in SiHanKu and AranSini

5.2 Prepositions in the Register of the Chinese Legal Texts

Prepositions are considered to be the function words, which indicate the relationship between nouns and verbs, other nouns or adjectives. They are usually connected with verbs and their position in sentence is usually between subject and predicate.¹⁸

During the analysis of prepositions, we found out that their occurrence in the legal texts differed from their frequency of occurrence in the AranSini corpus, where we could observe different i.p.m. values of certain prepositions. For example, the prepositions *ànzhào* 按照 and *yīzhào* 依照 were not present among the most frequented prepositions of the AranSini corpus, the frequency of occurrence of the preposition *zài* 在 was also different from its frequency in SiHanKu. The preposition *zài* 在 was the preposition with the highest level of

18 Po-Ching Yip and Don Rimmington, *Chinese: A Comprehensive Grammar* (London: Routledge, 2004), 159.

i.p.m. value in AranSini, in the case of the legal texts of the SiHanKu corpus, we did not discern any major differences among the i.p.m. values of the first three prepositions. Prepositions *yóu* 由 and *xiàng* 向 were also the prepositions with a high level of i.p.m. in the Chinese legal texts, yet they did not figure among the 10 most frequent prepositions of AranSini corpus.

SiHanKu		AranSini	
Prepositions	i.p.m.	Prepositions	i.p.m.
<i>duì</i> 对 ‘to’	5 624,10	<i>zài</i> 在 ‘in’	2 882,36
<i>zài</i> 在 ‘in’	5 505,49	<i>duì</i> 对 ‘to’	851,29
<i>yóu</i> 由 ‘by’, ‘by means of’	4 515,09	<i>wèi</i> 为 ‘for’, ‘for the sake of’	744,8
<i>xiàng</i> 向 ‘towards’	2 463,14	<i>yǐ</i> 以 ‘by’, ‘with’	487,93
<i>ànzhào</i> 按照 ‘according to’	2 382,09	<i>yǔ</i> 与 ‘with’	451,79
<i>bèi</i> 被 ‘passive’	2 206,15	<i>bèi</i> 被 ‘passive’	447,18
<i>yīzhào</i> 依照 ‘according to’	2 113,24	<i>cóng</i> 从 ‘from’	421,28
<i>wèi</i> 为 ‘for’, ‘for the sake of’	1 729,73	<i>bǎ</i> 把 [accusative marker]	324,09
<i>jīng</i> 经 ‘through’, ‘after’	1 504,37	<i>gěi</i> 给 ‘for’, ‘to’	295,93
<i>yǐ</i> 以 ‘by’, ‘with’	1 460,88	<i>yòng</i> 用 ‘through’	277,41

Table 4

List of Prepositions with the Highest Level of i.p.m. in SiHanKu and AranSini

During our research we could also observe different values of i.p.m. of the specific prepositions in both corpora. Prepositions *duì* 对 and *zài* 在 represented the most frequent prepositions in both corpora, major difference could be seen in the case of the preposition *yóu* 由, which i.p.m. value in the legal texts of SiHanKu, was the third highest, yet did not reach such significant value in AranSini corpus. The same situation could also be observed among other prepositions, such as *xiàng* 向, *ànzhào* 按照, *bèi* 被, *yīzhào* 依照, *wèi* 为, *jīng* 经. Lower value of i.p.m. was seen in the legal texts of SiHanKu when analysing the preposition *wèi* 为, whereas the i.p. m. of the preposition *yǐ* 以 was similar in both corpora.

SiHanKu		AranSini	
Prepositions	i.p.m.	Prepositions	i.p.m.
<i>duì</i> 对 ('to')	5 624,10	<i>duì</i> 对 ('to')	851,29
<i>zài</i> 在 ('in')	5 505,49	<i>zài</i> 在 ('in')	2 882,36
<i>yóu</i> 由 ('by, by means of')	4 515,09	<i>yóu</i> 由 ('by, by means of')	217,54
<i>xiàng</i> 向 ('towards')	2 463,14	<i>xiàng</i> 向 ('towards')	208,59
<i>ànzhào</i> 按照 ('according to')	2 382,09	<i>ànzhào</i> 按照 ('according to')	51,62
<i>bèi</i> 被 [passive]	2 206,15	<i>bèi</i> 被 [passive]	447,18
<i>yīzhào</i> 依照 ('according to')	2 113,24	<i>yīzhào</i> 依照 ('according to')	12,24
<i>wèi</i> 为 ('for, for the sake of')	1 729,73	<i>wèi</i> 为 ('for, for the sake of')	744,8
<i>jīng</i> 经 ('through, after')	1 504,37	<i>jīng</i> 经 ('through, after')	76,88
<i>yǐ</i> 以 ('by, with')	1 460,88	<i>yǐ</i> 以 ('by, with')	487,93

Table 5

Comparison of the most frequented prepositions in SiHanKu and in AranSini

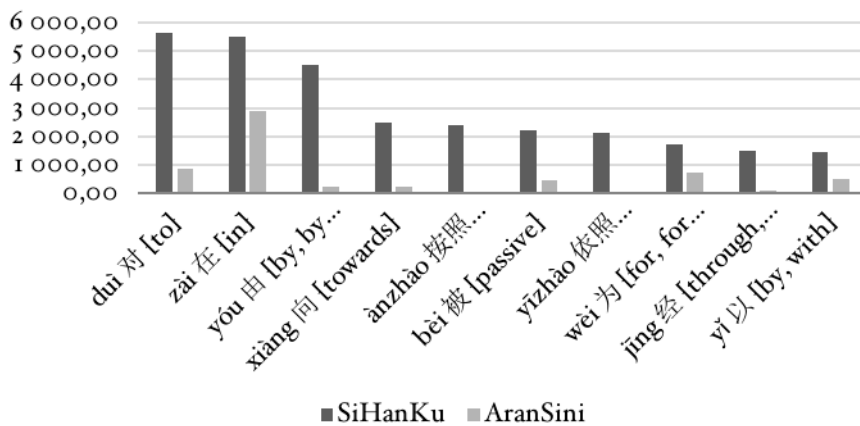


Chart 4

Prepositions in SiHanKu and AranSini

5.3 Particles

Particles in Chinese language,¹⁹ such as *de* 的, *de* 得, *de* 地, *zhe* 着, *le* 了 etc., are considered to be syllables, which express the grammatical meaning.²⁰ Particles in sentence can usually be seen connected with another part of speech, in which case they convey the grammatical meaning.²¹

Throughout our research we could also observe differences among the frequencies of the particles in the Chinese legal texts of SiHanKu and the particles present in the AranSini corpus. Based on our research we can assume, that the most frequent particles in the legal texts were structural particles, whereas in AranSini corpus we could observe that the most frequent particles, besides the structural particle *de* 的, were the particles *le* 了 and *zhe* 着. Therefore, based on our research, we can conclude that the most frequent particles in the legal texts were structural particles.

SiHanKu		AranSini	
Particles	i.p.m.	Particles	i.p.m.
<i>de</i> 的	55 677,57	<i>de</i> 的	41 555,10
<i>děng</i> 等	1 674,38	<i>le</i> 了	1 405,88
<i>suǒ</i> 所	1 536,00	<i>zhe</i> 着	346,43

19 Chinese particles are divided into three groups based on their functions. Therefore, we may distinguish the structural particles *jiégòu zhùcí* 結構助詞, aspect particles *dòngtài zhùcí* 動態助詞, modal particles *yǔqì zhùcí* 語氣助詞. Particles such as *de* 的, *de* 地, *de* 得, *suǒ* 所, *děng* 等, *gěi* 给, *shìde* 似的 are considered to be structural particles, whereas the particles *le* 了, *zhe* 着, *guo* 过 belong to aspect particles and the particles *a* 啊, *ma* 吗, *ne* 呢, *ba* 吧, etc., to modal particles. See Liu Yuehua 刘月华 & al. *Shiyong xiandai Hanyu yufa (zengding ben)* 實用現代漢語語法 [Practical Grammar of Contemporary Chinese (revised and enlarged edition)] (Beijing: Shangwu yinshuguan, 2004), 354.

20 Ross and Ma, *Modern Mandarin Chinese Grammar: A Practical Guide*, 19.

21 Luo Anyuan 罗安源, *Jianming xiandai Hanyu yufa* 簡明現代漢語語法 [Concise Grammar of Contemporary Chinese] (Beijing: Zhongyang Minzu Daxue chubanshe, 1999), 160.

<i>zhī</i> 之	1 067,49	<i>děng</i> 等	265,87
<i>de</i> 地	903,41	<i>de</i> 地	239,62
<i>guo</i> 过	126,52	<i>zhī</i> 之	237,06
<i>le</i> 了	86,98	<i>suǒ</i> 所	166,97
<i>de</i> 得	73,14	<i>de</i> 得	149,86
<i>zhèngzài</i> 正在	57,33	<i>guo</i> 过	120,09
<i>zhe</i> 着	23,72	<i>yīyàng</i> 一样	40,07

Table 6

List of particles with the highest level of i.p.m. in SiHanKu and AranSini

Consequently, we also managed to find out that the i.p.m. value of analyzed particles in the legal texts of SiHanKu was different from their value of i.p.m. in the AranSini corpus. The structural particles were considered the most frequent particles both in SiHanKu and AranSini, the presence of aspect particles was much lower, compared to the structural particles.

Sihanku		AranSini	
Particles	i.p.m.	Particles	i.p.m.
<i>de</i> 的	55 677,57	<i>de</i> 的	41 555,10
<i>děng</i> 等	1 674,38	<i>děng</i> 等	265,87
<i>suǒ</i> 所	1 536,00	<i>suǒ</i> 所	166,97
<i>zhī</i> 之	1 067,49	<i>zhī</i> 之	237,06
<i>de</i> 地	903,41	<i>de</i> 地	239,62
<i>guo</i> 过	126,52	<i>guo</i> 过	120,09
<i>le</i> 了	86,98	<i>le</i> 了	1 405,88
<i>de</i> 得	73,14	<i>de</i> 得	149,86
<i>zhèngzài</i> 正在	57,33	<i>zhèngzài</i> 正在	24,03
<i>zhe</i> 着	23,72	<i>zhe</i> 着	346,43

Table 7

Comparison of i.p.m. of the most frequented particles in SiHanKu and in AranSini

To summarize, the most frequently used particles in the register of Chinese legal texts were the structural particles, whose frequency was higher than the frequency of the aspect particles. Modal particles were not present in the legal texts which proves the neutral tone of the register of the Chinese legal texts.

5.4 *Passive Markers in the Register of Chinese Legal Texts*

The most frequently used passive marker is *bèi* 被, which can create passive voice with or without the agent in the sentence. Other passive markers may include *ràng* 让 ('let'), *jiào* 叫 ('call'), *gěi* 给 ('give').²²

In addition to the markers mentioned above, there are also some other words in Chinese language, which can be considered to be the verbs expressing the passive voice, such as *shòu* 受 [suffer, be subjected to], *āi* 挨 [suffer], *zāo* 遭 [suffer, sustain]. These markers are often considered to be the content words with their own meaning and are also automatically considered to be the passive verb forms.²³

Based on the frequency of each marker, we could observe that the lower value of i.p.m. of the markers *gěi* 给, *ràng* 让, *jiào* 叫 compared to the marker *bèi* 被 in the legal texts corresponds with their occurrence in the register of spoken and written language. Low frequency of these markers confirms the legal texts to be part of the written register SMY. On the other hand, the high frequency of the marker *bèi* 被 proves its universal usage in both spoken and written register of Chinese language.

SiHanKu		AranSini	
Passive markers	i.p.m.	Passive markers	i.p.m.
<i>bèi</i> 被	2 206,15	<i>bèi</i> 被	1 243,50
<i>shòu</i> 受	1 312,62	<i>shòu</i> 受	247,36

22 Compared to the marker *bèi* 被, other markers were not fully grammaticalized, therefore the above mentioned markers are all considered to be the content words. Richard Xiao and Tony McEnery, *Corpus-Based Contrastive Studies of English and Chinese* (New York: Routledge, 2010), 87.

23 *Ibid.*, 88.

<i>gěi</i> 给	551,54	<i>gěi</i> 给	901,26
<i>ràng</i> 让	175,94	<i>ràng</i> 让	1 114,91
<i>jiào</i> 叫	1,98	<i>jiào</i> 叫	247,75
<i>zāo</i> 遭	1,98	<i>zāo</i> 遭	25,11
<i>āi</i> 挨	0	<i>āi</i> 挨	14,16

Table 8

Comparison of i.p.m. of the most frequented passive markers in SiHanKu and in AranSini

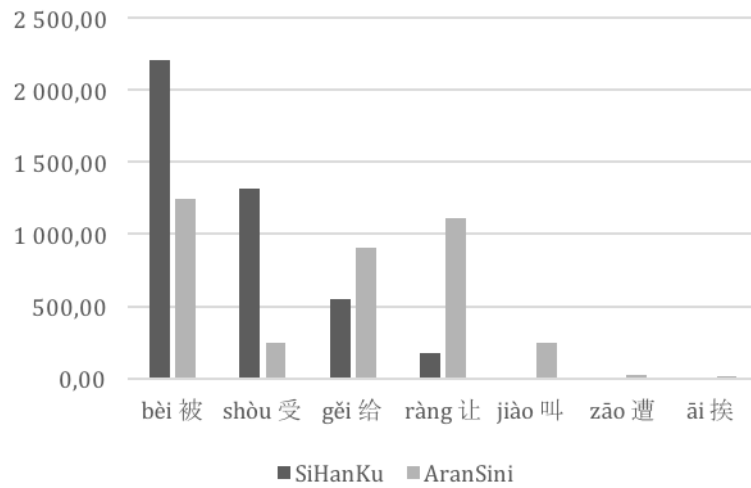


Chart 5

Passive Markers in SiHanKu and AranSini

Based on the obtained values of i.p.m., we can assume that the occurrence of marker *bèi* 被 with the highest value of i.p.m. among all other markers in both corpora, confirmed its universal usage in both registers. By analysing the collocations of the markers we managed to find out that they figured mostly as parts of nominal constructions present in the legal texts.

6 *Conclusions*

By this article we wanted to emphasize the difference between the usage of function words and the passive markers in the register of Chinese legal language and the register of written language, as we decided to use conjunctions, prepositions, particles and the passive markers as the object of our research, which were compared to their frequency of occurrence in the register of legal texts. The difference could be seen in the usage of specific conjunctions, prepositions, particles and passive markers in the register of the legal texts, which might also provide evidence for their specific functions in the register.

The results of our research should be analysed by future research in order to further discern the specifics of the legal texts and its differences from the register of written language. Therefore, we suggest to use the qualitative analysis of the legal texts which could provide us with additional information on this matter, as well as help us further differentiate the register of the legal texts from the register of written language.

Comenius University in Bratislava, Department of East Asian Studies