ISSN: 2688-8653

A Corpus-based Study on the Chinese Near-synonym: zhuānxīn and zhùyì

Xueshan Zhang

Hongkong Polytechnic University, China

zhxs19872016@163.com

Abstract

Comparative study of near synonyms is one of the most prolific areas of Chinese acquisition research. As a corpus incorporates a quite large quantity of data, it has been difficult for researchers to make full use of it. Chinese Word Sketch (CWS), which is powerful and can automatically provide words' grammatical relations, is developed to address this problem. This study aimed to differentiate the Chinese Near-synonym: zhuānxīn *concentrate on* and zhùyì *be aware of* based on the Gigaword Corpus and BLCU Corpus Center (short for BCC[2]) with CWS as an analysis tool. This study found that zhùyì was more frequently used than zhuānxīn and had more complicated grammatical relations. Additionally, zhuānxīn was frequently involved in high cognitive tasks and longer duration while zhùyì indicated low cognitive requirements and short duration. It was also found that the prosody of zhuānxīn tended to be positive and that of zhùyì was neutral. Finally, people can easily control the process of zhùyì while that of zhuānxīn was difficult to control.

Keywords

Chinese Near Synonyms; Chinese Word Sketch; The Gigaword Corpus; BLCU Corpus Center.

1. Introduction

In-depth understanding of synonym differentiation is crucial for second language learners' master of Chinese. Many researchers have noted that comprehensive and in-depth analysis of Chinese synonym words is needed to help language learners use Chinese expressions accurately (Wang and Huang, 2017; Wu & Wang, 2016; Zhao and Li, 2001; Lu, 2016; Li, 2017). Traditionally, researchers conducted comparative studies on Chinese synonyms via manual instances collections and introspection. With the development of corpus linguistics, more studies have drawn on the analysis of a large number of concordances. More recently, Chinese Word Sketch (CWS)[1] is used to automatically extract Chinese words' grammatical relations to explore nuanced Chinese synonyms distinctions (Wang and Huang, 2017). According to Wu and Wang (2016), CWS is a powerful tool that can automatically provide words' frequency, functional distribution, collocation, and synonyms comparison. However, only a limited number of Chinese synonym studies has employed CWS as an analysis tool.

In this case, this study aimed to differentiate the Chinese Near-synonym: 专心zhuānxīn *concentrate on* and 注意zhùyì *be aware of* with CWS. It is worth noting that data in this study was extracted from the Gigaword Corpus, BLCU Corpus Center (short for BCC[2]), Chinses Wordnet (搜文解字sōuwén jiězì) and Contemporary Chinese dictionary (Xiàndài hànyǔ cídiǎn). Moreover, the analysis framework proposed in this study was based on the Chinese synonym comparison frameworks suggested by Lu (2016) and Zhao (2001).

ISSN: 2688-8653

2. Literature Review

Zhao and Li (2001) suggested that synonym discrimination was one of the most critical and challenging topics in learning Chinses as a second language (CSL). Most of the current textbooks and dictionaries only provided rough and abstract meaning explanations which cannot effectively help second language learners of Chinses sort out appropriate words among Near-synonyms. Therefore, they suggest a systematic Chinese Near-synonyms analysis framework to address this problem. They propose that the comparison between Chinese Near-synonyms can be conducted from the semantic, grammatical constraints, and pragmatic perspective. It is worth noting that this theory-based analysis framework provides a solid basis for future related research. However, Lu (2016) pointed out that although Zhao and Li's analysis framework was comprehensive and essential for synonymous comparisons, it had little influence on teaching Chinese as second language due to its heavy dependence on theory.

In order to find more effective means for second language learners to differentiate Chinese synonyms, researchers turned to Corpus-based synonym analysis. In this case, studies on synonym analysis transited from the traditional introspection method to the empirical and statistical method. Lu (2016) suggested that the Corpus-based synonym analysis could not only provide many instances but also information about words' registers, frequency, collocations, and contexts. Therefore, learners can better understand words' usage with meanings and collocations represented via corpus examples. In addition, the researcher proposed more practical and effective Chinese synonyms comparison categories: grammatical positions, collocations, and register distributions. In a similar vein, Li (2017) conducted a Corpus-based Chinese synonyms differentiation study, investigating the similarities and differences between 难过nán guò *sad* and难受 nán shòu *uncomfortable* based on CCL corpus (Center for Chinese Linguistics PKU). The study found that 难过nán guò tended to emphasize the feelings in one's heart and was more prominent in terms of controllability than难受 nán shòu.

Nevertheless, it was tough for researchers to manually analyze words' grammatical distributions and relations from a huge number of data samples. Therefore, recent research employed an automatic and powerful grammatical relation extraction tool: the Sketch Engine (Kilgarriff et al. 2004). Yang (2016), adopting the Sketch Engine as an analysis tool, conducted a Corpus-based (the zhTenTen corpus with 1.7 billion words) analysis of Chinese epistemic modals: 可能kě *could* néng 也许yě xǔ *may*大概dà gài *it is probable* and 恐怕 kǒng pà *afraid of*. The results showed that frequencies and collocation patterns of these words were varied. This could effectively help second language learners learn to use them appropriately.

In 2005, the Chinese version of Sketch Engine was developed by the Academia Sinica group, named Chinese word sketch (CWS), which can automatically extract word collocations based on salience and the *MI log Frequency*. It was developed by attaching the Gigaword Corpus (with 1.12 million Chinese characters) to Sketch Engine (Gong, Ahrens & Huang, 2007). With the development of CWS, the research of Chinese synonym words transited into a new stage. Wu and Wang (2016) conducted a study applying CWS to compare the synonym words 接受 jiēshòu *accept*, and 接收jiēshōu *receive*. Finally, they summarized the similarities and differences between接受 jiēshòu and 接收jiēshōu in terms of co-occurrence words, grammatical functions, abstract or concrete objects, and use scopes. Results showed that 接受 jiēshòu had a close relation with mental activity while 接收jiēshōu was often collocated with objective things. Additionally, they applied the research results in learners' errors analysis and found that their findings can help learners correct most of the errors. Another synonym discrimination analysis adopting CWS as an analysis tool was conducted by Wang and Huang (2018). They employed CWS to differentiate the synonymous emotion words 愉快 yúkuài *pleasant* and 高兴 gāoxìng *happy*. Results showed that 愉快 yúkuài *pleasant* tended to last for a long time while高兴

ISSN: 2688-8653

gāoxìng *happy* was employed to describe the state change. It is worth noting that they demonstrated how to choose information from CWS in lexicographic studies.

It can be concluded that the number of Corpus-based studies on Chinese synonym words via CWS was still limited. This study aimed to address this paucity, extracting data from the Gigaword Corpus and BBC with CWS. Additionally, the researcher drew on a new analysis framework, which is a modification based on frameworks suggested by Lu (2016) and Zhao (2001). It is worth noting that Zhao's framework has been widely recognized as it is systematic and comprehensive. However, with the development of corpus, Lu (2016) suggested another application-oriented one with the basis of Zhao's. Nevertheless, both of them failed to analyze Chinese synonyms from the cognitive aspect. In this case, a more comprehensive analysis framework was proposed in this study.

3. Analysis Framework

Zhao (2001) proposed the following analysis framework for Chinese synonym words, which can be shown as follows:

Table 1. The analysis framework for Chinese synonym words (Zhao, 2001)

The analysis framework for Chinese synonym words (Zhao, 2001)		
Semantic characteristics	Grammatical constraints	pragmatic characteristics
∗ degree ∗ domain (scale; size) ∗ sense (single or multiple) ∗ focus ∗ collocation	∗ part of speech ∗ grammatical relations ∗ mood types of sentences (imperative; declarative; interrogative; exclamatory; bǎzìjù; bèizìjù……)	∗ genre and register ∗ semantic prosody ∗ geographic area

Note: Grammatical relations include grammatical roles of the keywords and other grammatical components having close relations with the keywords. For example, the grammatical relations of verbs include their subject, object, modifiers, modifies, and so on.

Lu (2016) suggested that Zhao's (2001)'s analysis framework was complicated, and results obtained through this framework can hardly be applied in second language teaching. Therefore, another analysis framework based on corpus study was proposed by Lu (2016), which can be shown in the following table 2.

Table 2. The analysis framework for Chinese synonym words (Lu,2016)

The analysis framework for Chinese synonym words (Lu,2016)		
Location of keywords in clauses ∗ simple clause (initial; middle; end) ∗ complex clause (the first clause; the second clause)	Collocation	Register (topic; mode)

Through the comparison of these two frameworks, it can be concluded that they have different analysis focus and classification standards. Moreover, both of them failed to analyze synonyms from their cognitive characteristics. As it has been discussed, the cognitive characteristics of

ISSN: 2688-8653

words can enhance second language learners' understanding of lexical semantics. Therefore, a revised framework was suggested as follows:

Table 3. The analysis framework for Chinese synonym words (verb) in this study

The analysis framework for Chinese synonym words in this study			
Semantic characteristics	Grammatic characteristics	Pragmatic characteristics	Cognitive characteristics
* senses * degree * lexical collocation	* frequency * part of speech * grammatical relations	* semantic prosody * register	* duration * controllability

Based on the suggested analysis framework, the analysis of Chinese synonym words was conducted from four central aspects in this study: semantic, grammatic, pragmatic and cognitive characteristics. Semantic characteristics were unfolded through three aspects: 1. Chinese synonym words could have multiple senses or a single sense, which was one way to differentiate them from each other. 2. The degree involved the intensity of words' meaning, such as the differences between respect and highly respect (Zhao and Li, 2001). 3. More detailed meaning differences of Chinese synonym words could be obtained via the meaning of words collocated with them. The second central aspect involved frequency, which referred to how many times the keywords appear in corpora, part of speech, and grammatical relations, consisting of grammatical roles of keywords and their related grammatical collocations. The third aspect comprised semantic prosody, which could be positive, negative, and neutral, and register, such as literature, scientific articles, news report. The final aspect would be analyzed from three sub-categories: duration, controllability, and effort. The first one is duration, which referred to whether the process of the verb can last for an extended period, short period, or transient. Controllability referred to whether the process can be controlled by people's willpower. The effort meant whether the subjects needed to pay tremendous or little effort to finish the process.

4. Data Collection and Analysis

4.1. Semantic Characteristics 专心zhuānxīn and 注意zhùyì

From different references, sense explanations of 专心 zhuānxīn and 注意 zhùyì can be summarized in the following Table 4.

From Table 4, it can be concluded that 专心zhuānxīn has two senses: 1. concentrate on; 2. united (working together) while 注意zhùyì has one sense: beware or pay attention. Through meaning explanations, it can be inferred that 注意zhùyì entails a more close relation with environmental factors. Individuals have to be aware of something due to external environments rather than based on an individual's willingness. In contrary, 专心 zhuānxīn implies that individuals devote more energies and efforts to one thing out of their own will, which indicates an energetic and positive attitude. Moreover, the antonym of 专心zhuānxīn is 分心fēn xīn *distracted* while the antonym of 注意zhùyì can be 大意dà yì *careless* or 疏忽 shū hu *negligent*. Therefore, the degree of专心zhuānxīn is more profound than that of 注意zhùyì.

The lexical collocation can give further information to understand the meaning of 专心zhuānxīn and注意zhùyì. Table 5 showed the high-frequency nouns with 专心zhuānxīn and注意zhùyì from the BCC. The BCC refers to the Beijing Language and Culture University-Corpus Center

ISSN: 2688-8653

(BLCU Corpus Center, BCC) Corpus, a full-text retrieval corpus with ten billion words, including nine languages, such as Chinese and English.

Table 4. Sense definitions of 专心zhuānxīn and 注意zhùyì

References	专心zhuānxīn concentrate one's attention on	注意zhùyì pay attention to
Contemporary Chinese Dictionary (xiàndài hànyǔ cídiǎn) 5th Edition	ADJ.: focus one's attention on; concentrate on; People should be attentive when they study.	Verb: pay attention to a particular aspect: attention; pay attention to safety; remind someone to pay attention to
On-line Chinese Dictionary http://cidian.aies.cn	Explanation: 1.be single-mined in 2. the single-mindedness 3. united	Explanation: 注意zhùyì refers to the mental activity with orientation and concentration to a particular object or event. It is a common psychological feature accompanied by mental processes such as perception, memory, thinking, and imagination. It has the characteristics of orientation and concentration. Orientation refers to the choice of specific stimuli that appear simultaneously; concentration refers to the inhibition of interfering stimuli. Its generation, scope, and duration depend on the characteristics of external stimuli and human subjectivity.
搜文解字 sōuwén jiězì	Meaning: concentrate one's mind on one thing	Meaning: to choose to pay attention to one or some particular stimuli in the environment.

Table 5. Top-10 high frequency nouns co-occurred with 专心zhuānxīn and注意zhùyì

专心zhuānxīn Collocation verb	Frequency	注意zhùyì Collocation verb	Frequency
专意	36	事项	5375
备考	19	身体	4727
事业	12	点	1054
学术	12	饮食	862
地	11	个人	414
学业	9	营养	275
联赛	9	交通安全	271
向学	8	学生	249
诚意	7	天气	208
做学问	7	义务	207

ISSN: 2688-8653

This table indicated that 专心zhuānxīn frequently co-occurred with processes requiring high cognitive demands, such as备考bèi kǎo *preparing exams*, 学术xué shù *academic research*, 学习 xué xí *study*, and 比赛 bǐ sài *contest*. On contrast, 注意zhùyì cannot be collocated with 备考bèi kǎo (* 注意备考), 学术xué shù (* 注意学术), 学习xué xí (* 注意学习) and 比赛 bǐ sài (* 注意比赛). For注意zhùyì, it tended to co-occur with two main types of events. The first type was related to 健康jiàn kāng *individual's physical health* and personal *daily life*, such as 饮食yǐn shí *diet*, 营养yíng yǎng *nutrition*, 安全ān quán *safety* and the second type of events was related to events which frequently changed, such as *whether*. In contrary, 专心zhuānxīn cannot be collocated with饮食yǐn shí (* 专心饮食), 营养yíng yǎng (* 专心营养), 安全ān quán (* 专心安全).

4.2. Grammatical Characteristics of 专心zhuānxīn and 注意zhùyì

4.2.1. Frequency and POS

In 搜文解字sōuwén jiězì, the frequency of专心zhuānxīn was 125, and the frequency of注意zhùyì was 1157, which was 9.2 times over the frequency of 专心zhuānxīn. In CWS, the frequency and POS (part of speech) of专心 zhuānxīn and 注意zhùyì can be shown in Table 6.

Table 6. The frequency and POS of 专心zhuānxīn and 注意zhùyì

Items	专心zhuānxīn	注意zhùyì
Frequency	3859	136547 (35.38 times over the frequency of 专心zhuānxīn)
POS	VH11 (stative intransitive verb)	VK1(stative verb with a sentential object)

From both 搜文解字sōuwén jiězì and CWS, 注意zhùyì had a much higher frequency than专心zhuānxīn, which can be explained through the semantic characteristics of these two words. The usage of 专心zhuānxīn was confined to high cognitive demand activities, such as study, preparing for exam, participating in the contest. In contradiction, 注意zhùyì can be associated with a wide range of events related to various aspects in personal and social life.

As for the part of speech, 专心zhuānxīn in Gigaword Corpus was stative intransitive verb and 注意zhùyì was a stative verb with a sentential object.

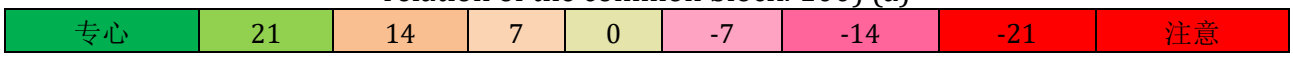
4.2.2. Grammatical Relations

Regarding the grammatical relations, the common patterns and only patterns (represented by Figure 1(a) (b) (c)) provided by CWS presented a vivid and comprehensive comparison between the synonym words: 专心zhuānxīn and 注意zhùyì.

Figure 1 demonstrates the common pattern of 专心zhuānxīn and 注意zhùyì, which means both 专心zhuānxīn and 注意zhùyì can serve as predicates and modifiers. The subjects and modifies in the lists can collocate with both of them. However, the tendencies of some co-occurring words were quite different. Words with the redder background tended to co-occur with 注意zhùyì, while Words with the greener background more frequently co-occured with专心zhuānxīn. More differences can be shown by the only patterns in Table 7 and Table 8.

ISSN: 2688-8653

Figure 1. The common patterns of 专心zhuānxīn and 注意zhùyì provided by the Sketch-Diff function of CWS (minimum frequency: 1 time; maximum number of items in a grammatical relation of the common block: 100) (a)



(b)

Subject	Frequency		Salience	
	专心 zhuānxīn	注意 zhùyì	专心 zhuānxīn	注意 zhùyì
业者 yè zhě residents in a community	1	43	3.0	15.1
我 wǒ I	6	6	13.9	2.8
他tā he	12	28	13.9	3.1
他们 tā men they	2	39	4.8	10.9
孩子hái zi children	3	1	10.7	0.3
捷运局jié yùn jú MRT Bureau	1	6	5.8	10.2
人员 rén yuan staff	1	54	1.6	9.4
小朋友 xiǎo péng yǒu children	2	4	9.0	5.7
政府zhèng fǔ government	2	90	2.4	8.8
她tā she	3	9	8.0	3.7

(c)

Modifies	Frequency		Salience	
	专心 zhuānxīn	注意 zhùyì	专心 zhuānxīn	注意 zhùyì
情况 qíng kuàng situation	1	51	4.0	16.7
态度 tài du attitude	2	2	9.3	1.4
话 huà words	1	8	5.7	8.8
弱点 ruò diǎn weakness	1	1	8.5	3.5
技巧 jì qiǎo skill	1	1	7.3	2.3
事件 shì jiàn event	1	10	4.4	6.0

Notes: CWS does not provide the pinyin and English translation of each word. They were added by the author for the convenience of readers. Moreover, frequency means how many times these collocations appear, and salience is the MI log Frequency, whose calculation is: MI log Frequency: MI – score × log fxy; MI-score: log2 fxyN / fxf y; fx = number of occurrences of word X, fy = number of occurrences of word Y, fxy = number of co-occurrences of words X and Y, (Kilgarriff et al. 2004)

ISSN: 2688-8653

Table 7. The only patterns of 专心zhuānxīn (minimum frequency: 1 time; the maximum number of items in a grammatical relation of the exclusive block: 50)

Subjects of专心 zhuānxīn	Frequency	Salience	Modifies of专心zhuānxīn	Frequency	Salience
科思·派瑞正 personal name	1	14.6	听众 tīng zhòng audience	2	13.5
吴永学 personal name	1	13.2	罗伟 personal name	1	12.41
戈伟托·蒂格劳 personal name	1	12.9	同路人 tóng lù rén fellow traveler	1	10.9
梁左 personal name	1	12.6	说明书 shuō míng shū instruction	1	8.7
罗桂玉 personal name	1	12.5	工作gōng zuò work	2	5.7
史国良 personal name	1	12.4	儿童ér tóng children	1	4.9
乔志温斯顿 personal name	1	11.6	训练 xùn liàn skill	1	4.9
谢玲玲 personal name	1	11.4	学生xué sheng students	1	4.2
朱蒂佛斯特 personal name	1	11.2	环境 huán jìng environment	1	4.0
胡军 personal name	1	10,1	比赛 bǐ sài contest	1	4.94

Note: This table only included the top-ten words shown by the only pattern of 专心zhuānxīn

Table 8. The only patterns of 注意zhùyì (minimum frequency: 1 time; the maximum number of items in a grammatical relation of the exclusive block: 50)

Subjects of注意zhùyì	Frequency	Salience	Modifies of注意zhùyì	Frequency	Salience
船只 chuán zhī ship	2350	97.7	焦点 jiāo diǎn focus	370	62.5
民众 mín zhòng republic	471	38.9	事项 shì xiàng point	212	52.9
驾驶人 jià shǐ rén drivers	50	33.3	现象 xiàn xiàng phenomenon	148	38.3
考生 kǎo shēng exam candidates	55	27.8	问题 wèn tí problem	458	34.6
公平会 gōng píng huì the name of an association	31	27.7	警讯 jǐng xùn news about emergency	23	30.0
家长 jiā zhǎng parents	53	25.5	事 shì event	89	28.0
投资人 tóu zī rén investors	35	22.6	苗头 miáo tóu sign	11	26.3
消费者 xiāo fèi zhě Investors	54	21.2	地方 dì fang place	103	25.5
国人guó rén Chinese people	24	17.4	动向 dòngxiàng tendecy	21	23.9
单位 dān wèi employers	87	16.9	重点 zhòngdiǎn key point	70	23.4

Note:This table only included the top-ten words shown by the only pattern of 注意zhùyì

ISSN: 2688-8653

Through comparison, it can be inferred that the subjects of 专心zhuānxīn tended to be the specific name of individuals. In contradiction, the subjects of 注意zhùyì tended to be a specific group of people or an organization, such as the group of people in ships, the republic, drivers, and parents. Also, 专心zhuānxīn can be used to modify working, training and participating contest and a specific group of people involving in one kind of perception, such as listening. Nevertheless, 注意zhùyì frequently co-occurred with abstract nous which needed further explanations such as events, situation, and change.

The Word sketch function in CSW can further explore grammatical relations. Therefore, the following grammatical relations of 专心zhuānxīn and 注意zhùyì can be obtained:

(1) 专心zhuānxīn subject, modifies

(2) 注意zhùyì: subject, object, modifies, SentObject, SentObject of, PP根据 gēn jù according to, 用 yòng employ, 从 cóng from, 与 yǔ with, 依yī according to, 把 bǎ used to put the object before the subject, 对 duì concerning, 在 zài used to show the time, location.

It was noted that 注意zhùyì has more grammatical relations than 专心 zhuānxīn, which meant 注意zhùyì covered a wider usage domain. For instances, 注意zhùyì can collocate with various prepositional phrases such as注意与+ object zhùyìyǔ+object, 注意依+ object zhùyìyī according to + object. However, it is not meaningful to say * 专心与+ object zhuānxīnyǔ+object and * 专心依+ object zhuānxīnyī according to + object. Moreover, compared with 专心zhuānxīn, 注意zhùyì can serve as an object, SentObject and SentObject of and the detailed information of these grammatical relations can be shown in Table 9.

Table 9. Other detailed grammatical relations of 注意zhùyì (Top-5 high-frequency words)

Object			SentObject			SentObject of		
Collocation words	Frequency	salience	Collocation words	Frequency	salience	Collocation words	Frequency	salience
事项 shì xiàng matter	3101	78.32	防范fáng fàn guard against	2334	85.31	请qǐng please	24334	115.04
动态 dòngtài dynamic change	641	52.06	防止fáng zhǐ prevent	308	48.56	疏于shū yú neglect	164	53.1
今发jīn fā news issued today	68	51.49	处理好chǔ lǐ hǎo deal with	147	47.81	小心xiǎo xīn take care of	139	39.17
自身 zì shēn oneself	574	48.06	避免bì miǎn avoid	362	45.12	希望 xī wàng hope	907	29.59
饮食 yǐnshí diet	428	47.86	行车xíng chē driving	179	45.02	开始kāi shǐ begin	367	26.69

The list of the objects showed that the object of 注意zhùyì with the highest frequency was 事项 shì xiàng *matter* and the rest words were either related to individual health and life or events with frequent change, which had been discussed before. However, it was ungrammatical to say 专心事项zhuānxīn shìxiàng or 专心动态 zhuānxīndòngtài, as专心zhuānxīn was an intransitive verb.

ISSN: 2688-8653

The sentObjects list indicated that 注意zhùyì tended to collocate with guard against, prevent and avoid, and then plus something which may cause disasters or terrible results. Nevertheless, people never say 专心防范zhuānxīn fángfàn or 专心防止zhuānxīn fáng zhǐ. The third column in the table suggests that 请qǐng please often occurred before 注意zhùyì. With further analysis, it was found that the construction of 请qǐng+注意zhùyì+ nouns appeared 181 times in the BCC and the construction of 请qǐng+注意zhùyì+ verbs appeared 238 times in the BCC. With further analysis, it was found that the structure of 请qǐng and注意zhùyì often served as a reminder to people.

However, the frequency of the structure 请qǐng+专心zhuānxīn + nouns was 0, and that of the structure 请qǐng+专心zhuānxīn + verbs was 6. In a nutshell, 请qǐng more frequently collocates with注意zhùyì and the structures of请qǐng+专心zhuānxīn can only collocate with verbs, as专心zhuānxīn was an intransitive verb.

4.3. Pragmatic Characteristics of 专心zhuānxīn and 注意zhùyì

4.3.1. Semantic Prosody

One of the advantages of Corpus-based study is that corpora can provide many instances that can provide a picture of the prosody of the keyword. From the instances provided by Giaword corpus as follows, 专心zhuānxīn has the positive prosody while 注意zhùyì is more neutral.

4.3.2. Register

The register information of 专心zhuānxīn and注意zhùyì can be extracted for the BCC shown as follows.

Table 10. Shows the frequency of 专心zhuānxīn and注意zhùyì in different registers

专心zhuānxīn		注意zhùyì	
register	frequency	register	frequency
文学 wénxué literature	1,847 (19%)	文学 wénxué literature	21,248(6%)
报刊bào kān newspapers and journals	3,887 (40%)	报刊bào kān newspapers and journals	17,7829 (50%)
微博wēi bó	3,156 (32%)	微博wēi bó	69,942(20%)
科技kē jì technology	848 (9%)	科技kē jì technology	85,362(24%)

Both 专心zhuānxīn and 注意zhùyì appears most frequently in the news report. However, with in-depth analysis, it can be found that 专心zhuānxīn in the news report was used to set the people who concentrated on their study and work as examples to be learned by others. In contradiction, 注意zhùyì in the news served as a reminder to make people beware of changes, dangerous situation and things which may cause negative results. 注意zhùyì was least used in literature while 专心zhuānxīn was least in technology.

4.4. Cognitive Characteristics of 专心zhuānxīn and 注意zhùyì

4.4.1. Duration

From the collocations of 专心zhuānxīn and 注意zhùyì, it can be concluded that 专心zhuānxīn tended to be collocated with the process with a long period, such as study and work, while 注意zhùyì needed short period, as it can be modified by 随时suí shí at any time and时时shí shí from time to time.

ISSN: 2688-8653

4.4.2. Controllability

From the BCC corpus, it was found that喜欢 xǐ huan *enjoy* can be collocated with 注意zhùyì, while no instance of the collocation of 喜欢 xǐ huan and 专心zhuānxīn was found. As喜欢 xǐ huan can indicate that people had a habit and the subjects of喜欢 xǐ huan can be controlled easily by people, 注意zhùyì was more controllable than 专心zhuānxīn. Furthermore, the controllability can also be explained by the proposition that专心zhuānxīn required intensified focus, which was not easy to be controlled by people while 注意zhùyì needed a less intensified one, easy to be controlled by people.

5. Conclusion and Pedagogical

Through the analysis, it can be found that 专心zhuānxīn and 注意zhùyì shared some similarities as well as differences. These two verbs indicated activities involving people's attention and the degree of 专心zhuānxīn is intensifier than that of注意zhùyì. 专心zhuānxīn was involved in high cognitive tasks which needed much more effort from people, such as study, work, and contest while 注意zhùyì tended to collocate with topics related to personal health and various aspects in daily life, such as safety, diet. Moreover, 注意zhùyì is more frequently used than 专心zhuānxīn. Although both of them can be predicators and modifier, 注意zhùyì had more complicated grammatical relations, frequently following by prepositional phrases. Moreover, the prosody of专心zhuānxīn was positive, appearing less frequent in technological articles while注意zhùyì was neutral and was least used in literature. Additionally, the duration of 专心zhuānxīn is longer than that of 注意zhùyì. Finally, people can easily control the process of 注意zhùyì while the process of 专心zhuānxīn is difficult to control.

With reference to the HSK[3] dynamic writings of second language learners (Hanyu shuiping ceshi dongtai yuliaoku), it can be found that second language learners misused the collocation * 注意学习zhùyì xuéxí, as 学习xuéxí is a task with high cognitive demand and long period.Based on the previous four-domain analysis and the error analysis, pedagogical implications can be summarized as: First, teachers can emphasize high collocation words of注意zhùyì and专心zhuānxīn, which should be pretty straightforward for students to master. Second, teachers should explain the cognition characteristics of 注意zhùyì and专心zhuānxīn, as they can be understood easily through learners' experience.

Notes: 1. http://wordsketch.ling.sinica.edu.tw/

2. http://bcc.blcu.edu.cn/

3. hsk.blcu.edu.cn

References

[1] Gong, S. P., Ahrens, K., & Huang, C. R. (2007). Chinese Sketch Engine and mapping principles: A Corpus-based study of conceptual metaphors using the building source domain. Paper presented In the Proceedings of the 8th Chinese Lexical Semantics Workshop.

[2] Kilgarriff, A., Pavel, R., Pavel, S., & David, T. (2004). The Sketch Engine. In Geoffrey, W. & Sandra, V., Lorient. (Eds), Proceedings of the 11th EURALEX International Congress (pp.105-116).

[3] Li, H.M. (2017). Hanyu jinyici bianxi: nán guò and nán shòu. [The analysis of Chinese synonym word: nán guò and nán shòu.] The journal of northern literature, 12(06), 96-97.

[4] Lu F.Z. (2016) Corpus based Near-synonym-teaching in TCSL. Journal of Yunnan Normal University, 14(5), 49-56.

[5] Wang, S., & Huang, C. R. (2017). Word sketch lexicography: new perspectives on lexicographic studies of Chinese Near-synonyms. Lingua Sinica, 3(1), 1-22.

ISSN: 2688-8653

[6] Wu, Y., & Wang, S. (2016). Applying Chinese word sketch engine to distinguish commonly confused words. Paper presented In Workshop on Chinese Lexical Semantics. Springer, Cham.

[7] Xun, E.D., Rao, G.Q., Xiao X.Y., Zang.J.J. Dashuju beijingxia BCC yuliaoku yanzhi. [The building up of BCC under the big data background]. Corpus linguistics3(1), 93-118.

[8] Yang, B. (2016). A Corpus-based analysis of Chinese epistemic modals: contrast among kě néng, yě xǔ, dà gài and kǒng pà. The journal of Guangdong foreign language and trade university 7(06), 33-41.

[9] Zhao, X. and, Li, Y. (2001) Synonymous analysis in TCSL. Journal of Jinan University huawen college 2, 16-21.