

A Comparable Corpus-Based Study of Cross-Strait Varieties: The Case of 注意 *Zhùyì* and 專心 *Zhuānxīn*

Jiexiu Liu

School of Foreign Languages, Northeast Petroleum University, Daqing, China

Yingying Chen

School of Foreign Languages, Northeast Petroleum University, Daqing, China

Abstract—Chinese Word Sketch (CWS) provides a tool to identify the semantic distinctions of the Chinese near synonyms in the natural language use situation. This study has the case of 注意 *Zhùyì* and 專心 *Zhuānxīn*, with Gigaword_CNA and Gigaword_XIN, two of the main sub-corpora in CWS, as the corpora for the research. Based on the comparison of senses and part of speech (POS) from Chinese Word Net and dictionaries, it is found that there are some disagreements of the POS between the ones prescribed by dictionary and in actual language use. What is more, some similarities and differences have been revealed based on the common and only patterns in the two sub-corpora. The study aims to reveal the variations of the two keywords in Taiwan Mandarin and Mainland Mandarin, which will contribute to the revision of the dictionary and facilitate a better understanding of cross-strait communication.

Index Terms—Chinese word sketch, corpus, near-synonyms, variations

I. INTRODUCTION

The Chinese language enjoys the status as an official language in mainland China and Taiwan. However, the language shows some degree of variation in these two areas. Some scholars have noticed the differences of Chinese among different regions and claimed they are due to political, cultural, historical, and economic reasons (Huang, Lin & Zhang, 2013). Therefore, lots of research has been conducted to reveal such differences to facilitate a better understanding of cross-strait communication (Diao, 2015; Jiang, 2018; Wang & Huang, 2018). Meanwhile, some studies focus on lexical variations. Zhou (2009) touched on names of people and places based on observation and introspection, and Zhao (2008) explored how one concept was expressed differently. More research has been carried out in recent years based on the corpus. According to Jiang (2018), corpus-based lexical comparisons between Taiwan and the Mainland are mainly about a thesaurus, celebrity coverage in the Chinese press, Chinese news headlines. However, the research on near synonyms based on corpus “has only focused on certain examples” (Wang & Huang, 2017). Hence it is far from sufficient. This study has the case of 注意 *Zhùyì* and 專心 *Zhuānxīn*, with Gigaword_CNA and Gigaword_XIN (two of the main sub-corpora in Chinese Word Sketch, CWS) as the corpora for the study. It is hoped that such a case study will not only reveal the variations of the two keywords but contribute to the revision of the dictionary.

II. RELATED RESEARCH

A. Research on Lexical Varieties between Taiwan Mandarin and Mainland Mandarin

The earlier researchers collected data manually based on observation and introspection. For instance, Ni (1995), Blount (1996) tried to discover the lexical differences between Mainland and Taiwan Mandarin. Zhao (2008) provided examples to show how one concept was expressed differently in the two regions. Moreover, some other research touched on the transliteration of persons and names (Zhou, 2009). More recent research has been carried out based on a corpus, especially with comparable corpora. Some lexical comparisons between Taiwan and Mainland have been conducted with topics ranging from Chinese news headlines (Chin, 2007), judgment terms (Kwong, et al., 2004), and celebrity reports in Chinese press (Tsou, Yuen, Kwong, Lai and Wong, 2005).

B. Research on Near Synonyms

Near-synonyms are words used in dictionaries to define each other for specific senses. “They are almost synonyms, but not quite” (Inkpen, 2004). They are not entirely interchangeable, for they differ in their denotation and connotation or in the senses they emphasize. They may also differ in the collocational patterns and grammatical categories. Some scholars tried to differentiate the near-synonyms from different aspects: Edmonds & Hirst (2002) put forward the

stylistic variations, denotational variations, structural variations, and expressive variations; Inkpen (2004) further divided them into three types, stylistic, denotational, and attitudinal. On top of that, some English near-synonyms have been evaluated, such as *begin* and *start* (Biber et al. 1998), *persist* and *persevere* (Phoocharoensil, 2021). There are many near-synonyms in Chinese, which pose a threat to learning and communication in life. Diao (2012) investigated 而已 *Éryǐ* with a microscopic comparison approach to reveal the different usages of the word in Taiwan and Mainland, on which he put forward the concept of speech community characteristic world. With the help of Word Sketch Engine, there is a more profound and thorough analysis of the semantic distinctions of the Chinese near-synonyms, such as the investigation on the HELP verbs (Gong & Wu, 2012), 鼓勵 *Gǔlì* and 勉勵 *Miǎnlì* (Wang & Huang, 2018). Their research indicates that comparable corpus-based lexical analysis can provide a solid foundation for differentiating the near synonyms in Chinese.

Although the above studies had some groundbreaking findings of the Chinese synonyms, using corpora to discriminate near-synonyms has only focused on limited examples (Wang & Huang, 2017). Moreover, it needs further exploration based on CWS instead of the mere KWIC method in comparing Chinese near-synonyms. Therefore, the research questions will be:

- 1) What are the senses of the two target words based on Chinese WordNet or dictionaries? Do the concordances reveal any differences in CWS?
- 2) What does the collocational analysis reveal about the similarities and differences of grammatical functions of the two words based on CWS?

III. COMPARATIVE SENSES OF THE TWO TARGET WORDS

The two words 注意 *zhùyì* “to pay attention to” and 專心 *zhuānxīn* “to concentrate” are challenging to be identified as synonymous pair, for they are interchangeable only in very limited given contexts in Chinese. However, it is quite interesting to note that “to concentrate on” is used to explain “pay attention to” in Webster’s dictionary.

According to Chinese WordNet (CWN), 注意 *zhùyì* means “集中精神，小心留意” *Jízhōng jīngshén, xiǎoxīn liúyì* (to pay attention to or to concentrate). It can be used both as a transitive VERB and a NOUN. For example,

- (1) 活動中除了注意安全外，請勿干擾上課之班級。

(Chinese Pinyin) *Huódòng zhōng chúle zhùyì ānquán wài, qǐng wù gānrǎo shàngkè zhī bānjí.*

(English translation) In addition to paying attention to safety during the activity, please do not interfere with the class.

- (2) 為便利各館使用，有下列三點注意事項請 館員留意。

Wèi biànlì gè guǎn shǐyòng, yǒu xiàliè sān diǎn zhùyì shìxiàng qǐng guǎnyuán Liúyì.

To facilitate the use of each library, there are the following three precautions for librarians to pay attention to.

When used as a NOUN, it means “特別專注的觀察” *Tèbié zhuānzhù de guānchá* (the focused observation).

- (3) 故宮展覽組組長周功鑫表示，據她觀察，雕刻、珍玩的確最能吸引小孩的注意。

Gùgōng zhǎnlǎn zǔ zǔzhǎng zhōugōngxīn biǎoshì, jù tā guānchá, diāokè, zhēnwán díquè zuì néng xīyǐn xiǎohái de zhùyì.

Zhou Gongxin, Forbidden City Exhibition Group leader, said that according to her observations, sculptures and treasures are indeed the most attractive to children.

Meanwhile, the Modern Chinese Dictionary (7th edition) points out that 注意 *zhùyì* can only be used as a VERB and means “把心思、思想放在某一方面” *Bǎ xīnsī, sīxiǎng fàng zài mǒu yī fāngmiàn* (to put your mind and thoughts on a particular aspect). For example,

- (4) 注意安全

Zhùyì ānquán

Pay attention to safety.

Since the meaning of 專心 *zhuānxīn* can not be searched in CWN, it is searched in *Guóyǔ cídiǎn* by the Ministry of Education of Taiwan, meaning “專一心思，集中心力” *Zhuānyī xīnsī jízhōng xīnlì* (to focus on). It is tagged as VH 11 (stative intransitive verb) in CWS. However, the Modern Chinese Dictionary defines 專心 *zhuānxīn* as an ADJECTIVE, denoting “集中注意力” *Jízhōng zhùyì lì* (to concentrate). For example,

- (5) 学习必须专心。

Xuéxí bìxū zhuānxīn

Study must be attentive.

In terms of sense, the two target words are quite similar in both regions. Nevertheless, the part of speech is different. 注意 *zhùyì* is defined as both a verb and noun in Taiwan Mandarin, while it is only described as a verb in Mainland Mandarin. The word 專心 *zhuānxīn* is used as a verb in Taiwan Mandarin, while it is an adjective in Mainland Mandarin.

IV. COMPARATIVE USAGE IN THE CORPORA

A. Distributional Differences in CNA and XIN

Gigaword_CNA and Gigaword_XIN are sub-corpora of CWS, the former including news texts from Xinhua News Agency of Beijing with 382, 881,000 tokens, and the latter from Central News Agency of Taiwan with 735, 499, 000 tokens (Wang & Huang, 2017). Based on CWS, it shows that 注意 *zhùyì* has a much higher frequency than 專心 *zhuānxīn* in Gigaword2. That indicates people use 注意 *zhùyì* more frequently in their language, so do the two sub-corpora. Moreover, both of the two words are tagged static verbs, 注意 *zhùyì* being a stative verb with a sentential object while 專心 *zhuānxīn* being a stative intransitive verb. (See Table 1).

TABLE 1
Frequency And Pos Of 專心 *Zhuānxīn* And 注意 *Zhùyì*

Node	Freq. in Gigaword2	Freq. in CNA	Freq. in XIN	POS
專心 <i>zhuānxīn</i>	3859	3143	516	VH11
注意 <i>zhùyì</i>	136547	113273	21616	VK1

A thorough analysis of the concordance lines will reveal different information about the part of speech of the two key words. 注意 *zhùyì* can not only be used as a transitive verb but also a noun as Concordance 1 and 2 have shown in Taiwan Mandarin, which is in accordance with the CWN. However, the corpus reveals that 注意 *zhùyì* can be a noun in Mainland Mandarin (see concordance 4 and 5), not only a verb prescribed in the Modern Chinese Dictionary. Moreover, there are some differences in the usage as well. In Taiwan Mandarin, the object of 安全 *Ānquán* (safety) after 注意 *zhùyì* is often omitted, while it is not in Mainland Mandarin.

The concordance examples from CNA

- | | | | |
|----------|--------------|-----------|---------------|
| (1) 0122 | 見解引起了專家學者們的 | <u>注意</u> | 。報導說，黃奇逸長期 |
| (2) 0057 | 增加的囚犯人口，經常引起 | <u>注意</u> | ，不過，自從一九七九年美國 |
| (3) 0035 | 六級，陣風八級，船隻請 | <u>注意</u> | 。審核：預報員：周中央 |

The concordance examples from XIN

- | | | | |
|----------|--------------|-----------|----------------|
| (4) 0117 | 引起了有關專家和學者的 | <u>注意</u> | ，他們認為這一觀點提出 |
| (5) 0059 | 已經引起世界各國經濟界的 | <u>注意</u> | 和興趣。哈巴羅夫斯克現在 |
| (6) 0085 | 她還在信中懇請救援人員 | <u>注意</u> | 自身安全。</p>羅總統離巴 |

As for 專心 *zhuānxīn*, it is used as a stative intransitive verb in concordance 7. However, it is used to modify other words, such as 教學, 開會, 搞好 in the eighth and ninth concordance, so it used as a transitive verb in those sentences. Moreover, it should be listed as an adverb in concordance 10. In the eleventh and twelfth example, it is an adjective as prescribed in the Modern Chinese Dictionary. However, it is evident that it should be listed as an adverb in the contexts of thirteenth and fourteenth examples.

The concordance examples from CNA

- | | | | |
|-----------|--------------|-----------|---------------|
| (7) 0033 | 戰戰爭的壓力，上課不能很 | <u>專心</u> | ，但緊張的程度顯然不若 |
| (8) 0199 | 獲得極大的改善，使老師能 | <u>專心</u> | 安心教學，五千三百多名學生 |
| (9) 0207 | 「可以不受電話干擾」， | <u>專心</u> | 開會、仔細檢討。會 |
| (10) 0119 | 堅持發展社會生產力， | <u>專心</u> | 地搞好現代化建設； |

The concordance examples from XIN

- | | | | |
|-----------|--------------|-----------|-------------|
| (11) 0009 | 國內均屬上乘，平時訓練既 | <u>專心</u> | 刻苦，又能開動腦筋，是 |
| (12) 0033 | ，這使他們在課堂上更加 | <u>專心</u> | ，更加刻苦，而且刺激了 |
| (13) 0059 | 歲開始，兩人又不約而同 | <u>專心</u> | 打羽毛球了。起初貽合 |
| (14) 0085 | 特慕名而來。他上車以後， | <u>專心</u> | 聽廣播，聽著聽著，激動 |

To put it in a nutshell, 注意 *zhùyì* is used as a verb and a noun as prescribed by CWS in Taiwan Mandarin. Nevertheless, it is not only used as a verb listed in the dictionary, but also as a noun as well in Mainland Mandarin, which should be listed in the Modern Chinese Dictionary. In addition, the word 專心 *zhuānxīn* can not only be used as a stative intransitive verb, but transitive verb and adverb as well in Taiwan Mandarin. While it is an adjective as listed in the dictionary, it is also an adverb in Mainland Mandarin.

B. The Common and Only Patterns

Based on “Show Diff” function of CWS, not only the common patterns of the two words but also their exclusive patterns are listed, with the minimum frequency 1. Table 2 and 3 are the common patterns of the two target words in CNA and XIN.

TABLE 2
COMMON PATTERNS IN CNA

專心	21	14	7	0	-7	-14	-21	注意
----	----	----	---	---	----	-----	-----	----

Subject	Frequency		Salience	
	專心 <i>zhuānxīn</i>	注意 <i>zhùyì</i>	專心 <i>zhuānxīn</i>	注意 <i>zhùyì</i>
他 <i>tā</i> he	11	19	14.3	1.1
業者 <i>yèzhě</i> person engaged in some industry or trade	1	43	2.9	14.9
孩子 <i>háizi</i> kid	2	1	9.2	0.8
人員 <i>rényuán</i> staff	1	49	2.0	9.2
捷運局 <i>Jié yùn jú</i> MRT Bureau	1	6	5.7	9.0
他們 <i>tāmen</i> they	2	24	5.7	8.3
Modifies				
情況 <i>Qíngkuàng</i> situation	1	38	4.6	16.7
態度 <i>Tàidù</i> attitude	2	2	9.5	1.4
技巧 <i>Jìqiǎo</i> skill	1	1	7.6	2.5
事件 <i>Shìjiàn</i> event	1	10	4.6	6.3
比賽 <i>Bǐsài</i> game	1	1	5.1	0.0

TABLE 3
COMMON PATTERNS IN XIN

Subject	Frequency		Salience	
	專心 <i>zhuānxīn</i>	注意 <i>zhùyì</i>	專心 <i>zhuānxīn</i>	注意 <i>zhùyì</i>
她 <i>tā</i> she	1	4	4.6	7.4
他 <i>tā</i> he	1	8	3.0	7.1
政府 <i>Zhèngfǔ</i> government	1	9	2.4	6.0
我 <i>wǒ</i> I	1	3	4.5	5.6
球員 <i>Qiúyǎn</i> player	1	1	5.1	2.3
Modifies				
弱點 <i>Ruòdiǎn</i> weakness	1	1	10.6	4.9

Different colors indicate different likeliness of collocation with other words. The words in the green parts strongly collocate with 專心 *zhuānxīn*, while those in the red parts strongly collocate with 注意 *zhùyì*. The words in yellow parts can collocate with both of them. It is interesting that both Taiwan Mandarin and Mainland Mandarin subjects are mainly personal pronouns, people, MRT Bureau or government. Most people will focus on something or be reminded to pay attention to something around them. The two target words are used to describe a kind of psychological state. When the subjects are MRT Bureau or government, metonymy is used in the contexts, for it is people who work in the MRT Bureau or government will concentrate on doing the affairs of the organizations.

Also, in Taiwan Mandarin, when 注意 *zhùyì* is used to modify other words, it has much stronger collocation with 情況 *qíngkuàng* (situation), as shown by the relatively high salience and frequency. 注意的情況 *zhùyì de qíngkuàng* (under the condition of paying attention to something) and its negative pattern 不注意的情況 *bù zhùyì de qíngkuàng* (under the condition of not paying attention to something) are common patterns in the corpus language use. When 專心 *zhuānxīn* is used as a modifier, it has stronger collocation with 態度 *tàidù*, with a much higher salience than 注意 *zhùyì*. That means 專心的態度 *zhuānxīn de tàidù* (focused attitude) is a commoner pattern than 注意的態度 *zhùyì de tàidù* (attitude to pay attention to). If people do things with their total concentration for some time, they are 專心 *zhuānxīn*. If people pay attention to something but do not fully focus on it, the word 注意 *zhùyì* will be used. That means a much longer psychological state in the connotation of 專心 *zhuānxīn* than 注意 *zhùyì*. In mainland Mandarin, only 弱點 *ruòdiǎn* is modified by both of the two target words. There are more collocations when the two target words are used as modifiers in Taiwan Mandarin than mainland Mandarin.

Although 專心 *zhuānxīn* and 注意 *zhùyì* have some similar collocations, they are more different than similar. Table 4 describes the only patterns of 專心 *zhuānxīn* and 注意 *zhùyì* respectively, which indicate that the words listed can only collocate with one of them in Taiwan Mandarin. The subject of 專心 *zhuānxīn* is usually one specific person. One will concentrate his attention when doing something. However, the subject of 注意 *zhùyì* with the highest frequency and salience is 船隻 *chuán zhī*. Taiwan is surrounded by sea, and ship transportation is essential for it. Weather forecast pays close attention to the wind power to warn the danger of wind scale, shown by the concordance examples extracted from the corpus. It is a typical reminder for transportation. Hence 船隻 *chuán zhī* has such high values. Also, most of

the other subjects are collective nouns, instead of individuals as 專心 *zhuānxīn* owns. The subjects cover people, drivers, examinees, consumers. Therefore, the subjects of 注意 *zhùyì* are usually a group of people, while the ones of 專心 *zhuānxīn* is primarily an individual.

TABLE 4
ONLY PATTERNS OF 專心 *ZHUĀNXÍN* AND 注意 *ZHÙYÌ* IN CNA

專心 <i>zhuānxīn</i>				注意 <i>zhùyì</i>			
Subject		Frequency	Salience	Subject		Frequency	Salience
羅貴玉	<i>Luō guìyù</i>	1	12.4	船隻	<i>chuán zhī</i>	ship	2349
史國良	<i>Shǐ guóliáng</i>	1	12.4	民眾	<i>mín zhòng</i>	the people	470
喬治溫斯頓	<i>qiáozhìwēnsīdùn</i>	1	11.4	駕駛人	<i>jiàoshǐrén</i>	person who drive	50
謝玲玲	<i>xiè línglíng</i>	1	11.4	考生	<i>kǎo shēng</i>	examinee	54
茱蒂佛斯特	<i>zhūdífósuītè</i>	1	11.1	公平會	<i>gōng píng huì</i>	Fair Trade Commission	31
Modifies		Frequency	Salience	Modifies		Frequency	Salience
聽眾	<i>tīngzhòng</i>	audience	2	焦點	<i>jiāo diǎn</i>	focus	336
同路人	<i>tónglùrén</i>	fellow traveler	1	事項	<i>shì xīng</i>	matter	158
工作	<i>gōngzuò</i>	work	2	現象	<i>xiàn xiàng</i>	phenomenon	104
兒童	<i>ér tóng</i>	child	1	警訊	<i>jǐng xùn</i>	warning sign	23
環境	<i>huánjìng</i>	surroundings	4.6	重點	<i>zhòng diǎn</i>	focus	65

When 專心 *zhuānxīn* is used to modify other words, half of them are people as well, such as audience, children, and students. These people have to do things wholeheartedly and need more attention for some time. When it is used to modify 工作 *gōngzuò* (work), it is an adverbial of manner. When used to modify 環境 *huánjìng* (surroundings), it is an attribute. While 注意 *zhùyì* is used to modify other words, most of them are not people anymore. Instead, they are things, like focus, matters, phenomena, and place. That means that in Taiwan Mandarin 專心 *zhuānxīn* can be used as adverbial and attribute while 注意 *zhùyì* can only be used as an attribute. In addition, 專心 *zhuānxīn* primarily modify people, yet 注意 *zhùyì* mainly modify focus, matter.

TABLE 5
ONLY PATTERNS OF 專心 *ZHUĀNXÍN* AND 注意 *ZHÙYÌ* IN XIN

專心 <i>zhuānxīn</i>				注意 <i>zhùyì</i>			
Subject		Frequency	Salience	Subject		Frequency	Salience
克思·派瑞正	<i>kèsīpàiruizhèng</i>	1	15.1	北京市	<i>běijīngshì</i>	Beijing	12
吳永學	<i>wúyǒngxué</i>	1	13.7	恭城縣	<i>gōngchéngxiàn</i>	Gongcheng County	2
戈韋托·蒂格勞	<i>gēwéituōdìgéláo</i>	1	13.4	大關村	<i>dàguāncūn</i>	Daguancun	2
梁左	<i>liángzuǒ</i>	1	13.1	人們	<i>rénmen</i>	people	9
密友	<i>mìyǒu</i>	1	11.1	威海市	<i>wēihǎishì</i>	Weihai	3
Modifies		Frequency	Salience	Modifies		Frequency	Salience
羅煒	<i>Luó wēi</i>	audience	1	事項	<i>shì xīng</i>	matter	41
				問題	<i>wèntí</i>	problem	226
				焦點	<i>jiāodiǎn</i>	focus	30
				現象	<i>xiàn xiàng</i>	phenomenon	44
				傾向	<i>qīngxiàng</i>	tendency	17

In mainland Mandarin shown in Table 5, the subject of 專心 *zhuānxīn* is also usually one specific person, as it is in Taiwan Mandarin. Nevertheless, the subject of 注意 *zhùyì* is either government or people, a general reference. It is not so targeted as is Taiwan Mandarin. When used as a modifier, 專心 *zhuānxīn* is not so diverse in collocation as it is in Taiwan Mandarin. While 注意 *zhùyì* is used to modify other words, most of them are things, such as matters, problem, phenomena, which is the same as Taiwan Mandarin.

TABLE 6
SIMILAR WORDS OF 專心 ZHUĀNXĪN AND 注意 ZHùyì

專心 <i>zhuānxīn</i> in CNA				專心 <i>zhuānxīn</i> in XIN			
word		similarity		word		similarity	
坦白	tǎnbái	be frank	0.425	難辦	nán bàn	difficult	0.238
遲疑	chí yí	to hesitate	0.413	傳告	chuán gào	to inform	0.235
不好意思	bù hǎoyìsi	sorry	0.408	不聞不問	bù wén bù wèn	indifferent	0.214
注意 <i>zhùyì</i> in CNA				注意 <i>zhùyì</i> in XIN			
word		similarity		word		similarity	
瞭望	liào wàng	keep a lookout	0.382	有利於	yǒu lì yú	beneficial	0.351
擔心	dānxīn	worry	0.336	注重	zhù zhòng	pay attention to	0.327
開心	kāixīn	happy	0.335	在於	zài yú	lie in	0.319

In all, the two target words are used to describe a kind of mental state. However, there is a much longer state in the connotation of 專心 *zhuānxīn* than 注意 *zhùyì*. Hence, used as a verb, 專心 *zhuānxīn* can be described as “to focus one’s attention on a particular activity or object for a long time”, while 注意 *zhùyì* means “to pay attention to a specific activity”. What is more, some similarities and differences have been revealed based on the common and only patterns in the two sub-corpora. The subjects in both regions are mainly personal pronouns, people, or government. When they are used as a modifier, there is much more variety of collocations in Taiwan Mandarin than mainland Mandarin. Based on the only patterns, it is concluded that the subject of 專心 *zhuānxīn* is usually one specific person in both regions. The subject of 注意 *zhùyì* is usually the specific group of people, such as drivers, examinees, in Taiwan Mandarin. Nevertheless, the subject is a general reference, including either government or people in Mainland Mandarin. When 注意 *zhùyì* are used as a modifier, it is used in the same way in both of the two parts, to modify matter, problem, phenomena.

C. Similar Words from Thesaurus

The maximum number of items is set 60 and the minimum similarity between cluster items is 0.60. The finding results are shown in Table 6, which illustrates the words that are 60 percent similar to 專心 *zhuānxīn* and 注意 *zhùyì* respectively in CNA and XIN. Their similar words are quite different. CWS further provides detailed information of those similar words in pairs. Take 注意 *zhùyì* for example, 瞭望 *liào wàng* to keep a lookout has the highest similarity with it in CNA. They have common patterns in sentence object, modifier, and other distribution of grammatical categories.

V. SUMMARY

Through a comparative study on the two near-synonyms 專心 (*zhuānxīn*) and 注意 (*zhùyì*) based on CNA and XIN of the CWS, it reveals some contextual similarities and differences in Taiwan Mandarin and Mainland Mandarin. Both of 專心 *zhuānxīn* and 注意 *zhùyì* share a related meaning of “to concentrate or concentration”, a kind of mental state, even though the core senses of the two key words in dictionaries seem not closely related with each other. The actual language use in the corpora indicates that there are some disagreements of their POS between dictionary and target language use. What is more, some similarities and differences have been revealed based on the common and only patterns in the two sub-corpora. It is hoped that such a study will help people be aware of the differences between Taiwan Mandarin and Mainland Mandarin in a microscopic way and help to revise the dictionary, especially the Modern Chinese Dictionary, based on the natural language use. However, there are some limitations in the study as well. The two subcorpora may not be large enough to cover all the usages. Secondly, the two subcorpora used in this research are only composed of the local newspaper, not other styles. Future studies may embrace more styles considering differences of words in cross-strait communication.

ACKNOWLEDGEMENTS

This work has been supported by a research grant under “Arts, Sports and Foreign Languages” Cultivation Scheme [Project Number: 2019YTW-W-05] from Northeast Petroleum University, China.

REFERENCES

- [1] Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- [2] Blount, B. G. (1996). The Structure of lexical variation: Meaning, naming, and context. *Cognitive Linguistics Research*, 5, 116-116.
- [3] Chief, L. C., Huang, C. R., Chen, K. J., Tsai, M. C., & Chang, L. L. (2000). What can near synonyms tell us? *International Journal of Computational Linguistics & Chinese Language Processing*, 5(1), 47-59.

- [4] Chin, A. (2007). A sociocultural comparison of Chinese news headline verbs between Hong Kong and Taiwan. In *Proceedings of The First International Workshop on Intercultural Collaboration*, pp. 444-450.
- [5] Diao, Y. B. (2012). Shi lun hai xia liang an yu yan de wei guan dui bi yan jiu [A microscopic comparison of cross-strait varieties: The case of Eryi]. *Journal of Beijing Normal University(Social Science Edition)*, 232 (4), 44-51.
- [6] Diao, Y. B. (2015). *Hai xia liang an ji gang ao di qu xian dai han yu cha yi yu rong he yan jiu* [A study on the difference and integration of modern Chinese across the Taiwan Straits, Hong Kong and Macau]. Commercial Press.
- [7] Diao, Y. B. (2016). Guan yu jin yi bu shen hua liang an si di yu yan dui bi yan jiu de si kao [Further deepening comparative study of Chinese dialects of the four cross-strait areas]. *Journal of Beijing Normal University*, 254 (2), 68-75.
- [8] Edmonds, P. (1996). Evoking meaning by choosing the right words. In *Proceedings of the First Student Conference in Computational Linguistics in Montreal*, pp. 80-87.
- [9] Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics Association for Computational Linguistics*, 28(2), 105-144.
- [10] Gong, S. P., & Wu, P. Y. (2012). Collocation, Semantic Prosody, and Near-synonymy: The HELP verbs in Mandarin Chinese. *International Journal of Computer Processing of Languages*, 24(1), 3-15.
- [11] Huang, C. R. (2007). *Tagged Chinese Gigaword*. Linguistic Data Consortium, Philadelphia.
- [12] Huang, C. R., Ahrens, K., Chang, L. L., Chen, K. J., Liu, M. C., Tsai, M.C. (2000). The module attribute representation of verbal semantics: From semantics to argument structure. *Int. J. Comput. Linguist. Chin. Lang. Process.* 5(1), 1-19.
- [13] Huang, C. R., Hsieh, S. K. & Chen, K. J. (2017). *Mandarin Chinese Words and Parts of Speech*. Routledge.
- [14] Huang, C. R., Kilgariff, A., Wu, Y., Chiu, C. M., Simon, S., Pavel R., Bai M. H., & Chen K. (2005). Chinese sketch engine and the extraction of grammatical collocations. In *Proceedings of the fourth SIGHAN workshop on Chinese language processing*, pp. 48-55.
- [15] Huang, C. R., Lin, J. X., & Zhang, H. R. (2013). World Chinese based on comparable corpus: The case of grammatical variations of jinxing. *Research on Macao language and culture*, 8(3), 397-414.
- [16] Inkpen, D. (2004). *Building a lexical knowledge-base of near-synonym differences*. University of Toronto Doctoral Thesis.
- [17] Jiang, M. (2018). *A comparable corpus-based study of grammatical variation between different variants of Mandarin Chinese*. The Hong Kong Polytechnic University Doctoral Thesis. <https://theses.lib.polyu.edu.hk/handle/200/9603>
- [18] Kwong, O. Y., Tsou, B. K., & Lai, T. B. (2004). Alignment and extraction of bilingual legal terminology from context profiles. *International Journal of Theoretical and Applied Issues in Specialized Communication*, 10(1), 81-99.
- [19] Lv, S. X., Ding, S. S. (2016). *Modern Chinese Dictionary*(7th edition) (Chinese Edition). The Commercial Press.
- [20] Ni, J. (1995). San wen yu yan de yi shu kao cha — hai xia liang an san wen bi jiao [An artistic investigation of prose language: A comparison across the Strait]. *Journal of Huaqiao University (Philosophy and social sciences edition)*. 3(1): 83-90+95.
- [21] Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, 7(1), 240-258. <https://doi.org/10.32601/ejal.911269>
- [22] Tsou, B. K., Yuen, R. W., Kwong, O. Y., Lai, T. B., & Wong, W. L. (2005). Polarity classification of celebrity coverage in the Chinese press. In *Proceedings of the International Conference on Intelligence Analysis* (pp. 2-4). McLean.
- [23] Wang, S., & Huang, C. R. (2017). Word sketch lexicography: new perspectives on lexicographic studies of Chinese near synonyms. *Lingua Sinica*, 3(1), 1-22. <https://doi.org/10.1186/s40655-017-0025-4>
- [24] Wang, X. A., & Huang, C. R. (2018). From near synonyms to power relation variations in communication: A cross-strait comparison of “Guli” and “Mianli”. In *Chinese Lexical Semantics*, (pp. 155-166). Springer International Publishing. https://doi.org/10.1007/978-3-030-04015-4_13
- [25] Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129. <https://doi.org/10.1093/applin/ami045>
- [26] Zhao, C. Y. (2008). A study on the “same word with different meanings” across the Strait — Taking Beijing dialect and Taiwan’s national language as examples. *Chinese Character Culture*, 27(3), 1-3.
- [27] Zhou, H. Q. (2009). Differences between Chinese and Putonghua in Singapore and their countermeasures. In J. Zhou & K. Dong (Eds.), *A study of language and language life across the Taiwan Strait* (pp. 135-156), Hong Kong Commercial Press.

Jiexiu Liu was born in Heilongjiang, China, in 1979. She received her Master’s degree in Linguistics from Harbin Normal University, China, in 2005. She has been the director of seven research projects, published more than ten papers and two textbooks in China.

Yingying Chen, was born in Henan, China, in 1996. She is a master candidate, majoring in Applied Linguistics.