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 已已 *Erǐ* 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), 鼓励 *Gǔlì* and 勉励 *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 *zhù yì* “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 *zhù yì* 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 zhù yì
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 *zhù yì* 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 Guoyu cidian 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) 學習必須專心。
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 *zhù yì* 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 zhùānxīn 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 zhùānxīn being a static intransitive verb. (See Table 1).

| Node     | Freq. in Gigaword2 | Freq. in CNA | Freq. in XIN | POS |
|----------|--------------------|--------------|--------------|-----|
| zhùānxīn |                    |              |              |     |
| zhùyì    |                    |              |              |     |

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 prescribed by CWS. Moreover, there are some differences in the usage as well. In Taiwan Mandarin, the object of 注意 zhùyì is often omitted, while it is not in Mainland Mandarin.

As for 專心 zhùānxīn, 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.

To put it in a nutshell, 注意 zhùyì is used as a verb and as 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 zhùānxīn 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.
Although 專心 zhuānxīn and 注意 zhù yì have some similar collocations, they are more different than similar. Table 4 describes the only patterns of 專心 zhuānxīn and 注意 zhù yì 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 zhù yì 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

---

**Table 2**

**COMMON PATTERNS IN CNA**

| Subject | Frequency | Salience |
|---------|-----------|----------|
| 他 tā | 11 | 14.3 |
| 他 tā | 19 | 1.1 |
| 他 tā he | 1 | 2.9 |
| 某者 yèzhě | 43 | 14.9 |
| 孩子 huá zi | 2 | 9.2 |
| 孩子 huá zi kid | 1 | 0.8 |
| 人 rán yuán | 49 | 9.2 |
| 細胞 jǐ bāo | 1 | 9.0 |
| 我 wǒ | 6 | 9.0 |
| 我 wǒ they | 2 | 8.3 |
| 情 qíng | 1 | 16.7 |
| 表示 biǎishì | 38 | 4.6 |
| 情 qíng kuàng | 1 | 4.6 |
| 劇 qù | 2 | 9.5 |
| 政府 zhèng fǔ | 1 | 1.4 |
| 政府 zhèng fǔ government | 1 | 2.5 |
| 事件 shì jiàn | 10 | 6.3 |
| 事件 shì jiàn event | 1 | 0.0 |

**Table 3**

**COMMON PATTERNS IN XIN**

| Subject | Frequency | Salience |
|---------|-----------|----------|
| 她 tā he | 1 | 4.6 |
| 她 tā he | 1 | 7.4 |
| 政府 zhèng fǔ | 8 | 3.0 |
| 政府 zhèng fǔ government | 9 | 7.1 |
| 我 wǒ | 3 | 2.4 |
| 我 wǒ I | 1 | 6.0 |
| 球員 qiú yuán | 1 | 4.5 |
| 球員 qiú yuán player | 1 | 5.6 |
| 球員 qiú yuán | 1 | 2.3 |

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 zhù yì. 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 zhù yì is used to modify other words, it has much stronger collocation with 情況 qíng kuàng (situation), as shown by the relatively high salience and frequency. 注意的態度 zhù yì de tā idù (attitude of paying attention when doing something) is a commoner pattern than 注意 zhù yì. That means 專心的態度 zhuānxīn de tā idù (focused attitude) is a commoner pattern than 注意的態度 zhù yì de tā idù. In mainland Mandarin, only 弱點 ruò diǎn 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.
the other subjects are collective nouns, instead of individuals as 注意 zhuānxīn owns. The subjects cover people, drivers, examinees, consumers. Therefore, the subjects of 注意 zhuānxīn are usually a group of people, while the ones of 專心 zhuānxīn is primarily an individual.

### Table 4

ONLY PATTERNS OF 專心 zhuānxīn AND 注意 zhuānxīn IN CNA

| Subject       | Frequency | Salience | Subject       | Frequency | Salience |
|---------------|-----------|----------|---------------|-----------|----------|
| 專心 zhuānxīn |           |          | 注意 zhuānxīn |           |          |
| 留貴玉       | 1         | 12.4     | 船隻          | 3249      | 95.4     |
| 史國良       | 1         | 12.4     | 民眾          | 470       | 35.5     |
| 喬治溫斯頓   | 1         | 11.4     | 駕駛人        | 50        | 30.9     |
| 謝玲玲       | 1         | 11.4     | 考生          | 54        | 25.8     |
| 萊蒂佛斯特   | 1         | 11.1     | 公平局        | 31        | 25.6     |

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 huán jìng (surroundings), it is an attribute. While 注意 zhuānxīn 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 zhuānxīn can only be used as an attribute. In addition, 專心 zhuānxīn primarily modify people, yet 注意 zhuānxīn mainly modify focus, matter.

### Table 5

ONLY PATTERNS OF 專心 zhuānxīn AND 注意 zhuānxīn IN XIN

| Subject       | Frequency | Salience | Subject       | Frequency | Salience |
|---------------|-----------|----------|---------------|-----------|----------|
| 專心 zhuānxīn |           |          | 注意 zhuānxīn |           |          |
| 克恩•漢瑞正  | 1         | 15.1     | 北京市        | 12        | 18.0     |
| 吳永學       | 1         | 13.7     | 恭城縣        | 2         | 16.1     |
| 戈偉托•蒂格勞 | 1         | 13.4     | 大關村        | 2         | 15.6     |
| 葉左       | 1         | 13.1     | 人們          | 9         | 14.5     |
| 蕭友       | 1         | 11.1     | 威海市        | 3         | 14.0     |

When 專心 zhuānxīn is also usually one specific person, as it is in Taiwan Mandarin. Nevertheless, the subject of 注意 zhuānxīn 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 zhuānxīn is used to modify other words, most of them are things, such as matters, problem, phenomena, which is the same as Taiwan Mandarin.

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 zhuānxīn 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 zhuānxīn is used to modify other words, most of them are things, such as matters, problem, phenomena, which is the same as Taiwan Mandarin.
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 zhù yì. 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 zhù yì 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 zhù yì 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 zhù yì 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 zhù yì 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 zhù yì for example, 瞭望 liàowàng 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.

| 專心 zhuānxīn in CNA | 注意 zhù yì in CNA | 專心 zhuānxīn in XIN | 注意 zhù yì in XIN |
|---------------------|------------------|---------------------|------------------|
| word                | similarity       | word                | similarity       |
| 坦白 tǎn bái        | 0.425            | 難辦 nán bàn         | difficult 0.238  |
| 遲疑 chí yí         | 0.413            | 傳告 chuán gào       | to inform 0.235  |
| 不好意思 bù hǎo yì | 0.408            | 不聞不問 bù wén bù wèn | indifferent 0.214 |

| 注意 zhù yì in CNA | 注意 zhù yì in XIN |
|-------------------|-------------------|
| word              | similarity        | word              | similarity       |
| 瞭望 liàowàng       | 0.382             | 有利 yǒu lì yú    | beneficial 0.351 |
| 擔心 dānxīn        | 0.336             | 注重 zhù zhòng     | pay attention to 0.327 |
| 開心 kāixīn        | 0.335             | 在於 zài yú       | lie in 0.319    |

| 專心 zhuānxīn in CNA | 注意 zhù yì in CNA | 專心 zhuānxīn in XIN | 注意 zhù yì in XIN |
|---------------------|-------------------|---------------------|-------------------|
| word                | similarity       | word                | similarity       |
| 0.319              |                   | 0.351               |                   |
| 0.351              |                   | 0.382               |                   |
| 0.382              |                   | 0.408               |                   |
| 0.408              |                   | 0.425               |                   |
| 0.425              |                   | 0.436               |                   |
| 0.436              |                   | 0.451               |                   |
| 0.451              |                   | 0.466               |                   |
| 0.466              |                   | 0.481               |                   |
| 0.481              |                   | 0.496               |                   |
| 0.496              |                   | 0.511               |                   |
| 0.511              |                   | 0.526               |                   |
| 0.526              |                   | 0.541               |                   |
| 0.541              |                   | 0.556               |                   |
| 0.556              |                   | 0.571               |                   |
| 0.571              |                   | 0.586               |                   |
| 0.586              |                   | 0.601               |                   |
| 0.601              |                   | 0.616               |                   |
| 0.616              |                   | 0.631               |                   |
| 0.631              |                   | 0.646               |                   |
| 0.646              |                   | 0.661               |                   |
| 0.661              |                   | 0.676               |                   |
| 0.676              |                   | 0.691               |                   |
| 0.691              |                   | 0.706               |                   |
| 0.706              |                   | 0.721               |                   |
| 0.721              |                   | 0.736               |                   |
| 0.736              |                   | 0.751               |                   |
| 0.751              |                   | 0.766               |                   |
| 0.766              |                   | 0.781               |                   |
| 0.781              |                   | 0.796               |                   |
| 0.796              |                   | 0.811               |                   |
| 0.811              |                   | 0.826               |                   |
| 0.826              |                   | 0.841               |                   |
| 0.841              |                   | 0.856               |                   |
| 0.856              |                   | 0.871               |                   |
| 0.871              |                   | 0.886               |                   |
| 0.886              |                   | 0.901               |                   |
| 0.901              |                   | 0.916               |                   |
| 0.916              |                   | 0.931               |                   |
| 0.931              |                   | 0.946               |                   |
| 0.946              |                   | 0.961               |                   |
| 0.961              |                   | 0.976               |                   |
| 0.976              |                   | 0.991               |                   |
| 0.991              |                   | 1.000               |                   |

V. Summary

Through a comparative study on the two near-synonyms 專心(zhuānxīn) and 注意(zhù yì) 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 zhù yì 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.
Jiesiu 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.