

ORIGINAL ARTICLE

Open Access



Word sketch lexicography: new perspectives on lexicographic studies of Chinese near synonyms

Shan Wang^{1*} and Chu-Ren Huang²

* Correspondence:

eduhk2018@gmail.com

¹Department of Chinese Language Studies, The Education University of Hong Kong, Tai Po, Hong Kong
Full list of author information is available at the end of the article

Abstract

Comparative study of near synonyms is one of the most productive research paradigms in Chinese lexicography. Empirical studies to discriminate near synonyms are either introspection-based or corpus-based. Yet, due to the large quantity of data in a corpus, lexicological studies of Chinese rarely make full use of the corpus data. To solve this problem, Kilgarriff's Word Sketch Engine is designed to automatically obtain grammatical and collocational relations of target words from corpora for researchers to further analyze them. Chinese Word Sketch (CWS), a language specific version of Word Sketch Engine, provides a tool to automatically identify grammatical information for Gigaword size corpora. Through a comparative study of the synonymous emotion words 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy', this paper illustrates how CWS can distinguish them and help lexicographers to discriminate their subtle differences. In particular, it focuses on the context where these synonymous words can be used to define each other and context where they should be differentiated. It also discusses how to select information from CWS such that the information represented would be suitable for lexicographic studies. Through the study of near synonyms, we propose that Word Sketch Lexicography will lead the next generation of dictionaries.

Keywords: Word sketch lexicography, Word sketch engine, Chinese word sketch, Near synonyms, Emotion words

1 Introduction

The Chinese language has a large number of synonyms. The teaching and learning of synonyms is difficult but important in language teaching. Synonym discrimination plays an important role in sorting out the nuances of their meanings, using a language accurately, and improving communication and writing abilities. From the perspective of research methods of distinguishing synonyms, the academic community has gone through two important stages: first, the manual data collection phase based on introspection, which mainly relies on the experience of the researchers and second the corpus-based phase, which obtains KWIC (key word in context) from a corpus. At present, the synonym discrimination is entering the third stage, which uses Word Sketch Engine to process the concordance lines from a corpus (Wang and Huang 2013a; Wu and Wang 2016). It obtains the grammatical and collocational relations of the target word, so researchers can further analyze it based on the results.

The major outcome of synonym discrimination is dictionaries. In the past decades, corpora have become popular with Chinese lexicographers. But the research method of discriminating synonyms is still mainly based on introspection or only to some extent using corpora. Although it is widely accepted that corpus-based approaches can help users obtain authentic data, it is difficult to utilize the retrieved results effectively due to the huge number of data and the limited human resources. Compared with the first two methods, the third method of using Word Sketch Engine can classify the corpus data according to the grammatical functions, which can reflect the differences and characteristics between the synonyms through authentic data. It in turn helps researchers quickly and prominently grasp the tendency of how to use the synonyms.

In collaboration with the Word Sketch Engine team, Chinese Word Sketch (CWS)¹ was developed by Academia Sinica. CWS has been proven to be a powerful tool in corpus-based linguistic studies, as illustrated by the long list of research papers supported by this tool (Gong et al. 2008; Huang et al. 2005; Kilgarriff et al. 2005; Wang 2012; Wang and Huang 2011, 2013b; Wang et al. 2012; Wu and Wang 2016). However, its application in the field of lexicography is rarely elaborated. In this paper, we will focus on how it can help distinguish synonyms and facilitate lexicography through a comparative study on two emotion words 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'. Based on the result, we propose that Word Sketch Lexicography will lead the next generation of dictionaries.

2 Related research

2.1 Research on near synonyms

Languages have many synonyms due to various reasons, but just as Cruse (1986: 270) pointed out, “..... natural languages abhor absolute synonyms just as nature abhors a vacuum”. Humans rationally make choices between one form and another to express a particular meaning, and thus there must be reasons why in the context, one word is preferred than another (Allan 2010).

Many scholars have pointed out different dimensions of discriminating near synonyms. For example, Cruse (1986) differentiates them from denotational variations, stylistic variations (including dialect and register), expressive variations (including emotive and attitudinal aspects), and structural variations (including collocational, selectional, and syntactic variations). Inkpen (2004) divides three types of distinctions: stylistic, attitudinal, and denotational. Much research was carried out to discriminate English near synonyms based on a corpus. For example, word frequency: *began* and *started* (Rundell and Stock 1992); collocation: *begin* and *start* (Biber et al. 1998), *of*, *at*, *from*, *between*, *through* (Kennedy 1998), *worry*, and *bother* (Lu and Ahrens 2006); genre: *sure* and *certain* in spoken and written texts (Summers 1993); register: *big*, *large*, and *great* in academic proeses and fictions (Biber et al. 1998).

In Chinese, near synonyms are difficult for learners. Hong and Chen 洪炜, 陈楠 (2013) examined Chinese as a second language (CSL) learners' acquisition of near synonyms. Hong and Zhao 洪炜, 赵新 (2014) examined the difficulty in learning different types of Chinese near synonyms. Zhang 张博 (2016) investigated CSL learners' word confusion and compared the commonality and specificity of confused words of *learners with different mother tongue* based on a large-scale Chinese interlanguage corpus, as

well as the data collected by themselves. Moreover, many near synonyms dictionaries were published to discriminate them in order to help Chinese language learners better use them (Mou and Wang 牟淑媛, 王硕 2004; Teng 鄧守信 2009; Yang and Jia 杨寄洲, 贾永芬 2005; Zhao and Li 赵新, 李英 2009).

Although these studies show the difficulty and importance of synonyms in Chinese learning, making use of corpora to discriminate near synonyms has only focused on certain examples (Chang 張哲璋 2015; Chung 鐘曉芳 2011; Liu et al. 2005; Zhang et al. 张文贤等 2012) and it is far from being a common practice. In addition, most studies only used the KWIC method, which did not propose an effective approach to deal with the large data set.

2.2 Development of corpus technology in lexicography

Collecting sufficient data to compile a dictionary is a common practice in lexicography. Before computer technology was used in lexicography, it took people many years to manually collect data to compile a dictionary. One example was *Oxford Advanced Learners Dictionary* (OALD) which took 50 years to complete.

Lexicography has been greatly influenced by computerized corpora. Instead of spending long time on obtaining suitable information, people can get the information of a word within seconds from an electronic corpus. One common technology is KWIC, from which the contexts that a word occurs in can easily get viewed. The limitation of KWIC is that it cannot efficiently sort out the data from a large-scale corpus. For example, in a corpus containing eight million words, using KWIC to search the word 'deal', there are up to 1500 sentences containing it (Biber et al. 1998). In recent years, the corpora are becoming larger and larger. In an era of big data, the largest challenge of using a corpus in lexicography is how to quickly and effectively deal with the huge retrieval lines available from a corpus.

Compared to the advancement of English lexicography through using corpora, there are limited successful cases in Chinese lexicography (Huang et al. 黄居仁等 1997; Su 苏新春 2006; Yu et al. 俞士汶等 2003). Past survey on Chinese lexicography underlined the conservative nature and lack of adaptation of corpora and other technological innovations (Huang et al. 2016). Even some recent work on how to discriminate near synonyms is mainly based on introspection, referring to a corpus only, rather than making full use of corpus data, such as (Zhao et al. 赵新等 2014).

In the following, we will discuss the shortcomings of the existing research on synonymous emotion words and introduce how to use CWS for an in-depth analysis with the examples of the words 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy', drawing on earlier lexical semantic studies on Chinese emotion words.

2.3 Synonymous emotion words

Emotion words refer to words that can denote human emotions (Pavlenko 2008). They differ from concrete words and abstract words in concreteness, imageability, and context availability (Altarriba and Bauer 2004). They are considered to have advantages over neutral words in the tasks of lexical decision (Kuperman et al. 2014) and memory (Ferré et al. 2015).

In recent years, with the development of network technology and the growing online resources, more and more users express their views in online comments. The analysis

of users' emotional tendencies (or sentiment) can be used to improve the quality of products and services, which have an important commercial value (Lu et al. 路斌等 2007; Xu et al. 徐琳宏等 2007; You et al. 游彬等 2013; Yuen et al. 2004). However, at present, the dominant paradigm of the study on emotion, especially in natural language processing, is only on polarity and does not provide deep insights to our understanding of the emotion content of these words.

In the field of lexical semantics, Chang et al. (2000) illustrated a consistent contrast in seven types of emotion verbs and also proposed a semantic interpretation to the contrast. These emotion verbs are divided into group A and B words. For example, group A words: 高興 *gāoxìng* '(1) glad; happy; cheerful (2) be willing to; be happy to', 開心 *kāixīn* '(1) happy; joyous; elated (2) amuse oneself at sb.'s expense; make fun of sb.', 難過 *nánguò* 'feel sorry; feel bad; be grieved', 痛心 *tòngxīn* 'pained; distressed; grieved', 傷心 *shāngxīn* 'sad; grieved; broken-hearted', 後悔 *hòuhuǐ* 'regret; remorse; repent', 生氣 *shēngqì* 'take offense; get angry', 害怕 *hàipà* 'be afraid; be scared', 擔心 *dānxīn* 'worry; feel anxious', 擔憂 *dānyōu* 'worry; be anxious', 憂心 *yōuxīn* 'a troubled heart'. Group B words: 快樂 *kuàilè* 'happy; joyful; cheerful', 愉快 *yúkuài* 'happy; joyful; cheerful', 喜悅 *xǐyuè* 'happy; joyous', 歡樂 *huānlè* 'happy; joyous; gay', 歡喜 *huānxǐ* 'joyful; happy; delighted', 快活 *kuàihuo* 'happy; merry; cheerful', 痛快 *tòngkuài* '(1) very happy; delighted; joyful (2) to one's heart's content; to one's great satisfaction (3) simple and direct; forthright; straightforward', 痛苦 *tòngkǔ* 'pain; suffering; agony', 沉重 *chénzhòng* 'heavy', 沮喪 *jǔsàng* 'dejected; depressed; dispirited; disheartened', 悲傷 *bēishāng* 'sad; grieved; sorrowful', 遺憾 *yíhàn* 'regret; pity', 憤怒 *fènnù* 'indignation; anger; wrath', 氣憤 *qìfèn* 'indignant; furious', 恐懼 *kǒngjù* 'fear; dread', 畏懼 *wèijù* 'fear; awe; dread', 煩惱 *fánnǎo* 'vexed; worried', 苦惱 *kǔnǎo* 'vexed; worried; distressed; tormented; troubled'. Generally speaking, group A words could be used to indicate inchoative states and thus are mainly used as a predicate and could be used transitively and in imperative and evaluative constructions. On the contrary, group B words can only indicate homogeneous states and thus show higher tendency of nominalization and are used as powerful modifiers in being a nominal modifier or an adjunct. Tsai et al. (1998) discussed the contrast between the synonym pair 高興 *gāoxìng* 'happy' and 快樂 *kuàilè* 'happy; joyful; cheerful' (Table 1), which shows that the syntactic behavior of verbs is semantically determined. They used Sinica Corpus² (Chen et al. 1996) to extract the two words' collocation. M. Liu (2016) identified three major lexicalization patterns for the emotion lexicon: Experiencer-as-subject, Stimulus-as-subject, and Affecter-as-subject. The three patterns highlight three distinct ways of conceptualizing emotions.

Although these studies have provided an insight to emotion words, the data is analyzed either through doing manual annotation or only giving several examples. With a

Table 1 Contrasts between 高興 *gāoxìng* 'happy' and 快樂 *kuàilè* 'happy; joyful; cheerful'

Word	Collocation	Sentential object	-le	Wish sentences	Evaluational sentences	Imperative sentences
高興 <i>gāoxìng</i> 'happy'	280	20 (7.1%)	2 (0.7%)	–	5 (1.8%)	3 (1.1%)
快樂 <i>kuàilè</i> 'happy; joyful; cheerful'	365	–	–	8 (2.2%)	–	–

small corpus, it is possible to manually annotate such data, but when a large corpus is available, it is hard to do it this way. Usually a large corpus is more convincing.

3 Emotions words: a case study of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'

This section selects a pair of near synonyms expressing the positive emotion of happiness, 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'. We first examine their usage in three corpora and five textbooks. Then we investigate how dictionaries discriminate them in order to find out the shortcomings.

3.1 Usage in corpora and textbooks

This section investigates whether 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' are widely used in three Chinese corpora and five textbooks. Table 2 shows their frequency in the Chinese corpora, namely, BCC,³ CCL,⁴ and Sinica Corpus. It is obvious that they are frequently used. It is also important to note that 高興 *gāoxìng* 'happy', as one of the most commonly used Mandarin term for happiness, is used more frequently than 愉快 *yúkuài* 'pleasant', in the ratio of roughly 3 to 1.

In addition to their high frequency in the three corpora, 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' are both first level words in *The Syllabus of Chinese Vocabulary and Characters Levels* (汉语水平词汇与汉字等级大纲) (Examination Center of The National Chinese Proficiency Test Committee 国家汉语水平考试委员会办公室考试中心 2001). In order to see how different usages of the two words are introduced to learners, we also investigated the frequency of them in five sets of popular CSL textbooks as indicated in Table 3. We found that both of them are widely used, with 高興 *gāoxìng* 'happy' more common than 愉快 *yúkuài* 'pleasant'. 高興 *gāoxìng* 'happy' is used about five to eight times of 愉快 *yúkuài* 'pleasant', even higher than the distribution in the corpora. This may due to language learning textbook's reliance and emphasis on more basic words.⁵

It is important to note that the textbooks do not explicitly differentiate the usages of the two near synonyms. However, even with textbook sentences, it is clear that they are not always interchangeable. For example, sentences (1)–(4) show how the two words are used in textbook *The Chinese Course* 汉语教程 (Deng et al. 邓懿等 1992-1993). In some cases, they are interchangeable [such as (1) (2)], while sometimes they are not [such as (3) (4)]. For the two common words, the textbook is lack of specific instructions about the differences between them. It is the same with other textbooks.

Table 2 Frequency of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' in three corpora

Corpora	愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'	Ratio of 高興 <i>gāoxìng</i> 'happy' over 愉快 <i>yúkuài</i> 'pleasant'
BCC	41,499	126,261	3.0
CCL	10,646	38,678	3.6
Sinica	460	1044	2.3
Average	17,535	55,327.7	3.0

Since none of the three corpora provide the accurate total number of words, the percentage of each word used in each corpus cannot be calculated

Table 3 Frequency of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' in different sets of CSL textbooks

Textbooks	愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'	Ratio of 高興 <i>gāoxìng</i> 'happy' over 愉快 <i>yúkuài</i> 'pleasant'
<i>The Chinese Course</i> 汉语教程 (Yang and Ma 杨寄洲, 马树德 1999-2003)	15	70	4.7
<i>The Chinese Course</i> 汉语教程 (Deng et al. 邓懿等 1992-1993)	11	53	4.8
<i>Chinese</i> 中文 (Chinese College of Jinan University 暨南大学华文学院 1997)	5	40	8.0
<i>Boya Chinese</i> 博雅汉语 (Li 李晓琪 2004-2008)	5	26	5.2
<i>Communicative Chinese</i> 交际汉语 (English Channel of CCTV英文央视频道 2003)	2	3	1.5
Average	7.6	38.4	4.8

- (1) 走在上边舒服极了，真让人高兴。

zǒu zài shàngbian shūfu jí le, zhēn ràng rén gāoxìng
walk_on_on-top-of_comfortable_extreme_ASP, really_make
people_happy

Walking on the road is very comfortable; (it) makes people happy.

- (2) 安娜说：“今天我在中国过年，跟在家里过圣诞节一样愉快。”

ānnà shuō : jīntiān wǒ zài zhōngguó guònián, gēn zài jiālǐ guò shèngdànjié yíyàng yúkuài

Anna say today I in China spend the Chinese-New-Year, with at home spend Christmas same pleasant

Anna said: "Today I spend the Chinese New Year in China, which is as pleasant as spending Christmas at home."

- (3) 妈妈和外公知道我学会用筷子了，当然更高兴。

māma hé wàigōng zhīdào wǒ xué huì yòng kuàizi le, dāngrán gēng gāoxìng

Mum and Grandpa know I learn can use chopstick ASP, of-course even happy

Mother and Grandfather knew I had learnt how to use chopsticks. Of course, (they were) happier.

- (4) 我函购了这本图册，工作余闲翻开来看看，老觉得新鲜有味，看一回是一回愉快的享受。

wǒ hángòu le zhè běn túcè, gōngzuò yúxián fānkāilái kànkān, lǎo juéde xīnxiān yǒu wèi, kǎn yī huí shì yī huí yúkuài de xiǎngshòu
I buy through-mailing ASP this CL picture-book, work leisure-time open read, always feel fresh have taste, read one CL be one CL DE enjoyment

I bought this picture book through mailing. I read it during my leisure time and feel it very interesting. Every time it is a happy enjoyment.

Why do the textbook introduce near synonyms without differentiating them? Our speculation is that this is because textbooks rely on dictionaries for sense definition and discrimination. Lexicographic work on Chinese near synonyms, unfortunately, do not adequately underline their usage differences. We will discuss this in more details in the next sections.

3.2 Problems with the sense definition of synonymous words in *Contemporary Chinese Dictionary* 现代汉语词典

The practice of using synonymous words in sense definition is common in lexicography. Although it is quick and easy to supply an equivalent word, it is difficult for learners to discriminate their differences, and thus easily leads to wrong usage. Table 4 shows word senses of some positive emotion words in *Contemporary Chinese Dictionary* 现代汉语词典 (7th Edition) (Dictionary Editing Room of Institute of Linguistics, China Academy of Social Sciences 中国社科院语言研究所词典编辑室 2016). We highlighted the same words with the same color. It is obvious that the different senses in Table 4, 愉快 *yúkuài* 'pleasant', 高興 *gāoxìng* 'happy', 快樂 *kuàilè* 'joyful', 舒暢 *shūchàng* 'ease', and 舒服 *shūfu* 'comfortable' are frequently used definition words. However, many of them are used in a circulatory way.

3.3 Problems with Chinese synonym dictionaries

We examined the explanations of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' in nine synonyms dictionaries (Cheng 程荣 2010; Ding 丁熙翰 1984; Fang 方琳 2000; Ge and Zhang 葛天红, 张福庆, 2000; Mu and Wang 牟淑媛, 王硕 2004; Shi 士彪 2002; Tai 泰芳 1997; Zhong and Yi 仲弋, 艺娜 1999; Zhu 朱景松 2009). It is found that the following problems are common. First, regarding the research method, it is mainly based on lexicographers' introspection and lack of corpus-based data analysis. Second, regarding providing examples, only a few examples are listed, which lacks support from large-scale corpora and it is difficult to reflect the common usage. Third, they cannot accurately provide grammatical function information. Fourth, some explanations are inaccurate, such as 高興 *gāoxìng* 'happy' is for the outer appearance, 愉快 *yúkuài* 'pleasant' is only in the heart. In view of these problems, in the following section, we use CWS to analyze their differences.

4 Use CWS to improve lexicography

There are three challenges of using corpus-based computational approaches in comparing near synonyms. First, researchers can only get large quantity of data of the targeted words in the form of concordance lines. Second, KWIC cannot show the relation between the targeted words. Third, it is hard to generate comparable and meaningful results immediately.

In view of different problems with using corpora directly in linguistic analysis, the Sketch Engine was developed. ".....the Sketch Engine, a corpus tool which takes as input a corpus of any language and a corresponding grammar pattern and which generates word sketches for the words of that language. It also generates a thesaurus and 'sketch differences', which specify similarities and differences between near-synonyms".

Table 4 Word senses and examples of some emotion words in *Contemporary Chinese Dictionary* 现代汉语词典 (7th Edition)

Word	Sense	Example
愉快 <i>yúkuài</i> 'happy; joyful; cheerful'	快意; 舒暢 <i>kuàiyì; shūchàng</i> 'happy; joyful; cheerful'	~地交談 ~ <i>de jiāotán</i> 'talk pleasantly' 心情~ <i>xīnqíng</i> ~ 'in a happy mode; in a cheerful frame of mind' 生活過得很~。 <i>shēnghuó guò dé hěn</i> ~ 'live a happy life.'
高興 <i>gāoxíng</i> '(1) glad; happy; cheerful (2) be willing to; be happy to'	① 形, 愉快而興奮 <i>xíng, yúkuài ér xīngfèn</i> adjective 'glad; happy; cheerful' ② 動, 帶著愉快的情緒去做某件事; 喜歡 <i>dòng, dài zhe yúkuài de qíngxù qù zuò mǒu jiàn shì; xǐ huan</i> verb 'be willing to; be happy to'	聽說你要來, 我們全家都很~。 <i>tīngshuō nǐ yào lái, wǒ men quánjiā dōu hěn</i> ~ 'Our entire family was glad to hear that you were coming.' 他就是~看電影, 對看戲不感興趣。 <i>tā jiùshì ~ kàn diànyǐng, duì kàn xì bù gǎn xìngqù</i> 'He is fond of films but not at all interested in the theatre.'
喜悅 <i>xǐyuè</i> 'happy; joyous'	形, 愉快; 高興 <i>xíng, yúkuài; gāoxíng</i> 'happy; joyous'	~的心情。 ~ <i>de xīnqíng</i> 'happy mode'
舒暢 <i>shūchàng</i> 'happy; entirely free from worry'	形, 開朗愉快; 舒服痛快 <i>xíng, kāilǎng yúkuài; shūfú tòngkuài</i> adjective 'happy; entirely free from worry'	心情~ <i>xīnqíng</i> ~ 'have ease of mind; feel happy' 車窗打開了, 涼爽的風吹進來了, 使人非常~。 <i>chēchuāng dǎ kāi le, liángshuǎng de fēng chuī jìn lái le, shǐ rén fēicháng</i> ~ 'Travelling in a car with its window open, one can enjoy a cool refreshing breeze.'
舒服 <i>shūfu</i> '(1) comfortable;(2) be well'	形 ① 身體或精神上感到輕鬆愉快 <i>xíng ① shēntǐ huò jīngshén shǎng gǎndào qīngsōng yúkuài</i> adjective 'comfortable; pleased; relaxed' ② 能使身體或精神上感到輕鬆愉快 <i>néng shǐ shēntǐ huò jīngshén shǎng gǎndào qīngsōng yúkuài</i> 'comfortable; pleasant'	睡得很~。 <i>shuì de hěn</i> ~ 'have a good sleep' 窑洞又~, 又暖和。 <i>yáodòng yòu ~, yòu nuǎnhuo</i> 'The cave dwelling is warm and comfortable.'
歡喜 <i>huānxǐ</i> 'joyful; happy; delighted'	① 形, 快樂; 高興 <i>xíng, kuàilè; gāoxíng</i> adjective 'joyful; happy; delighted' ② 動, 喜歡; 喜愛 <i>dòng, xǐhuan; xǐài</i> verb 'have a weakness for; be fond of'	滿心~ <i>mǎnxīn</i> ~ 'be filled with joy' 歡歡喜喜過春節 <i>huānhuānxǐxǐ guò chūnjié</i> 'spend a joyful Spring Festival' 她掩藏不住心中的~。 <i>tā yǎncáng bú zhù xīnzhōng de</i> ~ 'She failed to contain the joy in her heart' 他~打乒乓球 <i>tā ~ dá pīngpāngqiú</i> 'He likes to play table tennis.' 他很~這個孩子。 <i>tā hěn zhè ge háizi</i> 'He is very fond of the child.'
歡樂 <i>huānlè</i> 'happy; joyous; gay'	形, 快樂 (多指集體的) <i>xíng, kuàilè (duō zhǐ jí tǐ de)</i> adjective '(usu. of a collective) happy; joyous'	廣場上~的歌聲此起彼伏。 <i>guǎngchǎng shàng-de gēshēng cǐqǐbǐfú</i> 'Merry songs one after another from the square rose.'
開心 <i>kāixīn</i> '(1) happy; joyous; elated (2) amuse oneself at sb.'s expense; make fun of sb.'	① 形, 心情快樂舒暢 <i>xíng, xīnqíng kuàilè shūchàng</i> adjective 'exult; feel happy; rejoice' ② 動, 戲弄別人, 使自己高興 <i>dòng, xì nòng biéren, shǐ zìjǐ gāoxíng</i> verb 'amuse oneself at sb.'s expense; make fun of sb.'	大夥兒住在一起說說笑笑, 十分~。 <i>dàhuǒér zhù zài yìqǐ, shuōshuōxiàxiào, shífēn</i> ~ 'Everybody live together, talking and laughing, feeling very happy.' 別拿他~。 <i>bié ná tā</i> ~ 'Don't make fun of him!'

Table 4 Word senses and examples of some emotion words in *Contemporary Chinese Dictionary* 现代汉语词典 (7th Edition) (Continued)

快活 <i>kuàihuo</i> 'happy; merry; cheerful'	形, 愉快; 快樂 <i>xíng, yúkuài; kuàilè</i> adjective 'happy; merry; cheerful'	提前完成了任務, 心裏覺得很~。 <i>tíqián wánchéng le rènwu, xīn lǐ jué de hěn ~</i> '(I) felt very happy about accomplishing the task ahead of schedule.'
快樂 <i>kuàilè</i> 'happy; joyful; cheerful'	形, 感到幸福或滿意 <i>xíng, gǎndào xìngfú huò mǎnyì</i> adjective 'happy; joyful; cheerful'	~的微笑 ~ <i>de wēixiào</i> 'happy smile' 祝您生日~。 <i>zhù nín shēngrì ~</i> 'Happy birthday to you.'
痛快 <i>tòngkuài</i> '(1) very happy; delighted; joyful (2) to one's heart's content; to one's great satisfaction (3) simple and direct; forthright; straightforward'	形 ①舒暢; 高興 <i>xíng</i> ① <i>shūchàng; gāoxìng</i> adjective 'happy; joyful; delighted; gratified' ②盡興 <i>jìn xìng</i> 'to one's heart's content; to one's great satisfaction' ③爽快; 直率 <i>shuǎngkuai; zhíshuài</i> 'straightforward; frank and direct; forthright'	看見場上一堆一堆的麥子, 心裏真~。 <i>kànjiàn cháng shàng yī duī yī duī de màizi, xīnlǐ zhēn ~</i> 'I was delighted at the sight of stack after stack of wheat.' 這個澡洗得真~ <i>zhè ge zǎo xǐ dé zhēn ~</i> 'I have a very refreshing bath.' 痛痛快快地玩一場。 <i>tòngtongkuàikuài de wán yī chǎng</i> 'have a wonderful time' 隊長~地答應了我們的要求 <i>duìzhǎng ~ de dāying le wǒmen de yāoqiú</i> 'The team leader readily agreed to our request.' 他很~, 說到哪兒做到哪兒。 <i>tā hěn ~, dào nǎ er zuò dào nǎ er</i> 'He is a straightforward man, and does what he says he will.'
快意 <i>kuàiyì</i> 'pleased; satisfied; comfortable'	形, 心情爽快舒適 <i>xíng, xīnqíng shuǎngkuai shūshì</i> adjective 'pleased; comfortable'	微風吹來, 感到十分~。 <i>wēifēng chuī lái, gǎn dào shí fēn ~</i> 'A gentle breeze blowing, (I) felt refreshed.'
爽快 <i>shuǎngkuai</i> '(1) refreshed; comfortable (2) frank; straightforward; outright (3) with alacrity; readily'	形 ①舒適痛快 <i>xíng</i> ① <i>shūshì</i> <i>tòngkuài</i> adjective 'refreshed; comfortable' ②直爽; 直截了當 <i>zhíshuǎng; zhíjiéliǎodàng</i> 'frank; straightforward; outright'	洗個澡, 身上~多了 <i>xǐ gè zǎo, shēn shàng ~ duō le</i> 'feel much refreshed after a bath' 談了這許多話, 心裏倒~了些。 <i>tán le zhè xǔduō huà, xīnlǐ dào ~ le xiē</i> 'I feel relieved after talking things out so much.' 他是個~人。 <i>tā shì gè ~ rén</i> 'He is frank and straightforward.'
舒適 <i>shūshì</i> 'comfortable; cosy; snug'	形, 舒服安逸 <i>xíng, shūfu ānyì</i> adjective 'comfortable; cosy; snug'	環境~ <i>huán jìng ~</i> 'easy circumstances' ~的生活。 <i>~ de shēnghuó</i> 'comfortable life'

The translation in column I is from *A Modern Chinese-English Dictionary* 现代汉英词典 (Unit of Dictionaries, Foreign Language Teaching and Research Press 外语教学与研究出版社辞书部 2001). The translation in column II and III is from *The Contemporary Chinese Dictionary [Chinese-English Edition]* 现代汉语词典(汉英双语版) (Dictionary Editing Room of Institute of Linguistics of China Academy of Social Sciences 中国社会科学院语言研究所词典编辑室 2002). The *pinyin* are added by the authors.

(Kilgarriff et al. 2014; Kilgarriff et al. 2004). Based on it, the language specific version CWS was developed (Huang et al. 2005).

Two corpora, Academia Sinica Balanced Corpus of Modern Chinese (Sinica Corpus) (Chen et al. 1996) and Tagged Chinese Gigaword Corpus (2nd Edition⁶) (Huang 2009), are embedded in CWS. The former is a Mandarin Chinese corpus containing ten million words. The texts in this corpus are collected from different sources, such as philosophy, science, arts, etc. The later contains a total of 1.1 billion characters from Taiwan's Central News Agency, China's Xinhua News agency, and Singapore's Zaobao. Both corpora are word segmented and tagged with Part-of-Speech. When

using CWS, we can either use the two corpora directly or generate a sub-corpus from them according to the years, text types, text sources, and so on. Since Tagged Chinese Gigaword Corpus (2nd Edition) is much larger than Sinica Corpus, in the following part, we chose it to compare the emotion words 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'.

4.1 Use sketch-difference to get COMMON and ONLY patterns

Figure 1 shows the entry form of Word Sketch Differences. We set the *minimum frequency* to 5 times, *maximum number of items in a grammatical relation of the common block* to 100, and *maximum number of items in a grammatical relation of the exclusive block* to 100 as well.

After clicking on the button “Show Diff” (show differences), we will get not only the common patterns of the two words but also their exclusive patterns. This information is crucial to lexicographers.

Table 5 illustrates the *common patterns* of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'. The words listed in this table can collocate with both 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'. Though all these words can collocate with both the two emotion words, their tendency is different. This is indicated from the color chain from green to red. From 愉快 *yúkuài* 'pleasant' to 高興 *gāoxìng* 'happy', the greener a word appears, the more possible it collocates with 愉快 *yúkuài* 'pleasant'. By contrast, the redder a word appears, the more possible it collocates with 高興 *gāoxìng* 'happy'.

CWS automatically extracts collocations based on grammatical patterns that the word participates in (Kilgarriff et al. 2004; Kilgarriff and Tugwell 2002). These collocations are ranked by salience (Kilgarriff et al. 2004; Kilgarriff and Tugwell 2002; Rychlý 2008). Salience is the *MI log Frequency*, which is counted like this:

$$f_x = \text{number of occurrences of word X}$$

$$f_y = \text{number of occurrences of word Y}$$

Corpus: gigaword_alt

First lemma: 愉快

Second lemma: 高興

Sort grammatical relations:

Separate blocks: all in one block common/exclusive blocks

Minimum frequency: 5

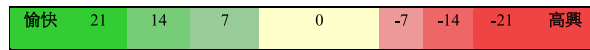
Maximum number of items in a grammatical relation of the common block: 100

Maximum number of items in a grammatical relation of the exclusive block: 100

Show Diff

Fig. 1 Word Sketch Differences entry form

Table 5 Common patterns of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'



Subject*	Frequency		Salience		Modifies	Frequency		Salience	
	愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'	愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'		愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'	愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'
心情 <i>xīnqíng</i> mood	40	7	38.5	13.1	事 <i>shì</i> thing	177	305	38.8	55.4
他 <i>tā</i> he	14	209	9.5	31.2	心情 <i>xīnqíng</i> mood	173	53	52.5	39.4
大家 <i>dàjiā</i> everyone	25	20	25.9	18.5	事情 <i>shìqíng</i> thing	143	107	45.5	47.6
人 <i>rén</i> people	22	54	12.3	14.7	日子 <i>rìzi</i> day	9	33	14.2	33.3
他們 <i>tāmen</i> they	6	23	7.4	13.3	神情 <i>shénqíng</i> look	8	5	19.6	17.3
總統 <i>zǒngtǒng</i> president	9	11	6	3.4	時刻 <i>shíkè</i> moment	7	12	10.5	18.6
					地方 <i>dìfāng</i> place	6	6	3.2	6

*CWS does not provide *pinyin* and English translation of each word. It is the same with other tables. We added them for the convenience of readers

f_{xy} = number of co-occurrences of words X and Y

MI-score: $\log_2 \frac{f_{xy}N}{f_x f_y}$

MI log Frequency: $MI - score \times \log f_{xy}$

Although 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' share many similarities, they are more different than similar. Table 6 depicts 愉快 *yúkuài* 'pleasant' *only patterns* and Table 7 shows 高興 *gāoxìng* 'happy' *only patterns*, which means that the words listed can only collocate with one of the two words.

The *subjects* of the two words indicate that 愉快 *yúkuài* 'pleasant' tend to have appearance and atmosphere as subjects, while 高興 *gāoxìng* 'happy' is apt to have human beings as subjects. The words in the *modifies* relation depict that 愉快 *yúkuài* 'pleasant' is inclined to modify time and atmosphere, while 高興 *gāoxìng* 'happy' tends to modify appearance and information.

Table 6 愉快 *yúkuài* 'pleasant' only patterns

Subject	Frequency	Salience	Modifies (first 10 words)	Frequency	Salience
諸君 <i>zhūjūn</i> everyone	8	33.5	夜晚 <i>yèwǎn</i> night	74	49.5
神情 <i>shénqíng</i> look	14	32.6	氣氛 <i>qìfēn</i> atmosphere	147	44
佳節 <i>jiājié</i> festival	17	30.9	假期 <i>jiàqī</i> vacation	79	40.4
各位 <i>gèwèi</i> everybody	13	30.3	時光 <i>shíguāng</i> time	35	38.2
旅途 <i>lǚtú</i> journey	8	26.4	假日 <i>jiàrì</i> holiday	53	36.7
氣氛 <i>qìfēn</i> atmosphere	14	20.9	佳節 <i>jiājié</i> festival	37	34.6
雙方 <i>shuāngfāng</i> both sides	17	18.6	春節 <i>chūnjié</i> Chinese New Year	72	33.6
精神 <i>jīngshén</i> spirit	13	14.7	回憶 <i>huíyì</i> recall	32	32.2
錢其琛 <i>qián qíchēn</i> Qian Qichen	5	12.8	經驗 <i>jīngyàn</i> experience	93	29.9
工作 <i>gōngzuò</i> work	11	7.4	笑聲 <i>xiàoshēng</i> laughter	14	26.2

Table 7 高興 *gāoxìng* 'happy' only patterns

Subject (first 15 words)			Frequency	Saliency	Modifies		Frequency	Saliency
我	<i>wǒ</i>	I	133	38.8	樣子 <i>yàngzi</i>	appearance	14	29.3
我們	<i>wǒmen</i>	we	92	32.2	眼淚 <i>yǎnlèi</i>	tear	13	28.9
中方	<i>zhōngfāng</i>	China	27	29.5	時候 <i>shíhou</i>	time	29	26.3
她	<i>tā</i>	she	64	26.9	表情 <i>biǎoqing</i>	expression	11	26.2
成就	<i>chéngjiù</i>	achievement	16	17.8	同時 <i>tóngshí</i>	at the same time	40	24.1
內心	<i>nèixīn</i>	heart	7	17.4	大事 <i>dàshì</i>	event	9	17.2
進展	<i>jìnzhǎn</i>	progress	12	14.1	理由 <i>lǐyóu</i>	reason	10	16.3
小朋友	<i>xiǎopéngyǒu</i>	child	6	11.9	話 <i>huà</i>	words	9	13.4
江澤民	<i>jiāng zémín</i>	Jiang Zemin	9	10.2	成果 <i>chéngguǒ</i>	achievement	11	12.3
呂秀蓮	<i>lǚ xiùlián</i>	Lu Hsiu-lien	5	10.1	消息 <i>xiāoxi</i>	news	10	11
老人	<i>lǎorén</i>	old people	7	9.8	原因 <i>yuányīn</i>	reason	7	9.5
夫婦	<i>fūfù</i>	husband and wife	5	9.1	進展 <i>jìnzhǎn</i>	progress	5	8.8
結果	<i>jiēguǒ</i>	result	8	7.9	人 <i>rén</i>	people	19	8.4
決定	<i>juédìng</i>	decision	6	7.5	信息 <i>xìnxī</i>	information	5	7.7

4.2 Use Word Sketch to get relations

The Word Sketch function can help us get the relations a word has and the salient words in a relation. Figure 2 shows the Word Sketch function. Through filling in the blank of Word Form, the relations that a targeted word can occur in will show up.

愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' enter the following grammatical relations respectively:

(a) 愉快 *yúkuài* 'pleasant': subject; modifies

(b) 高興 *gāoxìng* 'happy': subject, object, modifies, modifier, SentObject,⁷

SentObject of,⁸ PP 應 *yīng* 'should', 同 *tóng* 'with', 藉著 *jièzhe* 'make use of', 與 *yú* 'with', 和 *hé* 'used to indicate relationship, comparison, etc.', 從 *cóng* 'from (a time, a place, or a point of view)', 受 *shòu* 'suffer; be subjected to', 在 *zài* 'indicating time, place, scope, etc.', 用 *yòng* 'use; employ; apply', 為 *wéi* 'for the purpose of; for the sake of', 以 *yǐ* 'with; by means of', 由 *yóu* '(done) by sb.; by means of', 把 *bǎ* 'used when the object is placed before the verb, and is the recipient of the action; the sentence structure expresses disposition', 將 *jiāng* 'used to introduce the object before a verb', 向 *xiàng* 'towards; in the direction of', 對 *duì* 'with regard to; concerning; to'.

It is clear that 高興 *gāoxìng* 'happy' has more relation types than 愉快 'pleasant'. For example, 高興 *gāoxìng* 'happy' often occurs with a prepositional phrase, while 愉快 *yúkuài* 'pleasant' does not. These relations show that 愉快 *yúkuài* 'pleasant' collocates strongly with time and occasions, while 高興 *gāoxìng* 'happy' collocates with sentient

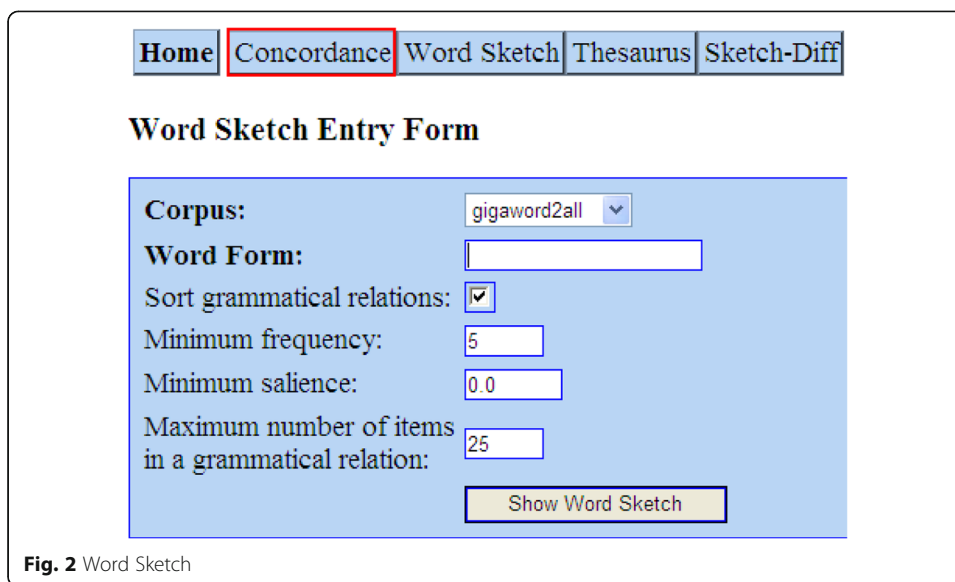


Fig. 2 Word Sketch

agents (typically human). Moreover, 高興 *gāoxìng* 'happy' describes change-of-state events involving sentient agents.

4.3 Providing similar words through *Thesaurus*

Figure 3 shows the Thesaurus entry form. We set *Maximum number of items* to 60 and *Minimum similarity between cluster items* to 0.60. After clicking the button *Show Similar Words*, the results are shown in Table 8.

Table 8 illustrates the words that are 60% similar to 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy', respectively. They can be either a synonym or an opposite. Their similar words are quite different from each other. Table 8 also shows that 愉快 *yúkuài* 'pleasant' patterns with state of happiness, while 高興 *gāoxìng* 'happy' patterns with change-of-state causal events.

4.4 Showing the CWS results in a dictionary

In the above sections, through the assistance of CWS, we rapidly got the contextual information of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'. It includes different word sketch relations, common and exclusive pattern words, and their similar words. It greatly contributes to the clustering of data and thus facilitates lexicographer's work.

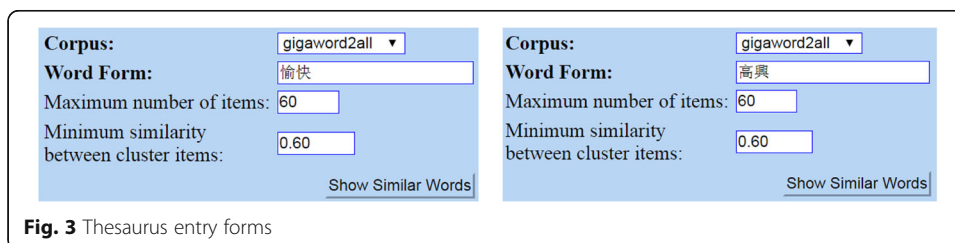


Fig. 3 Thesaurus entry forms

Table 8 Similar words of 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy', respectively

愉快 <i>yúkuài</i> 'pleasant'				高興 <i>gāoxìng</i> 'happy'			
Word			Similarity	Word			Similarity
快樂	<i>kuàilè</i>	happy	0.449	難過	<i>nánguò</i>	sad	0.315
輕鬆	<i>qīngsōng</i>	relaxed	0.397	期待	<i>qīdài</i>	look forward to	0.306
溫馨	<i>wēnxīn</i>	warm	0.335	欣慰	<i>xīnwèi</i>	gratified	0.302
歡樂	<i>huānlè</i>	joy	0.328	盼望	<i>pànwàng</i>	yearn for	0.293
美好	<i>měihǎo</i>	fine	0.283	樂意	<i>lèyì</i>	willing	0.279
愉悅	<i>yúyuè</i>	pleasure	0.281	驚訝	<i>jīngyà</i>	surprised	0.279
興奮	<i>xīngfèn</i>	excited	0.271	害怕	<i>hàipà</i>	be afraid	0.274
痛苦	<i>tòngkǔ</i>	pain	0.268	滿意	<i>mǎnyì</i>	satisfied	0.274
難忘	<i>nánwàng</i>	memorable	0.266	期盼	<i>qīpàn</i>	look forward to	0.264
悲傷	<i>bēishāng</i>	sad	0.248	感激	<i>gǎnjī</i>	grateful	0.246
浪漫	<i>làngmàn</i>	romantic	0.244	擔憂	<i>dānyōu</i>	worry	0.241
喜悅	<i>xǐyuè</i>	joy	0.24	遺憾	<i>yíhàn</i>	regret	0.238
尷尬	<i>gāngà</i>	awkward	0.24	覺得	<i>juéde</i>	feel	0.233
平靜	<i>píngjìng</i>	calm	0.234	理解	<i>lǐjiě</i>	understanding	0.229
幸福	<i>xìngfú</i>	happy	0.233	肯定	<i>kěndìng</i>	affirmative	0.226
祥和	<i>xiánghé</i>	peaceful	0.23	關心	<i>guānxīn</i>	care	0.222
激動	<i>jīdòng</i>	excite	0.228	樂於	<i>lèyú</i>	be happy to	0.221
嚴肅	<i>yánsù</i>	serious	0.226	感謝	<i>gǎnxiè</i>	thank	0.221
平安	<i>píngān</i>	safe	0.219	渴望	<i>kěwàng</i>	desire	0.218
無奈	<i>wú'nài</i>	helpless	0.218	深信	<i>shēnxìn</i>	firmly believe	0.217
親切	<i>qīnqiè</i>	cordial	0.217	在意	<i>zàiyì</i>	care	0.212
開心	<i>kāixīn</i>	happy	0.214	相信	<i>xiāngxìn</i>	believe	0.209
有趣	<i>yǒuqù</i>	interesting	0.214	清楚	<i>qīngchū</i>	clear	0.207
平和	<i>pínghé</i>	mild	0.212	記得	<i>jìde</i>	remember	0.206
緊張	<i>jǐngzhāng</i>	nervous	0.21	慶幸	<i>qìngxìng</i>	rejoice	0.204
悲痛	<i>bēitòng</i>	grieved	0.208	痛心	<i>tòngxīn</i>	distressed	0.204

Nevertheless, what kind of and how much information will be included in a dictionary depends on the type and use of the dictionary. A paper-based dictionary has limited pages and thus has to only cover the most important information; an electronic dictionary is more flexible and therefore can contain more usages. If a dictionary is for elementary learners, it is not good to include too complicated information; if a dictionary is for advanced learners, more valuable usage information will be more helpful.

The results in CWS show that both 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy' can modify 心情 *xīnqíng* 'mood' and 事情 *shìqíng* 'thing' type words. But 愉快 *yúkuài* 'pleasant' can also modify 氣氛 *qìfēn* 'atmosphere', 假日 *jiàrì* 'holiday', 假期 *jiàqī* 'vacation', and 神情 *shénqíng* 'look', while 高興 *gāoxìng* 'happy' modifies 消息 *xiāoxi* 'news' and 原因 *yuányīn* 'reason'. Moreover, syntactically both words can be attributive, resultative, and predictive. But 高興 *gāoxìng* 'happy' also tends to collocate with prepositional phrases. Suppose we are compiling a dictionary for

advanced-level Chinese learners, we can provide the following information regarding the comparison between 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'.

愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy'

【同】【Similarities】

二者都可以表示心情快樂。Both can express a happy mood.

語義方面:它們的主語常是表示心情或人的詞語;它們常用於修飾表示事情、心情、時光等的詞語。*Semantics*: Their subjects are often words that represent feelings or people; both of them are often used to modify words that express things, mood, time, and so on.

[例句] [Examples]

	語義 semantics	愉快 <i>yúkuài</i> 'pleasant'	高興 <i>gāoxìng</i> 'happy'
愉快、高興的主語 The subjects of 愉快 <i>yúkuài</i> 'pleasant' and 高興 <i>gāoxìng</i> 'happy'	表示心情的詞語作主語 the words that represent mood as subjects	目前正在維吉尼亞州結束巡迴演出的喬治溫斯頓今天心情顯然相當愉快。 mùqián_zhèngzài_wéijīniyà_zhōu_jiéshù_xúnhuì_yǎnchū_de_qiáozhìwēnsīdùn_jīntiān_xīnqíng_xiǎnrán_xiāngdāng_yúkuài currently_in-course-of_Virginia_state_end_concert-tour_George-Winston_today_mood_clearly_quite_pleasant <i>George Winston, who is currently ending a concert tour in Virginia, is clearly quite pleasant today.</i>	旅居里約熱內盧40多年的老華僑季福仁聽到北京申奧成功的消息後心情十分高興。 lǚjū_lǐyuèrènnèilú_40_duōnián_de_lǎo_huáqiáo_jìfúrén_tīngdào_běijīng_shēn'ào_chénggōng_de_xiǎoxi_hòu_xīnqíng_shífēn_gāoxìng living-overseas_Rio-de-Janeiro_40_more_year_DE_old_overseas-Chinese_jìfuren_heard_Beijing_bid_for-the-Olympic-Games_success_news_after_mood_utterly_happy <i>The old overseas Chinese Furen Ji, who has lived in Rio de Janeiro for more than 40 years, is very happy on hearing that Beijing successfully bid for the Olympic Games.</i>
	表示人的詞語作主語 the words that refer to people as subjects	他們愉快地回顧了訪華的情況。 tāmen_yúkuài_de_huīgù_le_fānghuá_de_qíngkuàng they_pleasantly_DE_review ASP_visit-China_DE_situation <i>They pleasantly reviewed the situation of visiting China.</i>	如果工黨能有這樣一個結果,他們自然會很高興。 rúguǒ_gōngdǎng_néng_yǒu_zhèyàng_yí_gè_jiēguǒ,tāmen_zìrán_huì_hěn_gāoxìng if_the-Labor-Party_can_have_like-this_one_CL_result,they_naturally_can_very_happy <i>If the Labor Party has such a result, they will naturally be happy.</i>
愉快、高興修飾的詞語 words that 愉快 <i>yúkuài</i> 'pleasant' and 高興 <i>gāoxìng</i> 'happy' modify	表示事情、心情、時光等的詞語 words that expresses things, feelings, time, etc	好友相聚是人生最愉快的事。 hǎoyǒu_xiāngjù_shì_rénshēng_zuì_yúkuài_de_shì good-friends_getting-together_be_life_most_pleasant_DE_thing <i>Good friends getting together is the most pleasant thing in life.</i>	電影年的活動進行一半,已聽到兩件值得高興的事。 diànyǐngnián_de_huódòng_jìnxíng_yībàn,yǐ_tīngdào_liǎng_jiàn_zhíde_gāoxìng_de_shì the-Year-of-the-Film_DE_activity_carry-on_half,already_hear_two_CL_worth_happy_DE_thing <i>When activities on the Year of the Film were carried out in half, (we) have heard two things worthy of pleasure.</i>

(Continued)

	我們現在以充滿愉快的心情離開漢城,因為我們已成功地完成會談。 wǒmen xiànzài yǐ chōngmǎn yúkuài de xīnqíng líkāi hànchéng, yīnwèi wǒmen yǐ chénggōng de wánchéng huìtán we now in be-filled-with pleasure DE mood leave Seoul, because we already successfully DE complete talk We are now leaving Seoul in a pleasant mood, because we have successfully completed the talks.	同學們難掩高興的心情。 tóngxué nányǎn gāoxìng de xīnqíng classmates cannot-conceal happy DE mood The classmates cannot conceal their happiness.
--	---	---

語法方面:都可以做定語、謂語、補語。 Grammar: They can both act as attributives, predicates, and complements.

[例句] [Examples]

句法成分 syntactic function	愉快 yúkuài 'pleasant'	高興 gāoxìng 'happy'
定語 Attribute	有愉快的心情,才有工作的朝氣。 yǒu yúkuài de xīnqíng, cái yǒu gōngzuò de zhāoqì 。 have pleasant DE mood, just have work DE vitality Having a pleasant mood, (you) will have the vitality for work.	他也以最誠懇、最高興的心情,祝福畢業同學鵬程萬裏,一帆風順。 tā yě yǐ zuì chéngkěn, zuì gāoxìng de xīnqíng, zhùfú biyè tóngxué péngchéngwànli , yīfānfēngshùn he also in most sincere, most happy DE mood, wish graduate classmate (lit. A roc can reach a destination of a myriad miles away at one jump) have a bright future, (lit. have a favorable wind throughout the voyage) Everything is going smoothly With the most sincere and happiest mood, he also wishes the graduates have a bright future and everything goes smoothly.
謂語 Predicate	閩獅漁號未涉案船員的返鄉歸期近了,船員們心情都很愉快。 mǐnshīyúhào wèi shè'àn chuányuán de fǎnxiāng guīqī jìn le, chuányuánmen xīnqíng dōu hěn yúkuài the-ship-Min-Lion-Fishing not involve-in-the-legal-case crew DE return-home return-date close ASP, crew mood all very pleasant The return home date of the crew of the ship Min Lion Fishing who are not involved in the legal case is close, and thus they are very pleasant.	有一百多位未婚男子爭搶綉球,搶到綉球的人很高興。 yǒu yībǎi duō wèi wèihūn nánzǐ zhēngqiǎng xiùqiú, qiǎngdào xiùqiú de rén hěn gāoxìng have one-hundred more CL unmarried-men compete-for colorful-silk-ball, grab colorful-silk-ball DE people very happy There are more than a hundred unmarried men competing for one colorful silk ball, and thus the man who can grab it is very happy.
補語 Complement	聊得很愉快 liáo de hěn yúkuài chat DE very pleasant chat pleasantly	說得很高興 shuō de hěn gāoxìng talk DE very happy talk happily

[常見搭配] [Common collocations]

高兴的/高興的 心情/事情/日子/神情/時刻

yúkuài de / gāoxìng de xīnqíng / shìqíng / rìzi / shénqíng / shíkè
pleasant / happy mood / thing / day / look / moment

【异】【Differences】

語義方面：“愉快”常與表示時光和氣氛的詞語搭配，狀態的持續時間較長；“高興”常用於描述施事的狀態變化。 *Semantics*: 愉快 yúkuài 'pleasant' often collocates with words that express time and atmosphere; the state can last for certain time. 高興 gāoxìng 'happy' is often used to describe the state of change.

[例句] [Sentences]

愉快 yúkuài 'pleasant'	高興 gāoxìng 'happy'
<p>預祝大家新春快樂，有一個充實愉快的春節假期。 yùzhù dàjiā xīnchūn kuàilè, yǒu yí gè chōngshí yúkuài de chūnjié jiàqī congratulate-beforehand everyone New-Year happy, have one CL full-of fruitful pleasant DE Spring holiday <i>I wish everyone a happy New Year and a fruitful and happy Spring Festival holiday.</i></p> <p>晚宴在愉快的氣氛中持續至午夜才結束。 wǎnyàn zài yúkuài de qìfēn zhōng chíxù zhì wǔyè cái jiéshù dinner in pleasant DE atmosphere in- the-process of continue till midnight end <i>The dinner ended in a pleasant atmosphere until midnight.</i></p>	<p>全體巴西人民都與羅納爾多、斯科拉裡和其他球員一起流下了激動和高興的眼淚。 quántǐ bāxī rénmin dòu yú luónàěrdūo, sīkēlǎi hé qítā qiúyuán yìqǐ liúxià le jīdòng hé gāoxìng de yǎnlèi all Brazil people all with Ronaldo, Scolari and other ball-player together fall ASP excite and happy DE tear <i>All the Brazilian people shed the excite and happy tears with Ronaldo, Scolari and other players.</i></p> <p>看到顧客高興的樣子，她也感到一種滿足。 kāndào gùkè gāoxìng de yàngzi, tā yě gǎndào yí zhǒng mǎnzú see customer happy DE look, she also feel one kind satisfaction <i>Seeing the customers' happy look, she also felt a kind of satisfaction.</i></p>

語法方面：“高興”後面常接介詞，如在、與、和等。 *Grammar*: 高興 gāoxìng 'happy' is often followed by prepositions, such as 在 zài 'in', 與 yú 'and', and 和 hé 'and'.

[例句] [Examples]

今天是大喜日子，很高興在北京見到您。

jīntiān shì dàxǐ rìzi, hěn gāoxìng zài běijīng jiàndào nín
today be great-job day, very happy in Beijing see you
Today is a day of great joy and I am glad to meet you in Beijing.

他相當高興與大家共度溫馨的時光。

tā xiāngdāng gāoxìng yǔ dàjiā gòngdù wēnxīn de shíguāng
he extremely happy with everyone spend-together warm DE time
He was so happy to spend the warm time with everyone.

董建華說，很高興和大家一起，見證香港的專業人士參與廣東省這項重大文化設施的建設。

dǒngjiànhuá shuō, hěn gāoxìng hé
dàjiā yìqǐ, jiànzhèng xiānggǎng de zhuānyè rénsì cānyù guǎngdōng
shèng zhè xiàng zhòngdà wénhuà shèshī de jiànshè
Dong Jianhua said, very happy and everyone together, witness Hong Kong's professional people participate in Guangdong's this major cultural facilities construction.

Tung-Chee-Hwa say, very **happy** with everyone together, witness Hong Kong DE professional people participate in Guangdong province this CL major culture facility DE construction
Tung Chee-Hwa said he was pleased to join us in witnessing the participation of Hong Kong professionals in the construction of this major cultural facility in Guangdong Province.

“高興”常用于使動結構,如令、讓、使。

高興 *gāoxìng* 'happy' is often used in a causative construction, such as 令 *líng* 'make', 讓 *ràng* 'let', 使 *shǐ* 'make'.

[例句] [Examples]

這次邀請賽有六國參加,令人高興。

zhè cì yāoqǐngsài yǒu liù guó cānjiā, lìng rén gāoxìng
 this CL tournament have six country participate, make people happy
It is a pleasure to have six countries to participate in this tournament.

沒有比在自己國家贏得冠軍更讓人高興的。

méiyǒu bǐ zài zìjǐ guójiā yíngdé guànjūn gēng ràng rén gāoxìng de
 no compare in one
 self country win champion even make people happy DE
There is no more happy than winning a champion in one's own country.

同胞如此熱誠地歡迎我們,使我非常高興。

tóngbāo rúcǐ rèchéng de huānyíng wǒmen, shǐ wǒ fēicháng gāoxìng
 compatriot like-this sincerely DE welcome us, make I feel **happy**
My compatriots welcome us so warmly, which makes me very happy.

[常見搭配] [Common collocations]

愉快的夜晚/氣氛/假期/時光/假日/佳節/春節/回憶/經驗/笑聲/節日/笑容
yúkuài de yèwǎn / qìfēn / jiàqī / shíguāng / jiàrì / jiājié / chūnjié / huíyì / jīngyàn / xiàoshēng / jiérì / xiàoróng
pleasant night / atmosphere / holiday / time / holiday / holiday / Chinese New Year / memory / experience / laughter / holiday / smile

高興的樣子/眼淚/時候/理由/事兒/泪水/話/口吻/模樣

gāoxìng de yàngzi / yǎnlèi / shíhou / lǐyóu / shìer / lèishuǐ / huà / kǒuwěn / múyàng
happy look / tears / time / reason / thing / tears / words / tone / look

【用法相近的詞】 【words that have similar usage】

[愉快 *yúkuài* 'pleasant'] 快樂 *kuàilè* 'happy; joyful; cheerful', 輕鬆 *qīngsōng* 'relaxed', 溫馨 *wēnxīn* 'warm', 歡樂 *huānlè* 'happy; joyous; gay', 美好 *měihǎo* 'fine; happy; glorious', 愉悅 *yúyuè* 'joyful; cheerful; delighted'

[高興 *gāoxìng* 'happy'] 難過 *nánguò* 'sad'

The above analysis shows that CWS can gather the relevant information of the targeted words, which greatly changes the data dispersion problem of only getting the

concordance lines by using a corpus alone. It can greatly facilitate researchers to do analysis. Therefore, we predicate that Word Sketch Lexicography will lead the direction of next generation of dictionaries.

5 Conclusions

Synonyms are difficult for language learners to distinguish. In the past decades, although the idea of using corpora in Chinese lexicography is widely accepted, KWIC alone has many limitations. Generalization and definitions in Chinese lexicography are typically still created without making full use of a corpus. With the arrival of the big data era, to make better use of large-scale data, the analysis with the corpus query tool Word Sketch Engine has become a necessity. CWS, as a language specific version of Word Sketch Engine, can facilitate researchers to get the most salient contextual information of Chinese words.

Through a comparative study on the synonymous words 愉快 *yúkuài* 'pleasant' and 高興 *gāoxìng* 'happy', this paper has illustrated how the CWS functions show their contextual information and facilitate lexicographers to discriminate their subtle differences. In particular, this paper has focused on the contexts where the synonymous words can both be used and contexts where they should be differentiated. It also discusses how to select information from CWS such that the information represented would be suitable for the targeted dictionary. Word Sketch Lexicography has more advantages than using a corpus alone, and thus this research predicts that it will lead the development of dictionaries of the next generation.

6 Endnotes

¹<http://wordsketch.ling.sinica.edu.tw/>

²<http://asbc.iis.sinica.edu.tw/>

³<http://bcc.blcu.edu.cn/>

⁴http://ccl.pku.edu.cn:8080/ccl_corpus/

⁵The one exception (1.5 times) has much shorter texts and smaller samples. Hence the lack of contrast may be due to the small samples, and thus it can be ignored.

⁶<http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2005T14>

⁷A sentient object follows 高興 *gāoxìng* 'happy'. For example,

我真的很高興看到這部書的順利出版。

wǒ zhēnde hěn gāoxìng kāndào zhè bù shū de shùnlì chūbǎn

I really very happy see this CL book DE smooth publication

I am really glad to see the smooth publication of this book.

⁸高興 *gāoxìng* 'happy' is the object of a sentient object. For example,

他對達成這筆交易感到高興。

tā duì dáchéng zhè bǐ jiāoyì gǎndào gāoxìng

he to achieve this CL deal feel happy

He felt happy with the deal.

Acknowledgements

We would like to thank the anonymous reviewers for their useful suggestions. An earlier version was presented at The 8th ASIALEX International Conference. This work is supported by Internal Research Grant of The Education University of Hong Kong (Project No.: 15214, Activity Code: R3733, Reference Number: RG 92/2015-2016) and General Research Fund (GRF) of the Research Grants Council of Hong Kong (Project no. 543810).

Authors' contributions

Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Chinese Language Studies, The Education University of Hong Kong, Tai Po, Hong Kong. ²Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.

Received: 7 April 2016 Accepted: 28 July 2017

Published online: 10 November 2017

References

- Allan, Keith, ed. 2010. *Concise encyclopedia of semantics*. Amsterdam: Elsevier.
- Altarriba, Jeanette, and Lisa M Bauer. 2004. The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology* 117(3):389–410.
- Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Chang, Zhe-Wei 張哲偉. 2015. A corpus-based lexical semantic study of Mandarin verbs of hanging: On the near synonym set: guà, xuán, diào 以語料庫為本之中文動詞近義詞「挂、懸、吊」之詞彙語意研究. MA dissertation. Hsinchu: National Chiao Tung University.
- Chang, Li-li, Keh-Jiann Chen, and Chu-Ren Huang. 2000. Alternation across semantic fields: A study on Mandarin verbs of emotion. *Computational Linguistics and Chinese Language Processing* 5(1):61–80.
- Chen, Keh-Jiann, Chu-Ren Huang, Li-Ping Chang, and Hui-Li Hsu. 1996. Sinica corpus: Design methodology for balanced corpora. In *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation (PACLIC-11)*, ed. Byung-Soo Park and Jong-Bok Kim, vol. 167, 167–176. Seoul: Kyung Hee University.
- Cheng, Rong 程榮. 2010. *A big dictionary of synonyms (Cihai edition) 同义词大词典 (辞海版)*. Shanghai: Shanghai Lexicographical Publishing House.
- Chinese College of Jinan University 暨南大学华文学院. 1997. *Chinese 中文*. Guangzhou: Jinan University Press.
- Chung, Siaw-Fong 鐘曉芳. 2011. A corpus-based analysis of "create" and "produce" 以語料庫為本分析近義詞. *Chang Gung Journal of Humanities and Social Sciences* 長庚人民社會學報 4(2):399–425.
- Cruse, Alan D. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- Deng, Yi, Rong Du, and Dianfang Yao 鄧懿, 杜榮, 姚殿芳. 1992-1993. *The Chinese course 汉语教程*. Beijing: Peking University Press.
- Dictionary Editing Room of Institute of Linguistics, China Academy of Social Sciences 中国社科院语言研究所词典编辑室. 2002. *The contemporary Chinese dictionary (Chinese-English Edition) 现代汉语词典(汉英双语)*. Beijing: Foreign Language Teaching and Research Press.
- Dictionary Editing Room of Institute of Linguistics, China Academy of Social Sciences 中国社科院语言研究所词典编辑室. 2016. *The contemporary Chinese dictionary 现代汉语词典 (7th ed.)*. Beijing: The Commercial Press.
- Ding, Xihan 丁熙翰. 1984. *Synonyms identification manual 同义词辨识手册*. Xi'an: Shaanxi People's Publishing House.
- English Channel of CCTV 英文央视频道. 2003. *Communicative Chinese 交际汉语*. Beijing: Science Popularization Press.
- Examination Center of The National Chinese Proficiency Test Committee 国家汉语水平考试委员会办公室考试中心. 2001. *The syllabus of Chinese vocabulary and characters levels 汉语水平词汇与汉字等级大纲*. Beijing: Economic Science Press.
- Fang, Lin 方琳. 2000. *An Illustrated synonym dictionary for student's 绘图学生同义词词典*. Chengdu: Sichuan Dictionary Publishing House.
- Ferré, Pilar, Isabel Fraga, Montserrat Comesaña, and Rosa Sánchez-Casas. 2015. Memory for emotional words: The role of semantic relatedness, encoding task and affective valence. *Cognition and Emotion* 29(8):1401–1410.
- Ge, Tianhong, and Fuqing Zhang 葛天红, 张福庆. 2000. *Latest student's practical synonym dictionary 最新学生实用同义词词典*. Beijing: Yanshan Publishing House.
- Gong, Shu-Ping, Kathleen Ahrens, and Chu-Ren Huang. 2008. Chinese word sketch and mapping principles: A corpus-based study of conceptual metaphors using the BUILDING source domain. *International Journal of Computer Processing of Languages* 21(1):3–17.
- Hong, Wei, and Nan Chen 洪炜, 陈楠. 2013. A study on the L2 acquisition of differences in similar sense and dissimilar sense of chinese near-synonyms 汉语二语者近义词差异的习得考察. *Applied Linguistics 语言文字应用* 2:99–106.
- Hong, Wei, and Xin Zhao 洪炜, 赵新. 2014. An empirical study of the difficulty of learning different types of chinese near-synonyms 不同类型汉语近义词习得难度考察. *Chinese Language Learning 汉语学习* 1:100–106.
- Huang, Chu-Ren. 2009. *Tagged Chinese Gigaword version 2.0*. (<http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2009T14>). Tagged from *Chinese Gigaword version 2.0*. (<https://catalog.ldc.upenn.edu/LDC2005T14>). Accessed on May 2013 and December 2016 through Chinese Word Sketch Engine.
- Huang, Chu-Ren, Keh-Jiann Chen, and Qing Xiong Lai 黃居仁, 陳克健, 賴慶雄 (eds.). 1997. *Mandarin Chinese classifier and noun classifier collocation dictionary 國語日報量詞典*. Taipei: Mandarin Daily Press.
- Huang, Chu-Ren, Adam Kilgariff, Yiching Wu, Chih-Ming Chiu, Simon Smith, Pavel Rychlý, Ming-Hong Bai, and Keh-Jiann Chen. 2005. Chinese sketch engine and the extraction of grammatical collocations. In *Proceedings of the fourth SIGHAN workshop on Chinese language processing*, 48–55. Stroudsburg: ACL.

- Huang, Chu-Ren, Lan Li, and Xinchun Su. 2016. Lexicography in the contemporary period. In *The Routledge encyclopedia of the Chinese language*, ed. Sin-Wai Chan, James Minett, and Wing-Yee Li, 545–562. New York: Routledge.
- Inkpen, Diana. 2004. Building a lexical knowledge-base of near-synonym differences. Ph.D. thesis. Toronto: University of Toronto.
- Kennedy, Graeme. 1998. *An introduction to corpus linguistics*. London: Longman.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The sketch engine: Ten years on. *Lexicography* 1(1):7–36.
- Kilgarriff, Adam, Chu-Ren Huang, Pavel Rychlý, Simon Smith, and David Tugwell. 2005. Chinese word sketches. Paper presented at *ASIALEX 2005: Words in Asian Cultural Context*, Singapore.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz, and David Tugwell. 2004. The sketch engine. In *Proceedings of the 11th EURALEX international congress*, ed. Geoffrey Williams and Sandra Vessier, 105–116. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- Kilgarriff, Adam, and David Tugwell. 2002. Sketching words. In *Lexicography and natural language: A festschrift in honour of B.T.S. Atkins (EURALEX 2002)*, ed. Marie-Hélène Corréard, 125–137. Grenoble: EURALEX.
- Kuperman, Victor, Zachary Estes, Marc Brysbaert, and Amy B Warriner. 2014. Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General* 143(3):1065–1081.
- Li, Xiaoqi 李晓琪. 2004–2008. *Boya Chinese* 博雅汉语. Beijing: Peking University Press.
- Liu, Mei-Chun, Ting-Yi Chiang, and Ming-Hui Chou. 2005. A frame-based approach to polysemous near-synonymy: The case with Mandarin verbs of expression. *Journal of Chinese Language and Computing* 15(3):137–148.
- Liu, Mei-Chun. 2016. Emotion in lexicon and grammar: Lexical-constructural interface of Mandarin emotional predicates. *Lingua Sinica* 2(1):1–47.
- Lu, Bin, Xiaojun Wan, Jianwu Yang, and Xiaoou Chen 路斌, 万小军, 杨建武, 陈晓鸥. 2007. Using *Tongyici Cilin* to compute word semantic polarity 基于同义词词林的词汇褒贬计. Paper presented at the *Proceedings of The 7th International Conference on Chinese Computing* 第七届中国信息处理国际会议, Wuhan, China.
- Lu, Wei-Lun, and Kathleen Ahrens. 2006. What corpora reveal about synonyms: A cognitive viewpoint on bother and worry. Paper presented at *The 1st International Symposium on Applied Linguistics*, Chiayi University, Taiwan.
- Mou, Shuyuan, and Shuo Wang 牟淑媛, 王硕. 2004. *A handbook of Chinese near-synonyms* 汉语近义词学习手册. Beijing: Peking University Press.
- Pavlenko, Aneta. 2008. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition* 11(2):147–164.
- Rundell, Michael, and Penny Stock. 1992. The corpus revolution. *English Today* 8(3):21–32.
- Rychlý, Pavel. 2008. A lexicographer-friendly association score. In *Proceedings of the 2nd workshop on recent advances in Slavonic natural languages processing*, ed. Petr Sojka and Aleš Horák, 6–9. Brno: Masaryk University.
- Shi, Biao 士彪. 2002. *A Primary school student's synonym dictionary* 小学生同义词词典. Beijing: Oriental Press.
- Su, Xinchun 苏新春. 2006. *A Sentence-making dictionary for student's* 学生造句词典. Shanghai: Shanghai Lexicographical Publishing House.
- Summers, Della. 1993. Longman/Lancaster English language corpus—criteria and design. *International Journal of Lexicography* 6(3):181–208.
- Tai, Fang 泰芳. 1997. *A carefully compiled new dictionary of synonyms and antonyms* 精编同义词反义词新典. Jinan: Tomorrow Press.
- Teng, Shou-Hsin 鄧守信. 2009. *A Chinese synonym's usage dictionary* 漢語近義詞用法詞典. Taipei: Bookman Books Co. Ltd.
- Tsai, Mei-Chih, Chu-Ren Huang, Keh-Jiann Chen, and Kathleen Ahrens. 1998. Towards a representation of verbal semantics—an approach based on near synonyms. *Computational Linguistics and Chinese Language Processing* 3(1):61–74.
- Unit of Dictionaries, Foreign Language Teaching and Research Press 外语教学与研究出版社辞书部. 2001. *A modern Chinese-English dictionary* 现代汉英词典. Beijing: Foreign Language Teaching and Research Press.
- Wang, Shan. 2012. Semantics of event nouns. Ph.D. thesis. Hong Kong: The Hong Kong Polytechnic University.
- Wang, Shan, and Chu-Ren Huang. 2011. Compound event nouns of the 'modifier-head' type in mandarin Chinese. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC-25)*, ed. Helena H Gao and Minghui Dong, 511–518. Singapore: Nanyang Technological University.
- Wang, Shan, and Chu-Ren Huang. 2013a. Apply Chinese word sketch engine to facilitate lexicography. In *Lexicography and dictionaries in the information age: Selected papers from the 8th ASIALEX international conference*, ed. Deny A Kway, Nur Wulan, and Lilla Musyaha, 285–292. Surabaya: Airlangga University Press.
- Wang, Shan, and Chu-Ren Huang. 2013b. The semantic type system of event nouns. In *Increased empiricism: Recent advances in Chinese linguistics*, ed. Zhuo Jing-Schmidt, vol. 2, 205–221. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Wang, Shan, Chu-Ren Huang, and Hongzhi Xu. 2012. Compositionality of NN compounds: A case study on [N_i+Artifactual-type event nouns]. In *Proceedings of the 26th Pacific Asia conference on language, information and computation (PACLIC-26)*, 70–79. Bali: Faculty of Computer Science, Universitas Indonesia.
- Wu, Yang, and Shan Wang. 2016. Applying Chinese word sketch engine to distinguish commonly confused words. In *Chinese lexical semantics*, ed. Minghui Dong, Jingxia Lin, and Xuri Tang, 600–619. Cham: Springer.
- Xu, Lin-Hong, Hong-Fei Lin, and Zhi-Hao Yang 徐琳宏, 林鸿飞, 杨志豪. 2007. Text orientation identification based on semantic comprehension 基于语义理解的文本倾向性识别机制. *Journal of Chinese Information Processing* 中文信息学报 21(1):96–100.
- Yang, Jizhou, and Yongfen Jia 杨寄洲, 贾永芬. 2005. *Comparison of the usage of 1700 synonym pairs* 1700 对近义词语用法对比. Beijing: Beijing Language and Culture University Press.
- Yang, Jizhou, and Shude Ma 杨寄洲, 马树德. 1999–2003. *The Chinese course* 汉语教程. Beijing: Beijing Language and Culture University Press.v
- You, Bin, Yue-Song Yan, Ying-Ge Sun, and Jing Liu 游彬, 严岳松, 孙英阁, 刘靖. 2013. Method of information content evaluating semantic similarity on HowNet 基于 HowNet 的信息量计算语义相似度算法. *Computer Systems and Applications* 计算机系统应用 22(1):129–133.

- Yu, Shi-Wen, Xue-Feng Zhu, Hui Wang, Hua-rui Zhang, Yun-Yun Zhang, De-Xi Zhu, Jian-Ming Lu, and Rui Guo 俞士汶, 朱學鋒, 王惠, 张化瑞, 张芸芸, 朱德熙, 陆俭明, 郭锐. 2003. *The grammatical knowledge-base of contemporary Chinese—A complete Specification 现代汉语语法信息词典详解 (2nd ed.)*. Beijing: Tsinghua University Press.
- Yuen, Raymond WM, Terrence YW Chan, Tom BY Lai, Oi Yee Kwong, and Benjamin KY T'sou. 2004. Morpheme-based derivation of bipolar semantic orientation of Chinese words. Paper presented at *the 20th international conference on computational linguistics*. Geneva, Switzerland.
- Zhang, Bo 张博. 2016. *A study on the distribution of confusing words and its causes of Chinese learners with different mother tongue background 不同母语背景的汉语学习者词语混淆分布特征及其成因研究*. Beijing: Peking University Press.
- Zhang, Wen-Xian, Li-Kun Qiu, Zuo-Yan Song, and Bao-Ya Chen 张文贤, 邱立坤, 宋作艳, 陈保亚. 2012. Corpus-based quantitative analysis on stylistic difference of Chinese synonyms 基于语料库的汉语同义词语体差异定量分析. *Chinese Language Learning 汉语学习* 3:72–80.
- Zhao, Xin, Wei Hong, and Jing-Jing Zhang 赵新, 洪炜, 张静静. 2014. *The study and teaching of Chinese synonyms 汉语近义词研究与教学*. Beijing: The Commercial Press.
- Zhao, Xin, and Ying Li 赵新, 李英. 2009. *Chinese synonyms dictionary of The Commercial Press 商务馆学汉语近义词词典*. Beijing: The Commercial Press.
- Zhong, Yi, and Na Yi 仲弋, 艺娜. 1999. *A Chinese student's synonym dictionary 中华学生同义词词典*. Shantou: Shantou University Press.
- Zhu, Jingsong 朱景松. 2009. *A Modern Chinese dictionary of synonyms 现代汉语同义词词典*. Beijing: Language and Culture Press.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
