

# Corpora-Based Comparative Analysis of Synonyms “Situation, Environment and Circumstance”

Sun Jing<sup>\*</sup>, Yin Zhuhui

School of Foreign Studies, Beijing Information Science & Technology University, Beijing, China

**Email address:**

1757298300@qq.com (Sun Jing)

<sup>\*</sup>Corresponding author

**To cite this article:**

Sun Jing, Yin Zhuhui. Corpora-Based Comparative Analysis of Synonyms “Situation, Environment and Circumstance”. *International Journal of Literature and Arts*. Vol. 8, No. 4, 2020, pp. 206-216. doi: 10.11648/j.ijla.20200804.15

**Received:** May 7, 2020; **Accepted:** May 25, 2020; **Published:** June 4, 2020

---

**Abstract:** This study compares the frequency, collocation and semantic prosody of *situation*, *environment* and *circumstance* in Corpus of Contemporary American English (COCA) and Written English Corpus of Chinese Learners (WECCL). Vocabulary is the cornerstone of second language learning and its development is one of the hot topics in second language acquisition researches. Huge in size, subtle in semantic difference and often presented in minimal contexts, synonyms in English remain a permanent challenge for Chinese English learners. Traditional synonym differentiation usually relies on item-by-item analysis of lexical meanings and introspective qualitative methods like intuition, experience, etc. However, in practice, such guesswork is far from satisfactory. With the rapid development of information technology, corpus linguistics provides a key to the study of synonyms. Based on corpus data, this paper compares and analyzes the selected synonyms using Antconc, Chi-square Calculator and BFSU Collocator. Findings indicate: 1) In terms of use frequency, Chinese students tend to overuse synonyms compared with native speakers; 2) In terms of salient collocations, Chinese students use the synonyms with lots of semantic ambiguity but few collocation types; 3) In terms of semantic prosody, on the Chinese students' side, inadequate accuracy is comparatively conspicuous as well as semantic prosody misuse. Preliminary cause analysis points to two main factors influencing Chinese students' mastery of synonyms, respectively L1 negative transfer and the misleading effect of Chinese-English dictionaries. Based on the above analysis, this paper puts forward some suggestions for teaching synonyms and compiling dictionaries.

**Keywords:** Corpus, Synonyms, Frequency, Collocation, Semantic Prosody

---

## 1. Introduction

This study compares the frequency, collocation and semantic prosody of *situation*, *environment* and *circumstance* in Corpus of Contemporary American English (COCA) and Written English Corpus of Chinese Learners (WECCL). Mastery of the subtle differences between synonyms is a key standard for English proficiency. Yet it remains a prominent problem for learners of English considering the huge size of synonyms. Traditional means of synonym differentiation, tedious and lacking accuracy, relies on item-by-item analysis of lexical meaning or introspective qualitative methods. But modernized approach emerges with the help of Corpus linguistics, a rapidly developing discipline with information technology and massive data. For synonym analysis, it provides authentic linguistic contexts and a possibility of more objective observation with greater accuracy than the

somewhat subjective introspection method.

In this paper, the author observes three synonyms in two corpora, WECCL and COCA, focusing on the following questions with some necessary data analysis methods: What are the similarities and differences in terms of the use frequency, collocation and semantic prosody of “situation, environment, and circumstance” by Chinese English learners and English native speakers?

## 2. Literature Review

### 2.1. Synonym Studies Abroad

The rapid development of corpus linguistics drives enthusiastic synonym studies. Corpora prove a powerful tool for researching lexical usages, collocations and so on. With authentic and specific contexts of language use they provide

fairly objective language data particularly in terms of synonym differentiation, traditionally an objective guess work.

Corpus linguistics research abroad has experienced 50 years of development, and great achievements have been made in the domain of corpus-based vocabulary study. Since the beginning of the 1960s when N. Francis and H. Kucera began to design and build BROWN, scholars began to conduct large-scale computer-based corpus research, and corpus linguistics gradually developed. Kennedy [1] believed that words and grammar could not be treated separately; that teaching and learning should feature an organically integrated “lexicogrammar”, hence the corpus study along the same line. Saeed’s study [2] distinguished several sets of synonyms involving conceptual meaning, connotation meaning, emotional meaning and stylistic meaning respectively. He distinguished words *press and cupboard* in dialects; *police, officer, cop and copper*; *naïve, gullible and ingenuous* according to their connotations; *wife, spouse, old lady and Missus* according to the register. Harward and Etienne [3] observed synonyms in different contexts. They discussed *pavement and sidewalk* in details under the influence of different dialects (British English and American English), and the formality and contextual style of *beauty and pulchritude*.

Hoey [4] proposed ten hypotheses on lexical triggering, among which three involved synonyms: 1) any word will be triggered to appear in a specific semantic group, which is its semantic association; 2) the differences between synonyms are mainly reflected in their different collocations with other words, different grammatical and semantic associations; 3) each word will be triggered to appear in a specific discourse relation, which is its textual collocation. Shahzadi, Asghar and Javed [5] applied Sketch Engine (SkE) to analyze the collocation, concordance, word sketches and sketch difference of synonyms *arrive* and *reach* in British National Corpus (BNC), and discussed how to effectively teach different functions of synonyms in the naturally generated discourse of corpus. Their study found that there is more meaning associated with *reach* than with *arrive*, indicating the widespread occurrence and use of *reach*. SkE data also showed that the frequency of *reach* in BNC is much higher than that of *arrive*. It proves that data analysis based on a naturally generated corpus can be used as an effective strategy to distinguish and teach synonyms *arrive* and *reach*.

## 2.2. Domestic Synonym Studies

Lu [6] referred to English corpus FLOB and FROWN in comparison to CLEC corpus with *cause, lead to, result in/from* in terms of frequency, collocation and semantic prosody for analysis. It was found that learners had difficulties in distinguishing these features. Different groups of learners showed significant differences in their word-use characteristics and acquisition patterns of synonyms, as expressed in varied frequencies of synonym uses and semantic conflicts in the use.

Sun [7] studied the pair of synonyms “*affect*” and “*influence*” and compared their semantic similarities and

differences in the native English corpus FROWN and the learner corpus CLEC. The author focused on three parameters: semantic prosody norm, semantic prosody polarity and semantic prosody strength, and analyzed the pragmatic meaning of the two words in the two corpora.

Literature review suggests the following interpretations. First, there are generally two methods of synonym differentiation by scholars at home and abroad, namely the traditional method and the method based on corpus. Second, corpus-based research generated abundant statistics in terms of quantity. Third, synonyms are usually studied and compared by corpus retrievals involving use frequency, register, collocation and semantic prosody at several levels.

In the meantime, the author makes a categorical search of domestic literature and finds that the research on English synonyms based on corpus covers a comprehensive range of parts of speech, while the majority research focuses on synonymous verbs and adjectives. In addition, the research on synonymous nouns usually only involves characteristics of semantic prosody but seldom investigates different features of collocation. In view of this, corpus, as a new generation of empirical research tool, has not fully played its significant advantages to a large enough extent. Moreover, the author thinks that the research of nouns based on corpus is not deep enough, calling for further research and exploration. Based on this, from the perspective of corpus linguistics, this study intends to discover some significant differences in frequency, collocation and semantic prosody based on corpora of Chinese learners and native speakers by taking a group of synonymous nouns as examples.

## 3. Theoretical Framework

### 3.1. Corpus Linguistics

Corpus linguistics emerged in the 1980s as an important breakthrough in research methodology and went through three developmental stages. It updated linguistic description framework and linguistic views [8] and liberated lexical research methods from card making and manual retrieval. The first stage featured the world’s first established Corpus of Brown University Standard Corpus of Present Day American English (BROWN) with nearly one million words of British and American English, covering 15 written styles. The second stage was marked by corpus establishment in different parts of the world and symbolized by joint creations of international corpora. In China, the 500,000-word GPEC corpus (Guangzhou Petroleum English Corpus) was built then. Worldwide the BNC Corpus is the best known with a size of about 100 million words in the 1990s. The third stage is characterized with three aspects: large-scale and multi-type corpus; developed corpus processing; universal application of corpus in all fields of language. With vast amount of genuine linguistic data corpus linguistics tries to reveal the complexity of natural language from a new perspective.

### 3.2. Dimensions of Synonym Differentiation in Corpus Linguistics

Kennedy [9] proposed four levels of corpus analysis: (1) Lexical level: to explore a word with view to frequency, the context and adjacent collocation. (2) Syntactic level: to quantitatively analyze various vocabulary combinations and sentence patterns with grammar and part of speech. (3) Discourse structure: to study the coherence and cohesion features of oral and written languages. (4) Disciplinary types: to explore different types of language features to discover emergent standards or sub-standards.

Frequency is an indispensable theoretical concept in the study of lexical collocation in linguistics. Word frequency statistics, both manual and computerized, reveals and analyzes the occurrences of words to get lexical rules. It is often used to study vocabulary regularities, word frequency, teaching and so on.

Another essential concept is collocation, studied for more than 50 years from different angles in different periods. Collocation is the co-occurrence of two or more adjacent words in a text [10]-1 usually described by three important terms: node words, collocation, and span. Calculation of collocations aims at typicality, which usually involves Z-score or T-score of the original number about the node word. The higher the score, the more typical the collocation is.

The third key concept, semantic prosody, is defined in various ways by Sinclair [10]-2, Louw [11], and Stubbs [12]. In a certain context lexical items manifest a strong semantic preference due to the pragmatic purpose. Semantic prosody selection/mechanism strongly restricts the choice of collocation words [13]-1, which could be roughly divided into three categories: positive, neutral and negative prosody [14]. The study of semantic prosody can play an important role in the differentiation of synonyms. And applied to vocabulary teaching, it will effectively supplement the traditional teaching methods.

## 4. Corpora

### 4.1. Corpus of Contemporary American English (COCA)

The Corpus of Contemporary American English (COCA), the largest corpus of American English is freely available online. It contains more than one billion words (20 million words each year 1990-2019) from texts of eight genres: spoken, fiction, popular magazines, newspapers, academic texts, and (with the update in March 2020): TV and Movies subtitles, blogs, and other web pages, each sub-corpus basically balanced. Users can retrieve from the entire corpus or any sub-corpus. A good reference resource for researchers, English teachers and students, it provides a window to inquiry and observation of the uses and changes of American English. Its simple operation interface brings great convenience to researchers.

### 4.2. Corpus of Contemporary American English (COCA)

Written English Corpus of Chinese Learners 2.0

(WECCL2.0), as a substantial sub-corpus of Spoken and Written English Corpus of Chinese Learners 2.0 (SWECCL 2.0), is based on WECCL 1.0 published in 2005 by Foreign Language Teaching and Research Press. Edited by Wen [15] and representing Chinese college students' performance in writing expository essays, WECCL has a great influence on linguistics, it is a database to display the features and essence of interlanguage, which provides insights into second language acquisition and foreign language teaching.

## 5. The Present Research

### 5.1. Research Tools

The analysis tools used in this study are Antconc, Chi-square Calculator and BFSU Collocator. AntConc is a retrieval tool with several basic functions such as extracting the contextualized co-occurrence of node words and use frequency. Chi-square ( $X^2$ ) test calculator is used to judge statistically significant differences. BFSU Collocator is a collocation analysis software calculating collocation intensity by MI (mutual information), MI3, Z-score, T-score, Log-log and Log-likelihood.

### 5.2. Research Findings

#### 5.2.1. Frequency

Corpus sizes of WECCL and COCA are 1,248,476 and 1,001,610,938 respectively. Because WECCL features written language, its equivalence is supposed to be the ACADEMIC genre in COCA for the sake of this study, size being 120,988,348.

i. The Frequency Data of Synonyms in WECCL and COCA

Usage frequencies of situation, environment and circumstance are listed respectively in Table 1:

Table 1. The frequency of synonyms in WECCL and COCA.

Word/frequency	WECCL	COCA
Situation	550	19646
Environment	1091	31709
Circumstance	59	933

According to the above data, *environment* is the word most frequently used by both Chinese students and native speakers. Next comes situation, about 50% (for Chinese learners or 60% for native speakers) the frequency of *environment*; *Circumstance*, has a frequency of about 5% (or 3% for native speakers) of *environment*. Both groups most often use *environment*, and then *situation*; *circumstance* is the least used.

ii. The Normalized Frequency Data of Synonyms in WECCL and COCA

Different corpus sizes entail the use of normalized frequency for the sake of comparison. Normalized frequency presents the proportional figure [16] of actual observation (raw) frequency of a certain retrieval item over the total frequency. In a frequency normalization operation, the ratio is standardized by 1 thousand (or 10 thousand, 1 million) to get the average frequency per thousand (or per 10 thousand, 1

million). The frequency standardization of this study is in units of frequency per million, as shown in the following table:

**Table 2.** The normalized frequency of synonyms in WECCL and COCA.

word	WECCL (1,248,476) Token/per million	COCA (120,988,348) Token/per million
	Native Speakers	English Language Learners
Situation	550/440.54	19646/162.38
Environment	1091/873.87	31709/262.08
Circumstance	59/47.26	933/7.71

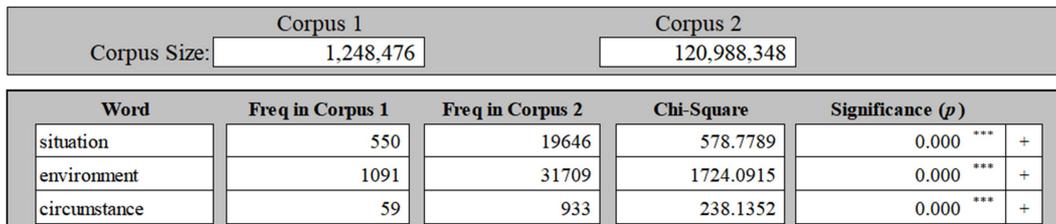
The normalized frequencies indicate that usages of *situation*, *environment* and *circumstance* by Chinese learners are significantly different from those of native speakers. The standard frequencies of native speakers are 162.38, 262.08 and 7.71, which are far less than the standard frequencies of Chinese learners, 440.54, 83.87 and 47.26. Among them, the standard frequency of the word *environment* in the two corpora is higher than the other two. Moreover, the standard frequency of *situation* is about 2.7 times that of native

speakers, the standard frequency of *environment* is about 3.3 times that of native speakers, and *circumstance* has about 6 times frequency that of native speakers'. Why? It calls for further studies.

iii. The Chi-Square Analysis Data of Synonyms in WECCL and COCA

In order to determine whether there is a significant difference in the use frequency by native speakers and English learners, Chi-Square test is attempted as in Figure 1.

### Chi-Square Calculator



**Figure 1.** The Chi-Square test of synonyms (corpus1: WECCL, corpus2: COCA).

It suggests that scores of *situation*, *environment* and *circumstance* are 578.7789, 1724.0915 and 238.1352, far higher than the critical score of 3.83. Since P-score is far lower than 0.05, we can be 95% confident of the significant differences of the synonym applications in the two corpora. Chinese students tend to overuse *situation*, *environment* and *circumstance*.

#### 5.2.2. Collocation

Analysis of “parts of speech” data is crucial for the study of collocation. Based on COCA and WECCL, this study compares percentages of varied types of parts of speech in collocations of *situation*, *environment* and *circumstance* as shown in the table below:

**Table 3.** Parts of speech and percentages of collocations in COCA and WECCL.

Sample Synonyms	Corpus	COCA		WECCL	
		Parts of Speech	Percentage	Parts of Speech	Percentage
situation		adjective	89.7%	adjective	89.8%
		noun	6.9%	determiner	3.4%
		determiner	3.4%	article	3.4%
environment		adjective	89.7%	verb	3.4%
		noun	6.9%	adjective	86.4%
		verb	3.4%	pronoun	3.4%
				article	3.4%
circumstance				verb	3.4%
		adjective	96.6%	preposition	3.4%
		determiner	3.4%	adjective	72.6%
				noun	17.2%
			determiner	6.9%	
			article	3.4%	

i. The Collocation Data of Synonyms in COCA

The study focus is the collocation with adjectives of a given synonym. Collocation strength between the node word and the collocation word is measured through the MI-score (mutual information). When the score is greater than 3, the

collocation is considered salient. In COCA’s query interface, the symbol span is set as [1/L, 0/R], and the minimum MI-score (mutual information) 3. Retrieved in COCA the salient collocations of *situation*, *environment* and *circumstance* (MI>3) are 51,86 and 29, respectively. 29 is the

largest possible number of collocations for circumstance, sample synonym are listed for comparison as table 4. therefore in this study, 29 salient collocations (MI>3) of each

Table 4. The collocation of situation in COCA.

Rank	Salient collocations	MI-score	Rank	Salient collocations	MI-score
1	no-win	11.88	16	unusual	5.08
2	win-win	10.64	17	competitive	4.87
3	stressful	8.33	18	economic	4.74
4	precarious	8.10	19	particular	4.73
5	dire	7.41	20	colonial	4.64
6	real-life	7.38	21	bad	4.54
7	hypothetical	7.26	22	employment	4.50
8	geopolitical	7.16	23	living	4.49
9	paradoxical	7.14	24	unique	4.47
10	current	5.73	25	similar	4.47
11	emergency	5.57	26	security	4.37
12	dangerous	5.54	27	present	4.34
13	financial	5.52	28	this	4.29
14	ideal	5.52	29	difficult	4.27
15	problematic	5.45			

As is seen, the salient collocations of situation are mainly adjectives, nouns and determiners. The adjectives are: *no-win*, *win-win*, *stressful*, *precarious*, *dire*, *real-life*, *hypothetical*, *geopolitical*, *paradoxical*, *current*, *emergency*, *dangerous*, *financial*, *ideal*, *problematic*, *unusual*, *competitive*, *economic*, *particular*, *colonial*, *bad*, *living*, *unique*, *similar*, *present* and

*difficult*, accounting for about 89.7% (see table 3) of the total salient collocation words; The nouns are: *security* and *employment*, accounting for about 6.9% (ibid); the determiner is *this*, accounting for about 3.4% (ibid). Therefore, ADJ. +*situation* has the highest proportion.

Table 5. The collocation of environment in COCA.

Rank	Salient collocations	MI-score	Rank	Salient collocations	MI-score
1	healthful	9.94	16	aquatic	6.30
2	restrictive	8.07	17	harsh	6.15
3	cleaner	7.70	18	learning	6.14
4	noisy	7.70	19	classroom	5.99
5	nurturing	7.48	20	competitive	5.89
6	welcoming	7.45	21	enabling	5.84
7	built	7.16	22	surrounding	5.84
8	hostile	7.12	23	arctic	5.62
9	supportive	6.93	24	external	5.61
10	marine	6.85	25	virtual	5.58
11	safe	6.65	26	unfamiliar	5.51
12	indoor	6.64	27	inclusive	5.45
13	ERP	6.61	28	healthy	5.43
14	atmospheric	6.37	29	clean	5.40
15	natural	6.32			

As is seen, the salient collocations of *environment* are mainly adjectives, nouns and verbs. The adjectives are: *healthful*, *restrictive*, *noisy*, *nurturing*, *welcoming*, *hostile*, *supportive*, *marine*, *safe*, *indoor*, *atmospheric*, *natural*, *aquatic*, *harsh*, *learning*, *competitive*, *enabling*, *surrounding*, *arctic*, *cleaner*, *external*, *virtual*, *unfamiliar*, *inclusive*,

*healthy*, and *clean*, accounting for about 89.7% (see table 3) of the total salient collocations; the nouns are: *ERP*, *classroom*, accounting for about 6.9% (ibid); the verb is *built*, about 3.4% (ibid). Therefore, ADJ. +*environment* accounts for the highest proportion.

Table 6. The collocation of circumstance in COCA.

Rank	Salient collocations	MI-score	Rank	Salient collocations	MI-score
1	unfurnished	15.39	16	external	5.37
2	aggravating	11.96	17	changing	5.14
3	mitigating	9.45	18	changed	5.12
4	singular	9.08	19	unique	4.84
5	fortunate	8.85	20	this	4.77
6	unfortunate	8.45	21	specific	4.54
7	exceptional	7.64	22	material	4.51
8	unusual	7.34	23	similar	4.31
9	extraordinary	7.11	24	given	4.18
10	latter	6.26	25	difficult	3.94

Rank	Salient collocations	MI-score	Rank	Salient collocations	MI-score
11	special	6.00	26	economic	3.74
12	historical	5.99	27	local	3.52
13	particular	5.95	28	individual	3.32
14	any	5.52	29	political	3.24
15	every	5.44			

As is seen, *circumstance's* salient collocation words are mainly adjectives and determiners. The adjectives are: *unfurnished, aggravating, mitigating, singular, fortunate, unfortunate, exceptional, unusual, extraordinary, latter, special, historical, particular, any, every, external, changing, changed, unique, specific, material, similar, given, difficult, economic, local, individual and political*, accounted for about 96.6% (see table 3) of the total salient collocations. The determiner is *this*, accounting for about 3.4% (ibid).

Therefore, ADJ.+*circumstance* accounts for the highest proportion.

ii. The Collocation Data of Synonyms in WECCL

With WECCL corpus through BFSU Collocator1.0, the span is set as [1/L, 0/R], and collocations with MI-score (mutual information)  $\geq 3$  and  $Z \geq 2$  are regarded salient collocations, 91,109 and 31, respectively. Again 29 of them (MI $\geq 3$  and  $Z \geq 2$ ) are listed as shown in the table below.

Table 7. The collocation of situation in WECCL.

Rank	Salient collocations	MI-score/Z-score	Rank	Salient collocations	MI-score/Z-score
1	current	18.40/122.54	16	sub-healthy	11.36/51.16
2	this	18.02/57.05	17	touchy	11.36/51.16
3	the	17.27/27.56	18	unconscious	11.36/51.16
4	present	14.01/38.47	19	different	11.29/13.63
5	harsh	13.36/51.10	20	embarrassed	11.19/34.06
6	serious	12.93/27.57	21	real	10.71/16.32
7	difficult	12.84/26.68	22	same	10.60/12.46
8	embarrass	12.77/59.06	23	similar	10.36/20.75
9	embarrassing	12.36/51.14	24	upsetting	10.36/36.16
10	bad	11.85/18.71	25	unfamiliar	9.90/21.73
11	economic	11.83/24.34	26	awkward	9.77/29.51
12	specific	11.60/27.72	27	occasional	9.77/29.51
13	educated	11.36/51.16	28	troubling	9.77/29.51
14	getting-worse	11.36/51.16	29	new	9.50/9.50
15	preset	11.36/51.16			

As is seen, the salient collocations of *situation* are mainly adjectives, determiners, articles and verbs. The adjectives are: *current, present, harsh, serious, difficult, embarrassing, bad, economic, specific, educated, getting-worse, preset, sub-healthy, touchy, unconscious, different, embarrassed, real, same, similar, upsetting, unfamiliar, awkward,*

*occasional, troubling* and *new*, accounting for 89.8% (see table 3) of the total salient collocations; the qualifier is *this*, accounting for about 3.4% (ibid); The article is *the*, accounting for about 3.4% (ibid); the verb *embarrass*, 3.4%. Therefore, ADJ. +*situation* has the highest proportion.

Table 8. The collocation of environment in WECCL.

Rank	Salient collocations	MI-score/Z-score	Rank	Salient collocations	MI-score/Z-score
1	our	20.56/82.11	16	studying	10.84/13.65
2	the	19.37/40.28	17	competitive	10.81/15.59
3	comfortable	15.37/46.77	18	vulnerable	10.58/27.57
4	protecting	14.99/47.87	19	alluring	10.39/36.54
5	quiet	14.65/46.00	20	controllable	10.39/36.54
6	good	14.42/22.08	21	homing	10.39/36.54
7	living	14.20/23.42	22	pollution-free	10.39/36.54
8	better	13.84/21.66	23	pretesting	10.39/36.54
9	new	12.86/18.79	24	rest-conductive	10.39/36.54
10	changeable	12.39/51.68	25	societal	10.39/36.54
11	natural	12.18/23.72	26	study-conductive	10.39/36.54
12	protect	11.74/17.96	27	teaching	10.39/36.54
13	peaceful	11.68/23.13	28	variable	10.39/36.54
14	of	10.91/3.02	29	wally	10.39/36.54
15	safe	10.89/17.41			

As is seen, *environment's* salient collocations are mainly adjectives, pronouns, articles, verbs and prepositions. The adjectives are: *comfortable, protecting, quiet, good, living, better, new, changeable, natural, peaceful, safe, studying,*

*competitive, vulnerable, alluring, controllable, homing, pollution-free, pretesting, rest-conductive, societal, study-conductive, teaching, variable* and *wally*, accounting for about 86.4% (see table 3) of the salient collocations;

The pronoun is *our*, about 3.4% (ibid). The article is *the*, accounting for about 3.4% (ibid); the verb, *protect*, about 3.4% (ibid); the preposition *of*, about 3.4% (ibid). Therefore, ADJ. +*environment* accounts for the highest proportion.

Table 9. The collocation of circumstance in WECCL.

Rank	Salient collocations	MI-score/Z-score	Rank	Salient collocations	MI-score/Z-score
1	male-dominated	13.57/110.30	16	dangerous	7.55/13.61
2	globalizing	11.99/63.67	17	good	7.50/7.40
3	adverse	11.76/58.95	18	new	7.30/8.65
4	gloomy	11.57/55.14	19	peace	7.22/12.14
5	academy	10.99/45.01	20	excellent	7.14/11.78
6	this	10.80/15.53	21	every	7.04/7.87
7	concrete	10.66/40.25	22	such	6.98/7.70
8	changeable	10.18/34.01	23	comfortable	6.81/10.50
9	ease	9.93/31.17	24	certain	6.56/9.61
10	superior	8.99/22.47	25	equal	6.45/9.25
11	social	8.73/14.41	26	language	6.20/8.47
12	warm	8.53/19.15	27	free	6.17/8.37
13	restaurants	7.97/15.78	28	private	6.15/8.31
14	any	7.89/10.69	29	working	5.97/7.79
15	the	7.63/3.45			

As is seen, circumstance's salient collocation words are mainly adjective, noun, determiner and article. The adjectives are: *male-dominated, globalizing, adverse, gloomy, concrete, changeable, superior, social, warm, any, dangerous, new, excellent, good, every, comfortable, certain, equal, free, private* and *working*, accounting for about 72.6% (see table 3)

of the total salient collocations; The nouns are: *academy, ease, restaurants, peace* and *language*, accounting for about 17.2% (ibid). The determiners are *this* and *such*, about 6.8% (ibid); the article *the*, about 3.4% (ibid). Therefore, ADJ.+*circumstance* accounts for the highest proportion.

iii. Comparative Analysis of Synonyms Collocation in COCA and WECCL

Table 10. The collocation of synonyms in COCA and WECCL.

	COCA	WECCL
situation	no-win, win-win, stressful, dire, precarious, real-life, hypothetical, geopolitical, paradoxical, current, emergency, dangerous, financial, ideal, problematic, unusual, competitive, economic, particular, colonial, bad, living, employment, unique, similar, security, present, this, difficult	current, present, harsh, serious, bad, difficult, embarrassing, economic, specific, educated, getting-worse, preset, sub-healthy, touchy, same, unconscious, different, embarrassed, real, similar, this, the, embarrass, upsetting, unfamiliar, awkward, occasional, troubling, new
environment	healthful, restrictive, noisy, hostile, nurturing, welcoming, supportive, marine, safe, indoor, atmospheric, natural, aquatic, harsh, learning, competitive, enabling, surrounding, arctic, cleaner, ERP, classroom, built, external, virtual, unfamiliar, clean, inclusive, healthy	comfortable, protecting, quiet, good, living, better, new, changeable, natural, peaceful, safe, studying, competitive, vulnerable, alluring, controllable, homing, pollution-free, pretesting, our, the, protect, of, wally, rest-conductive, societal, variable, study-conductive, teaching
circumstance	unfurnished, aggravating, singular, mitigating, fortunate, unfortunate, exceptional, unusual, extraordinary, latter, special, historical, particular, any, every, external, changing, changed, unique, specific, material, similar, this, given, difficult, local, economic, individual, political	male-dominated, globalizing, any, gloomy, concrete, changeable, superior, social, warm, adverse, dangerous, good, new, excellent, every, comfortable, academy, ease, restaurants, peace, this, such, the, certain, equal, language, free, private, working

Corpus-based analysis indicates that different synonyms have their own preferential collocations. The sample synonyms *situation, environment* and *circumstance* differ greatly in their salient collocation words in the two corpora. Compared with native speakers of English, Chinese learners use less connections of the Adj.+N type, but they use more types of parts of speech than native speakers.

Chinese learners tend to use "serious" to describe a situation. However, native speakers tend to use *dire situation* rather than *serious situation*. When Chinese learners choose words to modify *environment*, they are also significantly different from native speakers. For example, in the two

corpora, the collocation words "noisy" and "quiet" are both salient. Chinese students prefer to use the expression "quiet environment", while native speakers tend to use "noisy environment". Chinese learners use ADJ. + *circumstance* and N. + *circumstance*, which may be caused by L1 negative transfer, while native speakers mainly focus on ADJ. + *circumstance*.

To sum up, in view of the sample synonyms Chinese students mostly choose general adjectives in the salient collocations, such as *good, bad, serious*, which seem simple, limited and boring to a certain extent. Native speakers of English, on the other hand, have a wide range of salient collocations with semantically specific content words of

various kinds such as *dire*, *fortunate*, *precarious*. These words can vividly express semantic intentions. The lack of semantic specificity and vividness in Chinese learners use of English is most likely due to the limited mastery of vocabulary and idiomatic expressions.

### 5.2.3. Semantic Prosody

#### i. The Semantic Prosody Data of Synonyms in COCA

According to Sinclair [13]-2, semantic prosody means the trend of semantic preference driven by pragmatic purpose.

**Table 11.** The semantic prosody distribution of situation in COCA.

Type	Adjective Collocations (Frequency)	Total Frequency	Percentage	Semantic prosody
1	win-win (48), ideal (58)	106	5.4%	Positive
2	real-life (22), hypothetical (44), geopolitical (26), current (325), financial (157), unusual, competitive (35), economic (254), particular (162), living, similar (141), present (149), unique (49)	1462	74.6%	Neutral
3	no-win (21), stressful, dire (23), precarious (31), paradoxical (21), emergency (37), dangerous (38), problematic (33), colonial (33), bad (31), difficult (79)	392	20.0%	Negative

The above table shows the semantic prosody characteristics of the prominent adjective collocations in COCA of *situation*. Among them, the percentage of positive semantic prosody including *win-win*, *ideal* is 5.4%. Negative semantic prosody, including *no-win*, *stressful* etc., accounted for 20%. Neutral semantic prosody included *real-life*,

The tendency of semantic preference in turn strongly restricts the choice of collocation words, which results in homogeneous items of a limited number of semantic groups. The number of salient collocations ( $MI > 3$ ) (mutual information) of *situation*, *environment* and *circumstance* are 51, 86 and 29 respectively in COCA. The numbers of adjectives are 26, 26 and 28 respectively. The collocation words can be divided into groups of positive, negative and neutral semantic prosody.

*hypothetical*, etc., 74.6%. To sum up, *situation* shows a neutral semantic prosody feature in COCA. At the same time, the data gives an important message: *situation* has a strong tendency to combine with neutral or negative words, compared to just 5.4% of positive words that convey agreeable indications.

**Table 12.** The semantic prosody distribution of environment in COCA.

Type	Adjective Collocations (Frequency)	Total Frequency	Percentage	Semantic prosody
1	healthful (84), nurturing (42), welcoming (25), supportive (105), safe (170), learning (1202), clean (49), enabling (34), cleaner (27), healthy (80)	1818	53.0%	Positive
2	restrictive (127), marine (153), indoor (32), atmospheric (38), arctic (20), natural (581), aquatic (35), virtual (57), competitive (114), surrounding (90), external (161), inclusive (29)	1437	41.9%	Neutral
3	noisy (27), hostile (97), harsh (30), unfamiliar (21)	175	5.1%	Negative

The above table shows that the semantic prosody of adjectives used by native speakers to modify *environment* includes positive, negative and neutral categories. Among them, the percentage of positive category including *healthful*, *welcoming* and so on is 53.0%. Negative semantic prosody, including *noisy*, *harsh* accounted for 5.1%. Neutral semantic

prosody included *indoor*, *marine*, etc., 41.9%. To sum up, *environment* presents a strong trend of combining with positive semantic prosody features, sometimes with neutral semantic characteristics but very rarely with words of negative semantic indications.

**Table 13.** The semantic prosody distribution of circumstance in COCA.

Type	Adjective Collocations (Frequency)	Total Frequency	Percentage	Semantic prosody
1	mitigating (3), fortunate (3), extraordinary (3)	9	4.8%	Positive
2	unfurnished (3), singular (4), exceptional (4), unusual (5), latter (6), special (15), historical (12), particular (18), any (33), every (11), external (4), changing (3), changed (3), unique (3), specific (7), material (4), similar (6), given (7), economic (6), local (4), individual (4), political (6)	168	89.8%	Neutral
3	aggravating (4), unfortunate (3), difficult (3)	10	5.4%	Negative

As is seen, native speakers use a variety of adjectives to modify *circumstance*, common collocation words including *special* and *particular*. In addition, the percentage of 89.8%, the highest, highlights the feature of neutral semantic prosody. The positive category percentage is 4.8%, including words like *mitigating*, *fortunate*, etc. The lowest percentage is category of negative semantic prosody, only 5.4%, which included *blend*, *unfortunate* etc. In sum, *circumstance* presents a strong trend of combining with neutral semantic

prosody features.

#### ii. The Semantic Prosody Data of Synonyms in WECCL

In WECCL 2.0, the collocations of situation, environment and circumstance ( $MI \geq 3$  and  $N \geq 2$ ) were 91, 109 and 31, the numbers of adjectives being 26, 25 and 21 respectively. In this study, again 29 ( $MI \geq 3$  and  $N \geq 2$ ) are listed. These words can be divided into groups of positive, negative and neutral semantic prosody.

Table 14. The semantic prosody distribution of situation in WECCL.

Type	Adjective Collocations (Frequency)	Total Frequency	Percentage	Semantic prosody
1	educated (1)	1	0.76%	Positive
2	current (23), present (11), serious (10), economic (6), specific (4), preset (1), different (12), real (6), same (9), similar (3), occasional (1), new (7)	93	70.99%	Neutral
3	harsh (4), difficult (10), bad (10), embarrassing (2), sub-healthy (1), getting-worse (1), touchy (1), unconscious (1), embarrassed (2), upsetting (1), unfamiliar (2), awkward (1), troubling (1)	37	28.25%	Negative

If the percentage of semantic prosody of collocations is greater than 50%, the trend is obvious. In the above table the percentage of neutral semantic prosody is 70.99% mainly involving words like *current*, *present*, etc. While adjectives with negative semantic prosody, such as *harsh* and *difficult*,

which are often paired with *situation*, account for 28.25%. And only 0.76 percent of the words, with positive indications, are used to describe *situation*, the most frequent one being “*educated*”.

Table 15. The semantic prosody distribution of environment in WECCL.

Type	Adjective Collocations (Frequency)	Total Frequency	Percentage	Semantic prosody
1	comfortable (19), protecting (14), good (39), living (31), better (28), peaceful (6), safe (6), studying (9), pollution-free (1), teaching (1), study-conductive (1), wally (1)	156	73.24%	Positive
2	quiet (12), new (19), changeable (2), natural (8), competitive (7), societal (1), controllable (1), homing (1), pretesting (1), variable (1), rest-conductive (1)	54	25.35%	Neutral
3	vulnerable (2), alluring (1)	3	1.41%	Negative

Chinese students use a variety of adjectives to describe *environment*. Among them, the highest proportion involves positive semantic prosody (73.91%), including words such as *protecting*, *good*, etc. The proportion of words with neutral

semantic prosody is 25.35%, such as *quiet*, *new*, etc., and along the negative, 1.41%, e.g., *vulnerable* and *alluring*. Therefore “*environment*” collocations in WECCL favor positive semantic prosody characteristics.

Table 16. The semantic prosody distribution of circumstance in WECCL.

Type	Adjective Collocations (Frequency)	Total Frequency	Percentage	Semantic prosody
1	globalizing (1), superior (1), warm (1), good (3), excellent (1), comfortable (1)	8	29.63%	Positive
2	male-dominated (1), concrete (1), changeable (1), social (2), any (2), new (2), every (2), certain (1), equal (1), free (1), private (1), working (1)	16	59.26%	Neutral
3	adverse (1), gloomy (1), dangerous (1)	3	11.11%	Negative

As is seen, adjectives modifying *circumstance* in Chinese students’ language use gather around words with neutral semantic prosody, accounting for 59.26%. *Good*, *warm* and other words with positive semantics are in the second group, with a percentage of 29.63%. The lowest percentage was negative semantic prosody, including *adverse*, *exhausted*, etc., accounting for 11.11%. Therefore *circumstance* highlights the feature of neutral semantic prosody in WECCL.

### iii. Comparative Analysis of Synonyms Semantic Prosody in COCA and WECCL

Table 17. Comparison of semantic prosody synonyms in COCA and WECCL.

	COCA	WECCL
situation	5.4% positive	0.76% positive
	74.6% neutral	70.99% neutral
	20.0% negative	28.25% negative
environment	55.0% positive	73.24% positive
	41.9% neutral	25.35% neutral
	5.1% negative	1.45% negative
circumstance	4.8% positive	29.63% positive
	89.8% neutral	59.26% neutral
	5.4% negative	11.11% negative

It can be seen that there are both differences and

similarities in the semantic prosody features of the sample synonyms from the two corpora. First, concerning the semantic prosody of *situation*, both native English speakers and Chinese students “are” neutral. However, the proportions of the two semantic prosody are not the same. Native English speakers are 7 times more likely than Chinese students to use the positive semantic of *situation*. In view of this, Chinese students may make too little use of the positive semantic intention of *situation*.

Secondly, both native speakers and Chinese students use the word *environment* positively. However, the proportions of the two semantic prosody are not the same. The proportions of positive and neutral semantic prosody by native English speakers are quite equivalent, while the proportion of positive semantic prosody by Chinese students is about 3 times that of neutral semantic prosody. In view of this, Chinese students might be overusing the positive semantic prosody of *environment*.

Finally, as far as *circumstance* is concerned, both native speakers and Chinese students favor the neutral category. However, the proportions of the two semantic prosody are not the same. Chinese students are 6 times more likely than native English speakers to use *circumstance*'s with positive

semantic features. Thus, Chinese students might be overusing the positive semantic prosody of *circumstance*.

### 5.3. Analysis of Research Findings

It is insightful to analyze reasons for the significant differences in the three aspects studied with the sample synonyms: frequency, salient collocation and semantic prosody. As illustrated above, both similarities and significant differences are suggested with COCA and WECCL, and this part attempts to seek causes underlying the differences.

First, L1 negative transfer has an undeniable impact. In terms of the cognitive basis for L2 acquisition, existing L1 knowledge inevitably influences L2 development [17] at levels of vocabulary, syntax or discourse. The second language acquisition process of Chinese students is unavoidably affected by Chinese expressions, cultural characteristics and other factors. What's more, Chinese students are used to thinking in Chinese, while English is rarely activated in the initiating of ideas. And when Chinese students are uncertain about wording in English of a specific concept, they tend to turn to Chinese for help, which results in inaccurate use of English synonyms as well as other transferred expressions.

Second, traditional English vocabulary teaching method is also one of the factors that cause significant differences. Usually, when teaching vocabulary, English teachers stress grammar and key sentences, supplemented by a few typical examples. However, students may get confused about synonyms with non-key sentences neglected in language teaching. To be specific, students do not notice or grasp the particular features of application of words or the issue of semantic prosody in specific contexts.

Third, the somehow misleading effect of English-Chinese dictionaries may also cause significant differences. Guo [18] investigated 485 college students' use of dictionaries in 3 Chinese universities, and found that 44.9% used English-Chinese dictionaries for translation and writing tasks. Thus, as one of the essential reference books for English vocabulary learning, the English-Chinese dictionary plays an indispensable role. However, different English-Chinese dictionaries have different English synonyms listed under the same Chinese heading, some of which may not help students to distinguish and analyze synonyms. Furthermore, most English-Chinese dictionaries lack the information of synonym differentiation and do not explicitly list the differences about collocation or semantic prosody. Therefore, it is very difficult for Chinese students to complete the task of choosing the right words in a specific context.

### 5.4. Implications

Language is like a dress. We vary our dresses to suit the occasion. We don't appear at a friend's silver wedding anniversary in gardening clothes, nor do we go punting on the river in a dinner-jacket [19]. Judging from Potter's words, it is important to choose the right words in different contexts.

Whether written or oral, the proper selection of synonyms is the key to expressing our intentions.

This research brings some implications to teaching. Corpus data provides easily accessible information about real language use [20]. On the one hand, when English teachers discover unconventional interlanguage collocations of words in students' oral or written tasks, they should think about the reasons for such misuses. Therefore, attention should be paid to the explanation of frequency, typical collocation and semantic prosody features in vocabulary teaching. In this way, it is possible for Chinese students to approach idiomatic expressions to the greatest possible extent, reducing misuses and effectively lessening the impact of L1 negative transfer. On the other hand, Chinese college students shall learn to use language corpus independently. Nowadays, most college students are equipped with computers, and the Internet is very convenient. And corpus is a new platform for Chinese students' learning. Many corpus retrieval tools can be used, e.g., BNC (British National Corpus), COCA, LOB (Lancaster-Oslo/Bergen Corpus), etc. By searching the corpus of native speakers, learners can observe the characteristics of native speakers' use of vocabulary, such as typical collocation and semantic prosody features.

In terms of dictionary compilation, a large number of real and effective examples in corpus provide good materials for compilers. The lexical information based on corpus can provide usage characteristics of words in different contexts, hence the possibility for the differentiation of synonyms. The collocation and semantic prosody information obtained based on the analysis of corpus data can provide information for dictionary improvement, beyond definitions of words or mere lists of synonyms, with detailed synonym difference analysis. Therefore corpus-based research provides very useful information for the future development of Chinese-English lexicography.

## 6. Conclusion

This paper is based on a comparative study of synonyms "situation, environment, circumstance" in WECCL and COCA from three aspects: frequency, collocation and semantic prosody. The research findings provide some implications. First, in terms of frequency, Chinese students tend to overuse these synonyms compared with native speakers. For instance, where a native speaker might use "linguistic context", a Chinese learner might prefer "situation" instead. Secondly, in terms of salient collocations, comparatively Chinese students prefer synonyms with plenty of semantic ambiguity with few collocation types. Thirdly, in terms of semantic prosody, on the Chinese students' side, inadequate accuracy is conspicuous as well as semantic prosody misuse. Underlying these differences are two possible factors influencing Chinese students' mastery of synonyms, respectively L1 negative transfer and the misleading effect of Chinese-English dictionaries. Finally, data-based analysis enables some suggestions for synonym teaching and dictionary compiling.

---

## References

- [1] Kennedy, G. (1991). Between and through: The company they keep and the functions they serve. In K. Aijmer and B. Altenberg (Eds.), *English Corpus Linguistics: Studies in honour of Jan Svartvik* (pp. 95-110). London: Longman.
- [2] Saeed, J. (2000). *Semantics*. Beijing Foreign Language Teaching and Research Press.
- [3] Harward, J & Etienne, Z. (2000). *Words, Meaning and Vocabulary*. London: Wellington House.
- [4] Hoey, M. (2005). Lexical Priming: A New Theory of Words and Language. London & New York: Routledge. *International Journal of Lexicography*, 19 (3): 327-335.
- [5] Shahzadi, A., Asghar, A. & Javed, A. (2019). Effectiveness of Corpus in Teaching English Synonyms. *Corporum: Journal of Corpus Linguistics- CJCL*, 2 (1): 51-65.
- [6] Lu, J. (2010). Corpus-Based Study on the Collocation Features and Semantic Prosody In EFL Learners' Use of Synonyms. *Modern Foreign Languages* 33 (3): 276-286.
- [7] Sun, L. (2017) A New Interpretation of Semantic Prosody Features in Chinese EFL Learners Use of Synonyms ---Comparing *affect* and *influence*. *Contemporary Foreign Languages Studies* (1): 57-61.
- [8] He, A. P. (1997). On Corpus Linguistics. *Journal of Foreign Languages* (5): 22-27.
- [9] Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- [10] Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- [11] Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair* (pp. 157-176) Amsterdam: John Benjamins.
- [12] Stubbs, M. (1995). Collocations and semantic profiles: On the cause of trouble with quantitative studies: *Functions of language*, 2 (1): 23-25.
- [13] Sinclair, J. (1996). The Search for Units of Meaning. *Textus IX*: 75-106.
- [14] Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford Blackwell Publishers Ltd.
- [15] Wen, Q. F. (2008). *Spoken and Written English Corpus of Chinese Learners (2.0)*. Foreign Language Teaching and Research Press.
- [16] Liang, M. C. (2010). *Using Corpora: A Practical Coursebook*. Foreign Language Teaching and Research Press.
- [17] Ellis, R. (1999). *The Study of Second Language Acquisition*. Shanghai: Shanghai Education Press.
- [18] Guo, J. (2007). A Survey on the Needs and Use of College English Dictionary. *Education and Vocation* (12): 168-169.
- [19] Potter, S. (1959). *Our Language*. London: William Clowes and Sons Ltd.
- [20] Thornbury, S. (2002). *How to Teach Vocabulary*. Essex: Pearson Education Limited.