

A Corpus-assisted Study of Synonyms in EFL Teaching: Take *Preserve* and *Conserve* as Examples

Enyao Li

School of Foreign Studies, Guangzhou University of Chinese Medicine, China

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Synonyms, as an important part of English vocabulary, have always been a key but difficult point in EFL (English as a foreign language) teaching. In the traditional vocabulary teaching, teachers mainly distinguish synonyms via describing, translating and explaining with examples in the dictionary. However, most dictionaries can only provide separate word meanings and limited examples without explaining the different usage and meanings of the synonyms. Therefore, drawing upon Sinclair's model of extended lexical units (ELUs) and employing a corpus-assisted approach to the study of synonyms *preserve* and *conserve*, this study tries to analyse the similarities and differences of these synonyms from the perspectives of their frequency, colligation, collocation, semantic preference and semantic prosody. The findings show that corpus-assisted approach is more effective, objective and reliable in distinguishing synonyms and a number of implications can be concluded in English language teaching, learning, testing and research.

Keywords Corpus-assisted, Synonyms, EFL Teaching

1. Introduction

Vocabulary learning is the premise of improving English skills such as listening, speaking, reading and writing. Synonyms, as an important part of English vocabulary, have always been a key but difficult point in EFL (English as a foreign language) teaching. In the traditional vocabulary teaching, teachers mainly distinguish synonyms via describing, translating and explaining with examples in the dictionary. However, most dictionaries can only provide separate word meanings and limited examples without explaining the different usage and meanings of the synonyms, which would make students' confused. Therefore, the traditional way of teaching English

synonyms cannot meet the needs, and it is urgent to seek a more efficient and effective way. With the development of computer technology in the era of Big Data, corpus plays an increasingly important role in the study of linguistics and language teaching. Due to its large-capacity, high-speed and effectiveness, corpus can collect a large amount of verbal or written data and provide rich authentic usage information, which can help analyse the vocabulary meaning and usage in various aspects.

Drawing upon Sinclair's model of extended lexical units (ELUs) [1] [2] and employing a corpus approach to the study of synonyms *preserve* and *conserve*, the present study tries to analyse the similarities and differences of these synonyms from the perspectives of their frequency, colligation, collocation, semantic preference and semantic prosody, aiming to help teachers and learners better grasp the specific usage of synonyms by referring to the authentic corpus of native language users.

2. Literature Review

Synonymy is a significant but complicated sense relation in language and generally refers to "the phenomenon of two or more different linguistic forms with the same meaning" [3]. In this sense, synonyms can be considered as words or phrases that share the exact or nearly the same meaning. There are abundant previous studies on synonyms in different disciplines, such as lexicology, pragmatics, semantics, cognitive linguistics, and corpus linguistics. In this section, we will focus on the corpus approach to the study of synonyms.

Miller & Charles [4] examine the relationship between synonyms and substitutability on the basis of the Brown Corpus. Church et al. [5] conduct a corpus study of the synonyms *ask for*, *request* and *demand* and arrive at similar conclusions as Miller & Charles' [4]. Gries [6] investigates similarities and differences between English adjectives ending in *-ic* or *-ical* and focuses on the aspect of the overlap of their collocations. Besides, there are many other

synonyms that have attracted researchers' attention, for example, *urge* and *incite* [7], *big*, *large* and *great* [8], *get* and *have* [9], and *actually*, *genuinely*, *really*, and *truly* [10]. However, there are fewer studies on the corpus-approach to English synonyms by means of the Sketch Engine [11] [12].

3. Data and Research Methods

This section will mainly introduce the corpus data BNC (Section 3.1), corpus tool Sketch Engine (Section 3.2) and research procedures (Section 3.3) of this study.

3.1. Corpus Data

The corpus employed in the present study is the British National Corpus (BNC) which is a 100-million-word collection of language samples from a wide range of sources (e.g. academic books, magazines, newspapers, periodicals, and transcriptions of conversations), representing a wide cross-section of British English from 1980s to early 1990s, both spoken and written [13]. The nature of BNC is a sort of monolingual, synchronic, general and sample-based corpus, and it is "encoded according to the Guidelines of the Text Encoding Initiative (TEI) to represent both the output from CLAWS (automatic part-of-speech tagger) and a variety of other structural properties of texts (e.g. headings, paragraphs, lists etc.)" [12]. Besides, online BNC corpus is free access to the retrieval service, and it is very practical and useful corpus in language teaching, linguistic research, and natural language processing [13].

3.2. Corpus Tool

The corpus tool employed here is the Sketch Engine (SkE), which is a leading corpus tool and widely used in lexicography, language teaching and other research areas [14] [15]. SkE is "a corpus query system (CQS) which allows the user to view word sketches, thesaurally similar words, and 'sketch differences', as well as the more familiar CQS functions" [14] and it refers to two different aspects: the software and the web service which include "a large number of corpora pre-loaded and 'ready for use', and tools for creating, installing and managing your own corpora" [14]. This corpus tool has some core functions,

such as *Word Sketch*, *Word Sketch Difference*, *Thesaurus*, *Concordance*, *Wordlist*, and *Keywords*, most of which will be applied in this study and will be explained in the following sections.

3.3. Research Procedures

Based on the online BNC corpus and pre-loaded BNC corpus in the Sketch Engine, the main research steps are as follows:

- 1) Identify the similarity of the chosen words *preserve* and *conserve* by the "Thesaurus" function of Sketch Engine.
- 2) Analyse the overall frequency of the two synonyms and their distribution of genres by using BNC online search tool.
- 3) Extract the colligation and collocation by the "Word Sketch" and "Sketch Diff" functions of Sketch Engine.
- 4) Summarise the semantic preference and semantic prosody by examining the collocations.

4. Results and Discussion

This section will present the results of this study from the follow perspectives: similarity between *preserve* and *conserve* (Section 4.1), word frequency and genre distribution (Section 4.2), colligation (Section 4.3), collocation (Section 4.4), semantic preference and semantic prosody (Section 4.5).

4.1. Similarity between *Preserve* and *Conserve*

Since these two words *preserve* and *conserve* are from textbook exercises, we need to confirm their similarity before starting the analysis. SkE has the *Thesaurus* function, which can generate the synonyms immediately after you input the word. Figure 1 shows the results of the verb *conserve's* synonyms. As is clearly presented, *preserve* ranks the third, which means *preserve* and *conserve* are not most strongly synonymous. However, when examining the first two words (*safeguard* and *exhaust*), we can see it is much easier to distinguish *conserve* from them. Given the score, frequency and distinction, *conserve* and *preserve* are selected as examples in the present study.



Figure 1. Similarity between *preserve* and *conserve*

4.2. Word Frequency and Genre Distribution

In order to fully demonstrate the distribution of *preserve* and *conserve*, the frequency is counted separately through LIST function of BNC corpus, as illustrated in Table 1. The statistics shows that the total frequency of *preserve* and *conserve* is quite different: the frequency of *preserve* (3651) is over six times higher than that of *conserve* (604). In other words, *preserve* is much more frequently used than *conserve*. The vocabulary’s overall frequency is important, but it is of little help in understanding the word meaning and usage. Therefore, other retrieval functions need to be employed to gain a deeper understanding of these two synonyms.

Genre is generally defined as “a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or

academic community in which it regularly occurs” [16]. The BNC corpus can provide a wide range of genres, such as spoken, fiction, magazines, newspapers, academic. In order to analyse the genre distribution of *preserve* and *conserve*, we calculate the frequency through the CHART function of online BNC corpus, which is shown in Table 2.

Table 1. Frequency of *preserve* and *conserve* in BNC

preserve		conserve	
CONTEXT	FREQ	CONTEXT	FREQ
preserved	1554	conserve	311
preserve	1459	conserved	168
preserving	497	conserving	117
preserves	141	conserves	8
TOTAL	3651	TOTAL	604

Table 2. Genre distribution of *preserve* and *conserve* in BNC

GENRES	WORDS (M)	<i>preserve</i>		<i>conserve</i>	
		FREQ	PER MIL	FREQ	PER MIL
ACADEMIC	15.3	1054 (28.87%)	68.75	139 (23.01%)	9.07
NON-ACAD	16.5	891 (24.4%)	54.02	147 (24.34%)	8.91
MISC	20.8	895 (24.51%)	42.96	185 (30.63%)	8.88
MAGAZINE	7.3	224 (6.14%)	30.85	50 (8.28%)	6.89
NEWSPAPER	10.5	267 (7.31%)	25.51	42 (6.95%)	4.01
FICTION	15.9	235 (6.44%)	14.77	26 (4.3%)	1.63
SPOKEN	10	85 (2.33%)	8.53	15 (2.48%)	1.51
TOTAL	100	3651 (100%)	36.51	604 (100%)	6.04

In terms of the normalised frequency (PER MIL), both *preserve* and *conserve* can appear in all these seven genres, with the same rank from ACADEMIC, NON-ACAD, MISC, MAGAZINE, NEWSPAPER, FICTION to SPOKEN. On the basis of the statistics, both *preserve* and *conserve* are more likely to be used in written genres rather than in spoken one, but *preserve* is always more frequently used than *conserve* in all the above-mentioned genres.

In considering the percentage of frequency within each word (*preserve* or *conserve*), we can find more similarities and differences. *Preserve* has the highest frequency in academic genres (28.87%) and the lowest frequency in spoken genres (2.33%) whereas *conserve* has the highest frequency in miscellaneous genres (30.63%) and the lowest frequency in spoken genres (2.48%). Overall contrast shows that *preserve* is more frequently used in academic, newspaper and fiction genres while *conserve* is more frequently used in magazine and miscellaneous genres. Besides, they have almost the same frequency in non-academic and spoken genres.

4.3. Colligation Analysis

In corpus linguistics, colligation refers to “the relation of co-occurrence between the node and abstract grammatical categories” [17]. In practical language use, colligation is not just confined to co-occurrence with particular parts of speech, but also includes the “patterns of consistent

co-occurrence of a word with different syntactic contexts” [18].

As Kilgarriff et al. [14] argues, word sketch is “a one-page summary of a word’s grammatical and collocational behaviour”. In the Sketch Engine, we can get the colligation and collocation categories of automatically and quickly by the function of “Word Sketch” and “Word sketch difference”, as exemplified in Figure 2. This section will focus on the colligation categories of *preserve* and *conserve* in BNC, whose frequency and score (logDice coefficient) are summarised in Table 3.

As is demonstrated in the table, there are altogether 14 grammatical relations identified automatically by the corpus tool, 12 for *preserve* and 7 for *conserve*. The table also shows that *preserve* and *conserve* share 5 colligation patterns, including *object (V+N)*, *modifier (ADV+V)*, *pp_in-p*, *subject (N+V)*, and/or *(V+V)*, among which *preserve* tend to more frequently occur in the patterns of *modifier*, *pp_in-p*, *subject* while *conserve* in *object*, and/or patterns. In terms of these sharing patterns, we can identify their common grammatical features: both words are more likely to be used as transitive verbs. Besides, there are 7 patterns of *pp_by-p*, *pp_for-p*, *pp_as-p*, *pp_at-p*, *adj_comp*, *np_adj_comp*, and *pp_to-p* exclusively applied in *preserve* whereas only two patterns of *pp_through-p* and *part_over-a_obj* exclusively used in *conserve*, which means that *preserve* has more flexible and richer colligation patterns than *conserve*.



Figure 2. Colligation and collocation results of *preserve* and *conserve* in SKE (partial screenshot)

Table 3. Colligation categories of *preserve* and *conserve* in BNC

Grammatical relations	<i>preserve</i>		<i>conserve</i>		Examples (a. <i>preserve</i> ; b. <i>conserve</i>)
	FREQ	SCORE	FREQ	SCORE	
object	2596	69.95	432	70.24	a. ...clearly is a vital thing to <u>preserve the integrity</u> of good name... b. ...which needs to <u>conserve moisture</u> in the hot atmosphere...
modifier	572	15.41	78	12.68	a. ...and it is easier to find beautifully preserved examples of the older Rugosa... b. ...an amino acid sequence that seems to be highly conserved in nature with...
pp_in-p	295	7.95	28	4.55	a. ...whose works are preserved in the Epistulae Austrasiacae... b. ...which are absolutely conserved in all the other POU proteins...
subject	289	7.79	45	7.32	a. ...the gospel tradition preserves explicit commissions... b. ...and about half of the remaining residues conserved...
and/or	188	5.07	54	8.78	a. ...bureaucratic power was used to preserve and strengthen capitalism and the interests... b. ...taken by the architects to conserve and enhance the qualities of...
pp_by-p	66	1.78	0	0	a. ...is now preserved by a trust as the Kew Bridge Living Steam Museum...
pp_for-p	60	1.62	0	0	a. ...where the forest is preserved for the use of its...
pp_as-p	44	1.19	0	0	a. ...and they are easily preserved as fossils...
pp_at-p	27	0.73	0	0	a. ...they should be preserving at all costs...
adj_comp	20	0.54	0	0	a. ...the basic melody is generally preserved intact...
np_adj_comp	19	0.51	0	0	a. ...and she wished to <u>preserve her experience intact</u> in order to...
pp_to-p	7	0.19	0	0	a. ...a habit preserved to this day...
pp_through-p	0	0	3	0.49	b. ...and it is <u>conserved through</u> primate evolution...
part_over-a_obj	0	0	2	0.33	b. ...are very well conserved over large evolutionary distances...

4.4. Collocation Analysis

Collocation is an old term and has been defined in various ways. In the twentieth century, Firth [19] argued that “you shall know a word by the company it keeps” (p. 11) and brought the idea of collocation into its modern form, which is also called “an abstraction at the syntagmatic level” [20]. Sinclair [1] further develops the concept of collocation and describes it as “the occurrence of two or more words within a short space of each other in a text”. In other words, collocation is taken as “a co-occurrence pattern that exists between two items that frequently occur *in proximity* to one another-but not necessarily adjacently or, indeed, in any fixed order” [18].

The previous Section 4.3 shows that two verbs *preserve* and *conserve* share a number of grammatical and syntactical patterns, but the collocates in each pattern are

quite different. On the basis of the colligation statistics and the verb characteristics (Table 3 and Figure 2), this section will mainly focus on the analysis of collocates in the patterns of *object*, *modifier*, *subject*, and/or, which are respectively shown in Tables 4-7.

In the “object” pattern (V+N), there are more words used as objects of *preserve* than *conserve*, as is indicated in Table 4. Words like integrity, quo, anonymity, unity, independence, fossil, confidentiality, peace, neutrality, identity only collocate with *preserve* while words like heat, motif, nutrient, energy, epitope, Hd, residue, biodiversity, moisture only collocate with *conserve*. There are some words that can collocate with both verbs, such as specimen, heritage, hedgerow, symmetry, countryside, habitat, but they have different frequencies, for example, specimen is more frequently used with *preserve* while countryside is more frequently used with *conserve*.

Table 4. Collocates in “object” pattern of *preserve* and *conserve* in BNC

No.	Collocates	Freq		Score	
		<i>preserve</i>	<i>conserve</i>	<i>preserve</i>	<i>conserve</i>
1	integrity	32	--	7.6	--
2	quo	23	--	7.6	--
3	anonymity	13	--	7.1	--
4	unity	23	--	6.7	--
5	independence	26	--	6.4	--
6	fossil	11	--	6.3	--
7	confidentiality	8	--	6.2	--
8	peace	30	--	6.1	--
9	neutrality	7	--	6	--
10	identity	19	--	5.9	--
11	specimen	20	2	6.5	3.6
12	heritage	13	4	6	4.9
13	hedgerow	5	2	5.7	6
14	symmetry	3	2	4.7	5.3
15	countryside	12	15	5.5	6.2
16	habitat	5	6	5	6.1
17	heat	--	9	--	5.1
18	motif	--	3	--	5.2
19	nutrient	--	2	--	5.4
20	energy	--	42	--	6
21	epitope	--	2	--	6.8
22	Hd	--	2	--	7
23	residue	--	12	--	7.4
24	biodiversity	--	3	--	7.4
25	moisture	--	8	--	7.5

In the “*modifier*” pattern (*ADV+V*), Table 5 shows that there are many words (such as *beautifully*, *exquisitely*, *superbly*, *best*, *perfectly*, *marvellously*) that only collocate with *preserve* and there are several words (such as *highly*, *properly*) that only collocate with *conserve*. Words that can

collocate with both verbs include *well*, *either*, *poorly*, *absolutely* with different frequencies, among which *well* is prominent due to its much higher frequency of collocating with *preserve* than *conserve* and thus the idiomatic expression “*well preserved*” can be identified.

Table 5. Collocates in “*modifier*” pattern of *preserve* and *conserve* in BNC

No.	Collocates	Freq		Score	
		<i>preserve</i>	<i>conserve</i>	<i>preserve</i>	<i>conserve</i>
1	beautifully	21	--	8.7	--
2	exquisitely	5	--	7.8	--
3	superbly	6	--	7.8	--
4	best	37	--	7.7	--
5	perfectly	18	--	7.2	--
6	marvellously	3	--	7.1	--
7	magnificently	3	--	7.1	--
8	splendidly	3	--	7.1	--
9	thereby	12	--	7.1	--
10	lovingly	3	--	7	--
11	expressly	4	--	7	--
12	zealously	2	--	6.8	--
13	finely	4	--	6.8	--
14	religiously	2	--	6.7	--
15	preferentially	2	--	6.7	--
16	excellently	2	--	6.7	--
17	jealously	2	--	6.6	--
18	well	76	6	6.3	2.7
19	either	3	2	3.5	3
20	poorly	3	2	6.2	6.3
21	absolutely	2	3	3.9	4.6
22	completely	--	2	--	3.3
23	strongly	--	2	--	4
24	properly	--	2	--	4.2
25	highly	--	35	--	7

In the “*subject*” pattern (*N+V*), Table 6 shows the usage differences of these two words: there are a large number of words (such as *trust*, *tradition*, *museum*) that can collocate with *preserve*; by contrast, there are only three words

(*domain*, *soil*, *residue*) that can collocate with *conserve*. And we can observe that the collocates of *preserve* and *conserve* are more likely to be mutually exclusive in this grammatical pattern.

Table 6. Collocates in “*subject*” pattern of *preserve* and *conserve* in BNC

No.	Collocates	Freq		Score	
		<i>preserve</i>	<i>conserve</i>	<i>preserve</i>	<i>conserve</i>
1	mud	2	--	4.5	--
2	separation	2	--	4.2	--
3	transformation	2	--	4.1	--
4	timber	2	--	4	--
5	trust	7	--	3.8	--
6	designer	2	--	3.6	--
7	tradition	4	--	3.3	--
8	help	5	--	3.2	--
9	english	3	--	3.1	--
10	Edward	2	--	3	--
11	museum	3	--	2.6	--
12	God	3	--	2.3	--
13	means	2	--	2.2	--
14	text	2	--	2.1	--
15	church	4	--	1.8	--
16	method	3	--	1.7	--
17	resource	2	--	1.4	--
18	provision	2	--	1.4	--
19	family	5	--	1.3	--
20	act	3	--	1.2	--
21	section	2	--	1	--
22	rule	2	--	1	--
23	domain	--	2	--	4.1
24	soil	--	6	--	4.7
25	residue	--	2	--	5.1

In the “and/or” pattern (*V and/or V*), most words (such as *strengthen, recover, promote, create*) are exclusively used with *preserve* while only two words (*manage, collect*) are exclusively used with *conserve*. Words that collocate

with both words include *protect, restore, enhance, maintain, improve* which have different frequencies. For example, *protect/enhance* is more frequently used with *preserve* while *improve* is more often used with *conserve*.

Table 7. Collocates in “and/or” pattern of *preserve* and *conserve* in BNC

No.	Collocates	Freq		Score	
		<i>preserve</i>	<i>conserve</i>	<i>preserve</i>	<i>conserve</i>
1	pickle	2	0	7.9	--
2	bolster	2	0	6.4	--
3	strengthen	8	0	5.7	--
4	dry	3	0	5.2	--
5	recover	7	0	5.2	--
6	rebuild	2	0	5.1	--
7	guard	2	0	4.9	--
8	promote	7	0	4.4	--
9	secure	4	0	3.8	--
10	expand	2	0	3.6	--
11	defend	2	0	3.4	--
12	incorporate	2	0	3.1	--
13	create	7	0	2.6	--
14	record	3	0	2.6	--
15	identify	3	0	2.2	--
16	develop	3	0	1.4	--
17	present	2	0	1.3	--
18	pass	2	0	1.2	--
19	protect	15	5	5.2	3.6
20	restore	6	2	5	3.5
21	enhance	9	4	5.7	4.5
22	maintain	3	2	2.3	1.7
23	improve	2	4	2.1	3.1
24	manage	0	2	--	2
25	collect	0	2	--	2.4

4.5. Semantic Preference and Semantic Prosody

Both semantic preference and semantic prosody are related to the collocates of the selected words. Semantic preference refers to “the relation, not between individual words, but between a lemma or word-form and a set of semantically related words” [21]. An example of the verb *set in* was provided by Sinclair [1] to demonstrate its negative associations with the surrounding words, i.e. negative prosody. Stubbs [22] further divides semantic prosody into three types: positive, negative and neutral. Semantic prosody is mainly explored “by analysis of a concordance” [23] and exists “at the subconscious level” [24].

When it comes to the semantic preference of *preserve* and *conserve*, the “object” pattern is more prominent due to its frequency and the notional nouns in it. Through comparing the collocates (Tables 4 & 5) of *preserve* and *conserve*, we can arrive at the following conclusions: *preserve* has the semantic preference of something that is abstract but important, such as *integrity*, *anonymity*, *unity*,

independence and means “keep something as it is, without making any changes”. By contrast, *conserve* has the semantic preference of something that is related to natural resources, environment and ecology, such as *moisture*, *biodiversity*, *energy*, *heat*, with the meaning of “use as little of something as possible so that it lasts a long time”.

On the basis of McEnery and Hardie’s claim that “words or phrases are said to have a negative or positive semantic prosody if they typically co-occur with units that have a negative or positive meaning” [18], we will mainly discuss semantic prosody of collocates of *preserve* and *conserve* in the “object” and “modifier” patterns (Tables 4 & 5). By examining the concordance lines of these collocates, the statistical evidence will be provided to ensure the accuracy of interpretation, which is shown as follows in Figures 3 & 4.

In the “modifier” patterns, statistics in Figure 3 shows that the semantic prosody in *preserve* and *conserve* is similar: they create a positive semantic prosody in most cases, sometimes neutral, seldom negative. In negative semantic prosody, there is only one case *poorly identified*.

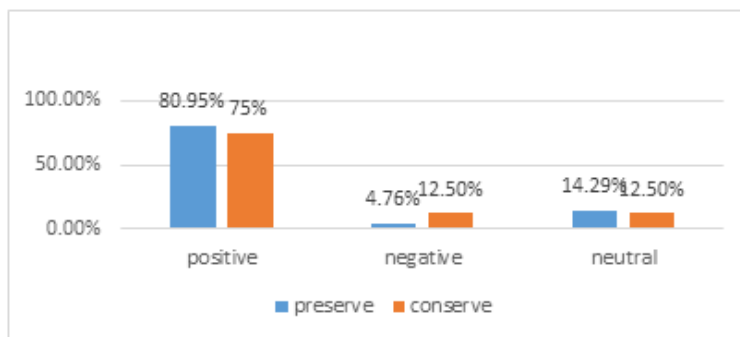


Figure 3. Semantic prosody of *preserve* and *conserve* in “modifier” pattern

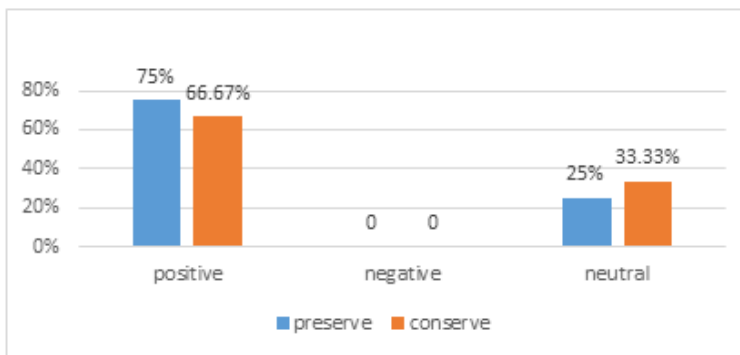


Figure 4. Semantic prosody of *preserve* and *conserve* in “object” pattern

As Figure 4 indicates, the semantic prosody of *preserve* and *conserve* in the “*object*” patterns is also similar: they create a positive semantic prosody in most cases, sometimes neutral, but never negative.

Combining the findings of semantic prosody in both “*modifier*” and “*object*” patterns, we can conclude that the overall semantic prosody created around *preserve* and *conserve* tend to be more positive or neutral, seldom negative. Referring to the above analysis, we can even say that the negative semantic prosody is generally viewed as marked.

5. Conclusions and Implications

This study mainly adopts a corpus-assisted approach to the study of synonyms in EFL teaching, taking *preserve* and *conserve* as examples in order to find out their similarities and differences in word frequency, genre distribution, colligation, collocation, semantic preference and semantic prosody, as previously discussed in Section 4. The findings have shown that corpus-assisted approach is more effective, objective and reliable in distinguishing synonyms and a number of implications can be concluded in English language teaching, learning, testing and research.

For teachers, rich corpora and various corpus tools can provide them with abundant materials on synonyms which could solve the problems existing in traditional ways of teaching and testing. Besides, corpus can offer new insights into the relationship between corpus linguistics and language teaching, which is especially helpful to many English teachers under the great pressure of doing research.

For learners, authentic sources from corpora can help them cultivate the consciousness of genre, colligation, collocation, semantic preference and semantic prosody. Since there are so many synonyms in English that it would be impossible for teachers to teach all of them, students should be able to learn how to make use of online corpora and corpus tools (such as Sketch Engine) to do their own analysis.

Although this study has its significance in many aspects of EFL teaching, there are still some limitations. By only examining synonymous words in the BNC, this study did not describe the Chinese learners’ actual use of synonyms. If two corpora are compared, we can get a more targeted and effective understanding of the similarities and differences between native language users and Chinese ones. In view of this, future studies can be concentrated on comparative analysis of synonyms in both native English speakers’ corpus and Chinese English learner’ corpus.

Acknowledgements

This study has been supported by the 2017 Project of Higher Education Teaching and Reform funded by

Guangzhou University of Chinese Medicine and the Project funded by Guangzhou Planning Office of Philosophy and Social Sciences (No. 2017GZGJ60).

REFERENCES

- [1] Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- [2] Sinclair, J. (2004). *Trust the text language, corpus and discourse*. London and New York: Routledge.
- [3] Stanojević, M. (2009). Cognitive synonymy: A general overview. *FACTA UNIVERSITATIS (Linguistics and Literature)*, 7(2), 193-200.
- [4] Miller, G. A., & Charles, W. G. (1991). Contextual Correlates of Semantic Similarity. *Language and Cognitive Processes*, 6(1), 1-28.
- [5] Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Exploiting on-line resources to build a lexicon* (pp. 115-164). Hillsdale, NJ: Lawrence Erlbaum.
- [6] Gries, S. (2001). A corpus-linguistic analysis of English *-ic* vs. *-ical* adjectives. *ICAME Journal*, 25(Dec), 65-108.
- [7] Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1(1), 75-98.
- [8] Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- [9] Gilquin, G. (2003). Causative *get* and *have*: So close, so different. *Journal of English Linguistics*, 31(2), 125-148.
- [10] Liu, D., & Espino, M. (2012). *Actually, Genuinely, Really, and Truly*: A corpus-based Behavioral Profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics*, 17(2), 198-228.
- [11] Hong, J. F. (2014). Chinese near-synonym study based on the Chinese gigaword corpus and the Chinese learner corpus. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8922, 329-340.
- [12] Hu, C., & Yang, B. (2015). Using Sketch Engine to Investigate Synonymous Verbs. *International Journal of English Linguistics*, 5(4), 29-41.
- [13] Leech, G. N. (1993). 100 million words of English: the British National Corpus (BNC). *English Today*, 9(1), 9-15.
- [14] Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In *Proceedings of Euralex, Lorient, France*, July 2004, pp. 105-116.
- [15] Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychly, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography ASIALEX*, 1(1), 7-36.
- [16] Bhatia, V. K. (1993). *Analysing genre: language use in professional settings*. London: Longman.

- [17] Stubbs, M. (2009). Memorial Article: John Sinclair (1933-2007): The Search for Units of Meaning: Sinclair on Empirical Semantics. *Applied Linguistics*, 30(1), 115-137.
- [18] McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- [19] Firth, J. R. (1957). *Papers in Linguistics 1934-51*. London: Oxford University Press.
- [20] Firth, J. R. (1968). A synopsis of linguistic theory 1930-55. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth, 1952-59* (pp. 168-205). London: Longmans.
- [21] Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- [22] Stubbs, M. (1996). *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell Publishers.
- [23] Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: in honour of John Sinclair* (pp. 157-176). Amsterdam: John Benjamins.
- [24] Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.