

# A Semantic Prosody Analysis on Two Synonymous Pairs in English Native Speakers' and Chinese Learners' Writings

Zhou Bi<sup>1</sup>

<sup>1</sup> Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Zhou Bi, Guangdong University of Foreign Studies, No. 2 Baiyun Avenue, Guangzhou City, China.

Received: June 1, 2019 Accepted: June 30, 2019 Online Published: July 2, 2019

doi: 10.5539/elt.v12n8p14 URL: <https://doi.org/10.5539/elt.v12n8p14>

## Abstract

The concept of semantic prosody has attracted great research interest in language teaching. Identifying learners' perception of semantic prosody and collocation may be beneficial to vocabulary teaching. This study analyzes two pairs of synonyms in English writings of Chinese students and English native speakers based on the ICNALE corpus. The analysis finds that Chinese writers' perception of semantic prosody is mostly consistent with that of native speakers. However, the use of collocation is quite different between them.

**Keywords:** semantic prosody, collocation, language teaching, vocabulary teaching

## 1. Introduction

This study explores the use frequency, collocation and semantic prosody of two pairs of synonyms by comparing words in Chinese learners' writings (CHN) with English native speakers' writings (ENS). The purpose of the study is to investigate whether there is difference in the use of words, collocation and most importantly, semantic prosody in Chinese students and English native speakers. The study is motivated by the research done by Xiao and McEney (2006), which compares the collocation and semantic prosody of near synonyms of Chinese and English. The research finds out that the near synonyms are normally not interchangeable in inter-languages as well as intra-languages because of different semantic prosody. Therefore, it might lead to difficulty in learning.

### 1.1 Previous Studies on Semantic Prosody

Sinclair (1991) proposed the idea that "many uses of words and phrases show a tendency to occur in a certain semantic environment". Later in 1993, Louw introduced the term "semantic prosody", then he conceptualized it as "a form of meaning which is established through the proximity of a consistent series of collocates." Up to now, many literature has been published on identifying the semantic prosody of certain words (Bednarek, 2008; Oster, 2010; Walker, 2011) and the cross-linguistic difference of semantic prosody (Wei & Li, 2014). However, many studies focus on the potential difference in semantic prosody. Few are done to investigate the actual difference between native speaker and foreign language learners.

Besides, the previous studies are mostly conducted based on a large general corpus, which could only reflect the semantic prosody in the more frequent environment where the word occurs. However, it is always found that words vary in meaning when occur in different contexts, that is to say, a word could occur not only in positive or only in negative semantic prosody. What matters is not to find out its possible semantic prosody but to judge in what condition it could be/should be used in negative or positive environment, especially for language learners.

### 1.2 Research Question

Based on discussion above, this study tries to answer the following question:

Is there difference in words' frequency of use, collocation and semantic prosody between Chinese learners' writings and English native speakers' writings?

To answer this question, the study will take both Chinese learners' writings and English native speakers' writings as analyzing source. The writings should focus on the same topic. The key issue of this study does not lie in finding out the semantic prosody of certain words, but in the actual performance of learners and native speakers. Therefore, as the writings are limited in a certain topic, the analysis of target words may not correspond to the result obtain from the general corpus.

## 2. Methodology

The corpus adopted in the present study is the International Corpus Network of Asian Learners of English (ICNALE), which is an international learner corpus developed by Dr. Shin Ishikawa (2013), Kobe University. The corpus comprises four modules: Spoken Monologue, Spoken Dialogue, Written Essays and Edited Essays. The sub-corpus of Written Essays contains around 130,000,000 tokens, which is of a relatively large size. The ICNALE rigidly controls prompt and tasks. Parameters, such as time for writing, the length of an essay, are well controlled, which helps to guarantee the reliability of a contrastive analysis. The ICNALE team has required all the learners to take Vocabulary Tests, English proficiency tests, such as TOEFL, as well as learner background survey, while ensuring the representativeness of the learners' corpus. The corpus includes the production data by native speakers of English, with a relative balance in the nationalities.

As this study tries to compare the writings of Chinese students with those of English native speakers, it only takes data from the Written Essays module to form a new corpus. The corpus in question is extracted by selecting writings from two topics (checking PTJ and SMK boxes) in CHN (checking Chinese students in all proficiency levels) and ENS (checking English Native Speakers in all types: students, teachers and else). There are only two topics, which are: (a) It is important for college students to have a part-time job (PTJ), and (b) Smoking should be completely banned at all the restaurants in the country (SMK). Table 1 shows the composition of this sub-corpus specifying the number of tokens in each topic category and participant category.

Table 1. Composition of the under-investigated corpus (number of tokens)

	CHN	ENS	sum
PTJ	110,224	4,5231	155,455
SMK	93,151	45,121	138,272
sum	203,375	90,352	293,727

Note: CHN: Chinese participants; ENS: English native speakers; PTJ: topic on part-time job; SMK: topic on smoking.

Although the number of tokens in Chinese learner's writings is much more than that in English native speakers, it is possible to compare their frequency difference by using Chi-square and log-likelihood ratio.

### 2.1 The Targeted Words

These words pairs were chosen based on the following criteria: (1) the frequency cut-off for each word form is 40 times per million words. Referring to other studies, this cut-off point is frequently used and it is also practical in this study. As the sub-corpus contains only 293,727 tokens in total, the raw frequency for the word form is around 12 times; words with lower frequency would increase the risk of inaccuracy for the present study thus they are sifted out; (2) their meanings are checked in the Collins Thesaurus and Webster-Merriam Online Dictionaries to make sure they share the same denotational meaning; (3) those words with more than one meanings are not carefully selected, which means that only those examples with related meaning to this study are included (e.g. *grow* will be selected only if it means "to increase or expand" but not "to develop to maturity"). Based on the criteria mentioned above, words chosen for this study are two groups of synonymous words chosen for comparison in this study are *increase, rise* and *decrease, reduce*.

### 2.2 Analysis

This study will compare the word frequency of usage, collocation and different semantic prosody in Chinese learners' and English native speakers' writings in the following ways.

The numbers of each word occurring in ENS and CHN writings are counted. As the numbers of writings in ENS and CHN are not equal, the frequencies will be standardized and compared their significance in difference by a Chi-square test and a log-likelihood ratio. This is to decide the different use frequency between learners' writings and native speakers' writings. The reason why this study takes it into consideration is that the difference in use frequency might shadowing different understanding in usage and semantic prosody.

The frequently occurred collocation will be selected within the span of 4 words to both the left and the right, with a mutual information of 3 (MI3) are selected. As this is a small corpus, it is conservative not to include low frequency words as collocation but as words occurring. Therefore, the methods of Z-score or MI is not adopted.

Instead, this study will consider MI3 and log-log (more suitable for detecting content words) as a reference. Hunston (2002: 71) suggested that collocation with an MI3 of 3 or higher can be regarded as significant.

The semantic pattern will be discerned after looking at each concordance line. Each concordance line will be examined as positive, negative or neutral based on its linguistic context. Those words with more than 100 concordance lines would be selected 100 randomly by excel (it is found in this study that no word has occurred more than 100 times). Ten percent of the concordance lines of each words will be randomly extracted and rated by another rater. If there is disagreement on certain line, a third rater would be involved. As Chinese learners' proficiency level varies, there might be some vague expressions in the writing which are hard to judge its semantic prosody. In this case, that concordance line would be eliminated. The rating criteria is as follows: (1) the rater should judge the sentence semantic prosody based on the meanings of its collocation or words on both sides, regardless of the intended meaning. For example, "reduce the cancer rate" is a good thing in general sense, but "cancer" is a negative thing. Therefore, the word "reduce" occur in a negative semantic environment and it should be counted as one example of negative semantic prosody.

### 3. Results and Discussion

All the concordance lines are evaluated based on the criteria, ten percent of which are evaluated by another rater, who is familiar with the criteria.

#### 3.1 Frequency Difference

Table 2 shows the different frequencies of each word occurring in ENS and CHN writings. As presented in the previous section, the number of tokens of CHN is much larger than that of ENS, it is necessary to compare the two by examine their significance level, rather than comparing their raw frequencies. Therefore, a chi-square and a log-likelihood are calculated to check if there is significant difference between the two. The table shows that the only word found to be different in frequency is *reduce* (with a significant level of 0.01), which is significantly overused in Chinese learners' writings. All the other words do not have significant difference. This result suggests that the Chinese learners' writings in applying the words *increase*, *rise* and *decrease* are much similar to those of English native speakers in the frequency perspective; but they overused the word *reduce* in writings.

Table 2. Frequencies of target words

	Freq in CHN	Freq in ENS	Chi-Square	log-likelihood	
increase	84	28	1.7457	1.81	+
rise	17	6	0.2359	0.24	+
reduce	88	18	9.4531 **	10.55**	+
decrease	36	12	0.2513	0.78	-

Note: "+" means "overuse"; "-" means "underuse"; The asterisks \*\* shows a significance level of 0.01.

#### 3.2 Collocation and Semantic Prosody

Table 3 shows the distribution of *reduce* and *decrease*, *increase* and *rise* in positive, negative and neutral semantic environment in Chinese learners' and English native speakers' writings.

Table 3. Semantic prosody of two pairs of words

	positive		neutral		negative	
	CHN	ENS	CHN	ENS	CHN	ENS
reduce	2 (.02)	0	21 (.24)	13 (.72)	64 (.73)	5 (.27)
decrease	2 (.06)	2 (.17)	18 (.50)	4 (.33)	16 (.44)	6 (.50)
increase	7 (.08)	6 (.21)	37 (.44)	15 (.54)	40 (.48)	7 (.25)
rise	1 (.06)	1 (.17)	9 (.53)	2 (.33)	6 (.35)	3 (.5)

Note: the bold number refers to a significant different in frequency. More information will be provided in the description part.

As the number of the concordance of each word is not very large, a chi-square is also calculated. It is found that *reduce* used in neutral semantic context shows a significant difference between CHN and ENS ( $p < 0.001, -$ ), which means that the use of *reduce* in neutral context in Chinese learners' writing is significantly less than that in English native speakers' writing. However, in negative semantic context, the use frequency of Chinese learners' writings is significantly more than that of English native speakers ( $p < 0.001, +$ ). This result indicates that Chinese learners tend to use *reduce* in a negative context. Besides, there is a significant difference in the use of *increase* in a negative semantic context ( $p = 0.036, +$ ). The data indicates that Chinese learners tend to use *increase* more in a negative semantic context than native speakers. The other words do not show difference in the distribution in semantic distribution, which means that the Chinese speakers could apply *decrease* and *rise* in much the same way as the native speakers do.

It is easy to found that two pairs of words in both ENS and CHN writings tend to have a neutral or negative semantic prosody. This may result in the topics of the data collected in the corpus. However, since this study only investigates the different use in words between Chinese learners and native speakers rather than the semantic prosody of words, the actual semantic prosody will not be discussed here. A description of the actual use of words will be provided so as to find out more about the difference between the two.

### 3.3 Reduce v.s. Decrease

From table 3, the word *reduce* in this ENS writings tends to occur more in neutral context. The typical collocations of *reduce* are as follows:

Neutral:

- V+NP (*reduce* as a transitive verb): loans, cost, income, time, number, risk, issue, reliance, cost;
- NP+be+V-ed (*reduce* in passive voice): tax revenue, place;

Negative:

- V+NP (*reduce* as a transitive verb): pressure, smoking, harmful pollutants;
- NP+be+V-ed (*reduce* in passive voice): debt;

Among all these words with which it co-occurs, the words *debt*, *cost* and *loads* show a significant relationship with *reduce* (MI3 = 4.00, 3.42, 3.42). The collocates show that *reduce* in native speakers' writings are usually used to express "the amount of something lessens".

The word *reduce* in CHN writings usually co-occurs with the following words or phrases:

Positive:

- V+NP (*reduce* as a transitive verb): learning proficiency, reputation;

Neutral:

- V+NP: chance, time, population, income, distance, use, appetite, cost, expanse, life, amount;
- NP+be+V-ed: time, number
- NP+VP (*reduce* used as an intransitive word): production, customers,

Negative:

- V+NP: burden, smoke, pressure, stress, disaster, harm, conflicts, danger;
- NP+be+V-ed: burden, smoking;

Those words with a MI3 of 3 or higher are *burden*, *time*, *money*, *pressure*, *smoking*, *stress*, etc. However, those words which occur more frequently in ENS are not included (*cost* only shows a MI3 of 2.27, *debt* and *loans* are not used). This indicates that Chinese learners use *reduce* in a very different way from native speakers. Most of the words co-occurring with *reduce* in CHN writings expressing a negative meaning. Also it can be seen that there are some unusual collocations like *reduce danger*, *reduce smoking*, etc.

As for *decrease*, the frequent collocates in CHN writing are as follows:

Positive:

- VP+NP (*decrease* used as a transitive verb): opportunity;
- NP + VP (*decrease* used as an intransitive verb): desire (for entertainment)

Neutral:

- NP + VP (*decrease* used as an intransitive verb): production;
- VP+NP (*decrease* used as a transitive verb): payment, time, selling, tax, action, study quality, chance, ability, time, immune system, life span/quality, population, customers;
- NP+Prep (a decrease in/of): sales

Negative:

- NP + VP (*decrease* used as an intransitive verb): cigarette;
- VP+NP (*decrease* used as a transitive verb): burden, pressure, risk, cigarette, bad effects, smokers, unemployment;
- NP+Prep (a *decrease* in/of): heart attack, cancer;

The MI3 value shows that *pressure*, *cost*, *smoker* have a strong occurrence with *decrease*. The semantic prosody is nearly half neutral and half negative, which shows a minor difference from the ENS writings, whose semantic prosody is more negative (.50). As the *decrease* collocations are similar in CHN and ENS, the detailed collocates will not be presented here. The frequent collocation with a MI3 of 3 or higher in ENS are *costs*, *diseases*, *lifespan*, *loans*, etc.

### 3.4 Increase vs Rise

The word *increase* in ENS occurs more in neutral meaning environment and equally occurs in positive and negative environment. The patterns are as follows:

Positive:

- VP+NP (*increase* used as a transitive verb): profit, health level, good performance, advantage;
- NP +VP (*increase* used as an intransitive verb): profit

Neutral:

- VP+NP (*increase* used as a transitive verb): litter and trash;
- V-ing +NP (*increasing* as an adj): awareness, cost, demands, packages, income, business, tax, expenditure, skills, number;
- NP + Prep (an *increase* of): customers;

Negative:

- NP +VP (*increase* used as an intransitive verb): sickness/ death, killing, workload;
- VP+NP (*increase* used as a transitive verb): workload, risk (of cancer);

The MI3 shows that the word *increase* has strong collocation with *workload*, *customers*, *business*, all of which express a neutral meaning. However, it is used more frequently in a negative context in CHN writings. The overall collocates are presented below:

Positive:

- VP+NP (*increase* used as a transitive verb): useful skills, efficiency, knowledge;

Neutral:

- NP +VP (*increase* used as an intransitive verb): quality of life, graduates, income, tuition fee/ price, cost;
- VP+NP (*increase* used as a transitive verb): students, tendency, income, chance, social experience;
- V-ing +NP (*increasing* as an adj): company, job, students, population, role, ability, number;
- NP + Prep (an *increase* of): class;

Negative:

- NP +VP (*increase* used as an intransitive verb): smokers, juvenile crime, unemployment, smoking, heart attack, depression/anxiety, people (killed);
- V-ing +NP (*increasing* as an adj): concern, smokers, heart disease, cancer rate, risk, possibility (of cancer);

There are more collocates in CHN probably because there are more writings than ENS. The strong collocation shown by MI3 are *risk*, *chances*, *experience*, *casualties*, etc. The semantic prosody in CHN writings is more negative, which is also revealed by frequency.

As for *rise*, the number of tokens of *rise* is too small in ENS, therefore, it is hard to decide their preference for

using *rise* in writing (also meaningless to calculate MI3). However, the semantic prosody of CHN writing of *rise* is still clear. The Chinese learners tend to use *rise* in a neutral meaning context (.53). The frequent collocates in CHN are *cost*, *death rate* and *disease*, which correspond to those in ENS. However, there is certain unusual use of collocation found in CHN, such as *a rise on life richness*.

#### 4. Conclusion and Implications

The study explores the use of frequency, collocation and semantic prosody of two pairs of synonyms by comparing the Chinese learners' writing with English speakers' writing. The results show that although there are certain minor differences in the collocational pattern between the two, the semantic prosody is similar. The major difference lies in the word *reduce*, which is more frequently used in a negative semantic meaning in Chinese learners' writing. This result indicates that, given the same topic, there might not be much difference in the users' perception of semantic prosody. Instead, the major difference lies in collocation of words. Therefore, it is suggested that semantic prosody may not necessarily lead to difficulty in learning, but for learners, collocation may still be a hard nut to crack.

Given the fact that the main difficulty in learning might lie in collocation of words rather than perception of semantic prosody, language teachers are suggested to pay more attention on helping learners with collocation of words. In this regard, teachers might enlighten students to compare the differences of collocations in synonymous word pairs to help understand the differences uses of words.

As the scope of this article only is limited to two pairs of words, it could only present a tip of the iceberg. Therefore, further research could try to investigate more synonymous pairs to see if learners behave the same with or differently from native speakers in perception of semantic prosody and the use of collocation.

#### References:

- Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics & Linguistic Theory*, 4(2), 119-139. <https://doi.org/10.1515/CLLT.2008.006>
- Hu, H. C. M. (2015). A semantic prosody analysis of three adjective synonymous pairs in COCA. *Journal of Language & Linguistics Studies*, 11.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world* (Volume 1, pp. 91-118). Kobe, Japan: Kobe University.
- Oster, U. (2010). Using corpus methodology for semantic and pragmatic analyses: What can corpora tell us about the linguistic expression of emotions? *Cognitive Linguistics*, 21(4), 727-763. <https://doi.org/10.1515/cogl.2010.023>
- Sinclair, J. (1999). *Corpus, concordance, collocation*. Shanghai Foreign Language.
- Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes*, 30(2), 101-112. <https://doi.org/10.1016/j.esp.2010.12.003>
- Wei, N., & Li, X. (2014). Exploring semantic preference and semantic prosody across English and Chinese: their roles for cross-linguistic equivalence. *Corpus Linguistics & Linguistic Theory*, 10(1), 103-138. <https://doi.org/10.1515/cllt-2013-0018>
- Xiao, R., & Mcenery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129. <https://doi.org/10.1093/applin/ami045>
- Zhang, C. (2010). An overview of corpus-based studies of semantic prosody. *Asian Social Science*, 6(6). <https://doi.org/10.5539/ass.v6n6p190>

#### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).