



A Semi-automatic Error Retrieval Method for Uncovering Collocation Errors from a Large Learner Corpus

以半自動化擷取方法探究大型學習者語料庫之搭配詞錯誤
Christine Ting-Yu Yang, et al. [full author details at the end of the article]

Received: 30 April 2019 / Revised: 15 August 2019 / Accepted: 23 September 2019 /
Published online: 11 November 2019
© National Taiwan Normal University 2019

Abstract

Previous studies on ESL/EFL learners' verb-noun (V-N) miscollocations have shed some light on common miscollocation types and possible causes. However, barriers to further understanding of learners' difficulties still exist, such as the limited amount of learner data generated from small corpora and the labor-intensive process of manually retrieving collocational errors. To provide researchers with a more efficient retrieval method, this study proposed the use of the *Sketch-Diff* function in the *Sketch Engine* (SKE) platform to semi-automatically retrieve collocation errors in large learner corpora. To test the feasibility of this semi-automatic retrieval method, a 7.4-million-word EFL learner corpus was investigated with *Sketch-Diff*, and 4541 tokens of common miscollocations were identified. Analysis of these miscollocations revealed that most errors were verb-based and often caused by negative transfer from the learners' L1, undergeneralization (e.g., ignorance of L2 syntactic rules), and approximation (e.g., the misuse of near-synonyms, hyper-/hyponyms, antonyms, and lexemes with similar sound/form). This study demonstrates that using *Sketch-Diff* to retrieve V-N miscollocations from a large learner corpus is both feasible and efficient. This method can be applied to other languages to further deepen our understanding of L2 learners' difficulties in collocation acquisition.

中文摘要

過去針對以英語為二語/外語的學習者的動詞—名詞(動-)錯誤搭配研究, 已揭示常見之錯誤類型與可能成因。然而, 此類研究多採用較為耗費人力之搭配詞檢索方式, 來探究小型語料庫之錯誤情形, 造成學者難以將此類擷取方法應用至大型語料庫。為提供研究者更有效率之搭配詞錯誤擷取方法, 本文提倡使用 *Sketch Engine* 語料處理平台的 *Sketch-Diff* 功能, 以半自動化方式自大型學習者語料庫中擷取搭配詞錯誤。為測試此半自動化擷取方式之可行性, 本文以 *Sketch-Diff* 檢驗一座七百四十萬字的學習者語料庫中的錯誤搭配錯誤情形, 並判別出 4,541 筆常見之動- 錯誤搭配詞。分析結果指出多數錯誤為動詞之誤用, 而歸咎其成因則多為學習者之母語負遷移、生成不足(如: 忽略二/外語句法規則)以及相近表達誤用(如: 近義詞誤用、上/下位詞誤用、反義詞誤用以及形/音近詞誤用)。本文研究結果表明 *Sketch-Diff* 功能可有效以半自動化方式自大型學習者語料庫擷取動—名搭配詞錯誤, 並建議應用此研究方法至其他語言, 以進一步加深對二/外語學習者搭配詞習得困難之理解。

Keywords Collocation · Error extraction method · Learner corpus · Error analysis

關鍵詞 搭配詞 · 錯誤擷取方法 · 學習者語料庫 · 錯誤分析

Past Corpus-Based Research on Learners' Collocational Errors

In the field of second language (L2) learning, collocations have been widely recognized as a key aspect of vocabulary competence ([5; Nattinger & [11, 36, 37, 44]), with many studies highlighting the significant role collocations play in developing a learner's mental lexicon [12, 20, 22]. Mastery of collocations not only facilitates a learner's linguistic production and overall comprehension [6, 15, 20, 27], but also enables the learner to achieve fluency in the target language [1, 34, 46] so as to satisfy "the [learner's] desire to sound [and write] like others" in certain registers ([43], p.75).

However, it is also widely acknowledged that collocations, especially lexical collocations (i.e., collocations constructed of two open-class components such as *adjective-noun* and *verb-noun*), are difficult for L2 learners to master. Studies have consistently revealed ESL/EFL learners' insufficient knowledge of English lexical collocations [2, 4, 7–10, 13, 14, 19, 23, 32]. Of these, verb-noun (V-N) collocations are widely recognized as the most significant structure because they "form the communicative core of utterances where the most important information is placed" ([3], p.227). Nevertheless, this structure is also reported to be particularly difficult for L2 learners to acquire [19, 21, 24, 32, 40, 45].

Given the importance of V-N collocations and L2 learners' difficulty in mastering them, researchers have been investigating how learners use V-N collocations. The goal of some of these studies has been to enhance learners' awareness of acceptable collocations by having the learners notice their own errors [20, 42]. To achieve this aim, Paquot & Granger [33] have proposed the use of learner corpora to investigate learners' miscollocations. They have argued that the languages examined in learner corpora are made of "continuous stretches of oral or written discourse (p.131)," and the wording is more naturally selected by learners in the form of pedagogically designed tasks. Furthermore, the electronic format of corpus data allows researchers to automatically extract collocations for further analysis with the help of a wide range of corpus tools. Because of these features, researchers have commented that learner corpora are ideal data sources to study learners' collocation use.

In fact, there has been increasing research into both European learners' (e.g., [18, 26, 30, 31, 41]) and Asian learners' V-N miscollocations (e.g., [16, 19, 29, 41]), and these studies have yielded quite consistent findings regarding the common types and causes of learners' collocational errors. Regarding categories of miscollocation, learners were found to frequently misuse the verb components of V-N collocations in their writing. As for the causes, negative L1 influence is reported to be the most influential factor in error production. For instance, in one of the most extensive studies on collocations, Nesselhauf [31] investigated V-N collocations produced by advanced German learners of English in the German subcorpus of the International Corpus of Learner English (GeCLEE). She manually extracted V-N combinations in 318 essays and identified 836 deviations. Her analysis of these errors suggested that most of the errors were verb-based, with congruent word-for-word translations from L1 to L2 as the major causal factor.

Even though previous studies have revealed some informative and useful findings, some limitations still remain. The first limitation concerns the limited size of the

corpora. The GeCLEE in Nesselhauf's [31] study, for instance, comprises only 154,191 words. The Israeli Learner Corpus of Written English in Laufer and Waldman's [19] study, another comprehensive study on V-N collocations, comprises 291,049 words. Another corpus containing more than 100,000 words is the 160,000-word corpus in Marco [26]. Other previously investigated corpora are all smaller than 100,000 words. The concern regarding the use of small corpora is whether the findings could comprehensively represent learners' V-N miscollocations. As argued in Paquot and Granger [33], the larger a corpus is, the higher the degree of the representativeness of the data and the generalizability of the results. Based on their argument, it is reasonable to believe that undertaking studies on larger corpora could reveal more generalizable results of learners' V-N miscollocations. Existing studies, however, have been mostly conducted based on relatively small corpora, and there is thus room for the undertaking of larger corpus studies to obtain more comprehensive as well as generative results.

The reason for using small corpora in previous studies might be attributed to the intensive manual work related to the process of retrieving and identifying V-N miscollocations in L2 learners' production. Most of the existing corpus-based studies on learners' V-N miscollocations were conducted by manually retrieving potential errors from raw data. The researchers would often firstly create a list of key verbs (e.g., [16, 18, 26, 29, 41]) or nouns (e.g., [19]) as the nodes to generate their concordance lines with corpus tools. The researchers would then manually scrutinize the generated concordance lines one by one to identify both V-N collocations and potential miscollocations. The acceptability of manually identified potential errors would be determined with the consultation of collocation dictionaries and/or reference native corpora to decide whether they were indeed V-N miscollocations or not. Substantiated V-N miscollocations would then be analyzed for their types and causes. This manual retrieval process often imposed a considerable burden on the researchers. For instance, Nesselhauf [31] manually identified 836 V-N miscollocations out of a base of 2082 V-N collocations. In the study by Laufer and Waldman [19], the researchers identified 561 errors out of a base of 18,415 generated V-N combinations. In these studies, the researchers invested a significant amount of time to pick out V-N miscollocations from a much greater number of acceptable combinations. The same manual procedure would be somewhat impractical for researchers aiming to examine miscollocations from larger learner corpora, for the number of to-be-examined V-N combinations may multiply by the increase of the corpus size, with the majority of these combinations actually well-formed items.

In addition to investigating corpora with raw data, several studies have retrieved data from error-tagged corpora (e.g., [24, 47, 48]). In Zhang and Yang's [48] study, for instance, the researchers retrieved combinations with the tag "CC3," which stands for V-N combination errors in the one-million-word Chinese Learner English Corpus (CLEC), for analysis. They discovered that the majority of the retrieved 1481 miscollocations were caused by a negative transfer from the learners' L1 and that the three most common types of miscollocation were inappropriate choice of verb collocates, the misuse of delexical verbs, and erroneous noun choice. For researchers, retrieving V-N miscollocations from error-tagged learner corpora seems convenient, yet the construction of an error-tagged corpus is itself time-consuming and labor-intensive in nature. The construction of the CLEC, for instance, took seven

years before being released to the public. In addition to the lengthy time required for construction, errors in the CLEC were mostly tagged by non-native speakers in China. The potential inconsistency and/or mistagging in error-tagging by non-native speakers' judgment may decrease the representativeness of the identified miscollocations.

With increases in the number of large learner corpora, such as the two-million-word-plus ICLE, the 1.3-million-word International Corpus Network of Asian Learners of English (ICNALE), and the 83-million-word-plus EF-Cambridge Open Language Database (EFCAMDAT), discovering a less labor-intensive retrieval method to generate more representative results in these large corpora is thus essential to ease researchers' burden when investigating V-N miscollocations.

Semi-automatic Error Retrieval with Sketch Engine

To overcome the above limitations, the current study proposes an innovative method that integrates computer-aided semi-automatic error retrieval with human inspection. The method is based on using the *Sketch-Diff* function of the commercial online platform *Sketch Engine* (<http://www.sketchengine.co.uk>), the user interface of which is illustrated in Fig. 1.

The *Sketch-Diff* function can systematically display three different types of collocational comparison. One is *Sketch-diff by lemma*, which presents the collocational similarities and discrepancies between two lemmas in the same corpus. Another is *Sketch-diff by word form*, which compares the collocational behaviors between two-word forms of the same lemma. The other is *Sketch-diff by subcorpus*, which compares and contrasts the collocates of the same headword in two different corpora, and is also the option that the researchers propose to retrieve potential V-N miscollocations. The mechanism of *Sketch-diff by subcorpus* is to compare the association strength (i.e., logDice score¹) of a headword's collocates in one subcorpus with that in another subcorpus. Collocates with statistically higher logDice scores in the former subcorpus will be presented in a green area, indicating that these items co-occur with the headword significantly more often. Similarly, collocates with significantly higher logDice scores in the latter corpus will be shown in a red area. For items with equal/similar logDice scores in both corpora, they will be presented in a white area, suggesting that these collocates co-occur with the headword equally often in the two sub-corpora. Take the verb *tell* for example. A user can use *Sketch-diff by subcorpus* to examine whether the noun collocates of *tell* in writing are different from those in speaking by assigning the written texts in BNC as one subcorpus and the spoken transcripts in BNC as the other subcorpus. As illustrated in Fig. 2, items in the green area are more likely to co-occur with *tell* in written discourse, whereas those in the red area co-occur with *tell* more frequently in spoken discourse. Collocates in the white area are then those that co-occur equally often in both registers.

With this unique feature, we propose that *Sketch-Diff* can be employed to retrieve potential V-N miscollocations by comparing the verb/noun collocates of

¹ The formula of logDice score and its features are presented in Rychlý [35].

Fig. 1 User interface of *Sketch-Diff*

targeted keywords in a designated learner corpus with those in a large native corpus. This newly proposed method might be more preferable than previously employed methods due to the fact that it is more automatic and more time-efficient. While methods in previous studies often required researchers to manually scrutinize the concordance lines of every V-N combination to retrieve potential miscollocations, the *Sketch-Diff* function can explicitly present a searched word's potentially erroneous collocates in a summary chart. This feature of *Sketch-Diff* will be illustrated in detail in a later section of this paper.

To examine the feasibility of adopting the *Sketch-Diff* to retrieve potential miscollocations in large learner corpora, a study was conducted using *Sketch-Diff* to investigate the V-N miscollocations from texts produced by Chinese-speaking EFL learners. Two research questions were raised:

1. Can the *Sketch-Diff* tool uncover potential V-N miscollocations by semi-automatically comparing a large learner corpus with a large native corpus? If so, to what extent is this method more efficient than traditional error retrieval methods?
2. Based on the findings retrieved by *Sketch-Diff* in the large learner corpus, what are the common error categories and the possible causes of V-N miscollocations by Chinese-speaking EFL learners?

tell (verb) Sketch diff for subcorpora Written/Spoken

Written	6.0	4.0	2.0	0	-2.0	-4.0	-6.0	Spoken
object	17,951	1,523	0.29	0.13				
reporter	<u>171</u>	0	8.1	--				
conference	<u>285</u>	0	7.9	--				
magistrate	<u>97</u>	0	7.1	--				
inquest	<u>82</u>	0	7.1	--				
court	<u>639</u>	<u>22</u>	8.1	3.6				
friend	<u>250</u>	<u>11</u>	6.8	2.6				
police	<u>292</u>	<u>14</u>	7.4	3.6				
story	<u>1,247</u>	<u>62</u>	9.8	6.1				
truth	<u>577</u>	<u>31</u>	9.4	6.5				
tale	<u>260</u>	<u>18</u>	8.5	6.3				
jury	<u>216</u>	<u>15</u>	8.4	6.6				
difference	<u>88</u>	<u>20</u>	5.9	4.4				
joke	<u>83</u>	<u>11</u>	7.0	6.0				
lie	<u>218</u>	<u>37</u>	8.4	8.1				
mum	<u>87</u>	<u>61</u>	6.7	7.4				
dad	<u>37</u>	<u>39</u>	5.5	7.0				
mummy	<u>9</u>	<u>14</u>	3.9	7.1				
Brian	0	<u>7</u>	--	5.0				
Wesley	0	<u>12</u>	--	7.3				
daddy	0	<u>23</u>	--	7.8				

Fig. 2 Sketch-Diff of tell in the written subcorpus and in the spoken subcorpus of BNC

The Corpora

For this study, a Chinese-speaking EFL learner corpus was uploaded to *Sketch Engine* (SKE) to conduct the proposed semi-automatic error retrieval. The learner corpus was composed of four sub-corpora—CLEC 1.0, the Written English Corpus of Chinese Learners (WECCL) 1.0 and 2.0, the Joint College Entrance Examinations Testees Corpus (JCEETC), and the Taiwanese College Learner Corpus (TCLC).

The first two sub-corpora contain data produced by EFL learners in Mainland China. CLEC, used in Zhang and Gao [47] and Zhang and Yang [48], is a one-million-word error-tagged corpus consisting of written texts produced by high school and college students. In this study, we employed the un-error-tagged version. The WECCL 1.0 and 2.0 are sub-components of the Spoken and Written English Corpus of Chinese Learners 1.0 and 2.0, currently the largest learner corpus in Mainland China. Since both the learner data and the reference corpus in this study were mainly composed of written texts, only the written components were uploaded to *SKE*. The other two learner corpora include articles written by EFL learners in Taiwan. The approximately two-million-word JCEETC consists of texts written by Taiwanese high school graduates during their college entrance exams. As for TCLC, it contains 1.8 million words written

by college students from six universities in Taiwan. The four sub-corpora were combined into one 7.4-million-word corpus, the size of which is considerably larger than previous studies', to examine whether the proposed method could smoothly process such a large amount of data and yield informative results regarding L2 learners' V-N miscollocations.

To retrieve potential miscollocations from the learner corpus, a native reference corpus was required to execute the semi-automatic retrieval of potential errors. In this study, two existing native speaker corpora, the BNC corpus and the Corpus of Contemporary American English (COCA), were utilized as references. The reason for using both BNC and COCA, rather than employing one of them only, was to prevent the possible usage/spelling differences between British and American English. Since we aimed to draw on more formal written language to compare with the learners' writing, only the BNC written corpus and news texts of COCA were selected and combined into one large reference corpus for comparison. This corpus, labeled as BNCCOCA, contains approximately 222 million words.

Retrieval of Potential Miscollocations Through Sketch-Diff

To retrieve potential V-N miscollocations through *Sketch-Diff*, a list of the most frequent nouns in the learner corpus was generated. The reason for generating a list of frequent nouns instead of verbs is that nouns tend to be the main indicators of learners' English V-N miscollocations [24]. A similar point was also made by Manning and Schütze [25], who used the term "focal word" to indicate the crucial role of nouns in V-N collocations. Hence, inspecting the verb collocates of a noun is a more efficient way to identify V-N misuse than looking into the noun collocates of a verb. To investigate the most frequent nouns within a manageable number of words for analysis, this study set the frequency threshold at 300 times.

Based on the frequency threshold, a list of 690 key nouns was generated. Verb collocates of these target nouns in both BNCCOCA and the learner corpus were then retrieved via the use of *Sketch-Diff*. Take the noun *knowledge* (appearing 6789 times in the learner corpus) for example. Through the use of *Sketch-Diff*, a summary chart illustrating the verb collocates with *knowledge* as the object in BNCCOCA and in the learner corpus was obtained. As illustrated in Fig. 3, the blue row presents the occurrences of *knowledge* as an object in BNCCOCA and in the learner corpus, which are 5135 and 2841 respectively. The red column shows verbs frequently used by Chinese-speaking EFL learners to collocate with *knowledge* and which rarely appear in the native corpus. For example, the collocation *master knowledge* appeared 58 times in the learner corpus, whereas it did not appear at all in BNCCOCA. Similarly, there were zero occurrences of the collocation *enrich knowledge* in BNCCOCA. In contrast, this collocation appeared eight times in the learner corpus. Combinations appearing quite frequently in the learner corpus yet never occurring in BNCCOCA were thus potential V-N miscollocations that were deemed worthy of further analysis. However, as the main goal of this study was to identify common errors made by learners, the V-N miscollocations to be investigated had to meet two criteria: the miscollocation appeared more than three times in the learner corpus, and the miscollocation was not found in BNCCOCA.

knowledge⁰ Sketch diff for subcorpora BNCCOCA/ChineseEFLLearners

	BNCCOCA	6.0	4.0	2.0	0	-2.0	-4.0	-6.0	ChineseEFLLearners
verbs with "knowledge" as object	5,126	2,673	0.29	0.43					
deny	103	0	8.0	--					
specialise	30	0	7.4	--					
test	36	0	6.6	--					
glean	16	0	6.6	--					
extend	37	0	6.6	--					
assume	26	0	6.3	--					
demonstrate	24	0	6.3	--					
advance	17	0	6.3	--					
further	14	0	6.3	--					
display	24	0	6.2	--					
pool	12	0	6.1	--					
derive	17	0	6.1	--					
disseminate	11	0	6.0	--					
broaden	12	0	5.9	--					
share	90	4	7.2	2.9					
work	57	3	5.9	1.8					
require	171	10	7.1	3.1					
possess	52	4	7.6	4.4					
lack	45	4	7.0	3.9					
increase	81	13	6.0	3.5					
apply	58	10	7.2	5.0					
have	1,184	275	6.0	3.9					
teach	5	30	3.3	6.1					
get	33	260	2.8	5.8					
grasp	3	23	3.9	7.5					
absorb	4	61	4.0	8.5					
know	3	76	0.5	5.2					
study	3	90	2.4	7.5					
learn	8	668	3.0	9.5					
recite	0	3	--	4.8					
enlarge	0	3	--	4.9					
comprehend	0	3	--	5.0					
acquire	0	3	--	5.2					
cram	0	4	--	5.3					
specialize	0	5	--	5.7					
enrich	0	7	--	5.9					
lean	0	8	--	6.3					
master	0	58	--	8.8					

Fig. 3 Sketch-Diff of knowledge between BNCCOCA and the learner corpus

Some people might question the appropriateness of setting a native speaker corpus as the norm to decide the acceptability of the retrieved collocations. It should be noted that, however, the size of the BNCCOCA is 30 times larger than the Chinese-speaking EFL learner corpus, and this vast amount of data should be adequate to generate possible V-N combinations used by native speakers. If a V-N combination never appeared in BNCCOCA, it is very likely that this combination is scarcely, if ever, used by native speakers and could cause a great difficulty for readers trying to comprehend its meaning. In addition, the current research set out to investigate V-N miscollocations in learners' writing so as to prevent the learners from continually making the same mistakes. Since languages in the written form are often considered more formal, it is preferable for EFL learners to produce V-N collocations that are also used by native speakers in their writing so that the writing is comprehensible to readers around the world².

Here we would like to elaborate on how this proposed method is more efficient than the manual error retrieval methods applied in previous research, and hence answer the first research question. For example, let's again consider the noun *knowledge*. In the learner corpus, there are 2841 instances of *knowledge* as an object. If we adopted the traditional error retrieval method, we would have to inspect these 2841 concordance

² In the context of World Englishes (WE), some people might question the suitability of using native speakers as the norm to decide the acceptability of V-N collocations produced by language learners. Some might argue that learners' atypical collocations (i.e., collocations that are rarely/never used by native speakers) could still be considered acceptable as long as their meanings are comprehensible. In the context of English writing, however, this might not be the case. As pointed out by Matsuda and Matsuda [28], many teachers have a stricter standard in students' written production and tend to assign lower scores on writing assignments with language features that are deviant from the native norm. Even students expect more corrective markings from their teachers on their written texts. Since the current study aims to uncover English learners' collocation uses in writing, we argue that deviant collocations generated from the comparison with native speakers' written texts should be considered problematic expressions that need to be dealt with for the better teaching/learning of English writing.

lines to firstly differentiate potential erroneous *V-knowledge* collocations from the well-formed ones. The acceptability of the potential *V-knowledge* miscollocations would then be judged by the researchers with the use of collocation dictionaries and native speaker corpora. By employing the *Sketch-Diff* function, however, the error retrieval process is more time-efficient in two ways. First, we only needed to examine the concordance lines of the nine verb collocates that occurred more than three times in the learner corpus but which did not appear in BNCCOCA (see Fig. 4 for the nine verb collocates), and the number of to-be-examined concordance lines plummeted to 96 occurrences. In addition to the reduced number of concordance lines, consultation with native speaker corpora for identifying miscollocations that are not used by native speakers is also excluded from the error identification process because the nine collocates generated by *Sketch-Diff* had already been found missing in BNCCOCA due to the corpora comparison.

To further prove the efficiency of the proposed semi-automatic error retrieval method, the researchers calculated the total occurrences of the 690 key nouns as objects in the learner corpus and obtained the result of 311,915 occurrences. Again, if we adopted the traditional error retrieval method, we would have to spend a great amount of time inspecting all the 311,915 instances of V-N combinations one by one to identify erroneous collocations. On the contrary, by examining items in the red area that never occur in the native reference corpus, the number of concordance lines for inspection was reduced to 12,431 instances, which is only one twenty-fifth of the 311,915 instances. Based on the two figures, it is obvious that the proposed error retrieval method is more efficient than those in previous studies.

Even though this method helps to retrieve collocation errors more easily, false alarms (i.e., correct usages and/or non-V-N combinations wrongly marked as incorrect V-N collocations) might still be generated due to learners' misspelling or false POS tagging. Thus, to ensure that the 12,431 instances in the red column were indeed genuine miscollocations made by the learners, potential collocation errors in this area were further checked by one coder. However, to ensure that the designated coder could accurately identify which instances were miscollocations and which were false alarms, around 10% ($n = 1239$) out of the 12,431 instances were randomly selected and examined by four other coders for a reliability test. If an instance labeled as a false alarm by the designated coder was also identified as a false alarm by any two of the other four coders, the designated coder's identification of that instance would be recognized as correct. Cross-examination on the 1239 instances showed that 1170 out of the coded 1239 instances were judged the same by any of the other two coders, suggesting an agreement rate of 94.4%. After ensuring the reliability of the designated coder's judgment, the rest of the instances were then all examined by the designated coder.

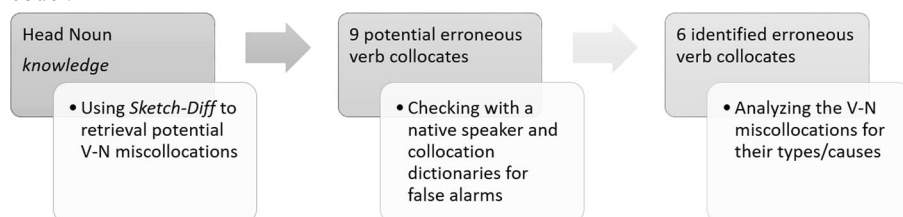


Fig. 4 Process of retrieving miscollocates of *knowledge* and eliminating false alarms

To more clearly illustrate the procedure of how possible errors were identified and how false alarms were managed, Figure 4 presents an example of how potential miscollimates with the head noun *knowledge* were identified and dealt with. Nine potential miscollimates of *knowledge* were retrieved by *Sketch-Diff*. The coder then consulted with a native English teacher and collocation dictionaries for confirmation. Three of the collocates were treated as false alarms, leaving only six verb collocates to be analyzed.

Examination of the 12,431 instances yielded 7890 instances of false alarms, which mostly resulted from misspelling (e.g., “*surfe the Internet”), false POS tagging (e.g., *sing + club* in “the singing club”), and others (e.g., *take + the Internet* in “take the Internet as an example”). While *Sketch-Diff* seems to generate many false alarms in the error retrieval process, these false alarms are mostly caused by the nature of the learner corpus itself and would also occur in the traditional error retrieval method. For instance, there are often lexical and grammatical errors in learner corpora, which can cause some difficulties for an English POS (part of speech) tagger. Because most English POS-taggers are designed to process data produced by native speakers, their accuracy rates are lower when used to process L2 learner languages and produce many false POS taggings (cf. [39]). These false POS taggings can lead to the occurrence of many false alarms when employing *Sketch-Diff*. If a non-verb word in a learner corpus is wrongly tagged as a verb by a POS-tagger, *Sketch-Diff* will mistakenly treat this word as a potential verb collocate of a head noun and compare the combination’s logDice score in the learner corpus with that in the native speaker corpus. Once the non-verb word in the native corpus is not mistagged as a verb, the logDice score of the combination in the learner corpus will be much higher than that in the native speaker corpus. This thus triggers *Sketch-Diff* to categorize this combination into the red area and cause the occurrence of false alarms. However, these false alarms would be retrieved by the traditional error retrieval method, because the traditional method also requires the learner data to be POS-tagged first before retrieving all the V-N constructions for error identification. Thus, filtering out these false alarms is an inevitable process in both the traditional and the proposed semi-automatic error retrieval methods.

Classification and Analysis of V-N Miscollimations

Examination of the semi-automatically retrieved potential V-N miscollimations ultimately identified 4541 instances of collocation errors, and the concordance lines of these errors were manually examined for the classification of error category and possible causes. Reviewing the classification systems of previous studies (e.g., [30, 31, 48]), learners’ V-N miscollimation errors could be generally divided into (1) verb-based deviations, (2) noun-based deviations, and (3) other deviations. However, the subcategories under these three types of deviation vary in different studies, causing difficulties in determining which subcategories to include. Instead of adopting/modifying an existing framework, the researchers classified the errors into verb-based, noun-based, and others at the preliminary stage of classification, and further categorized these errors into subgroups after a thorough analysis of the error.

As for the possible causes of each V-N miscollimation, the researchers referred to James’ taxonomy [17] of error diagnosis, which includes interlingual, intralingual, communication-strategy, and induced errors. Some modifications were made to the

framework before applying it to the errors. For instance, induced errors, referring to errors “that result more from the classroom situation than from either the students’ incomplete competence in English grammar or first language interference” ([38], p.256), were less likely to be noted since the researchers could not know how much the learners were negatively influenced by their classroom situation. This cause of error was hence excluded from the framework employed in this study. In addition, both intralingual and communication-strategy errors are caused by incomplete knowledge of the L2. The researchers therefore combined these two causes into intralingual errors; subcategories of intralingual errors, however, were described as follows: approximation (i.e., errors resulting from the misuse of another near-equivalent L2 item), overgeneralization (i.e., errors resulting from the overuse of one member of a set of forms and the underuse of others in the set), and undergeneralization (i.e., errors resulting from the incomplete rule application of an L2 item). The causes of miscollocations in the present study thus include (1) negative L1 transfer, (2) approximation, (3) overgeneralization, and (4) undergeneralization.

Results and Discussion

The first part of this section presents the common error categories of the V-N miscollocations retrieved by *Sketch-Diff*. Descriptive statistics of the results are given and compared with those of previous studies. The second part then presents the possible causes of these miscollocations and discusses each with examples extracted from the learner corpus.

Common Categories of Chinese-Speaking EFL Learners’ V-N Miscollocations

In this study, 4541 tokens (570 types) of V-N miscollocation were retrieved and identified with *Sketch-Diff* and human inspection. In addition to assigning these miscollocations into verb-based, noun-based, and other deviations, two other main categories were also defined, namely mixed deviation (i.e., errors contain two different deviation categories) and unclear meanings (i.e., errors where it is difficult to understand the intended meaning). Under these five main categories, 10 subcategories were then identified. The categorization of these errors is presented in Table 1.

Based on the findings, it was found that 83.7% of the 4541 V-N miscollocations were verb-based deviations, with only 12.2% and 1.7% of the errors being noun-based and other deviations, respectively. It should be noted that among the 101 occurrences of Mixed Deviations, 91 of them were verb-based errors with the other two main deviation categories showing a high extent of inappropriate verb uses. These findings clearly demonstrate that the Chinese-speaking EFL learners misused verb collocates more often than other components in their production of V-N collocations.

To examine whether the results of this study differ from those of previous studies, we compared the three most common categories of misuse identified in this study with those of Nesselhauf [31] and Zhang and Yang [48]. Table 2 presents the three most common categories of misuse identified in the three studies.

Despite different research targets and retrieval methods, the comparison shows that wrong choices of verbs and nouns are listed as the three most common categories of

Table 1 Distribution of V-N miscollocation types and tokens among the categories of misuse

Categories of misuse	Types <i>N</i> (%)	Tokens <i>N</i> (%)
Verb-based deviation		
Erroneous verb choice, e.g., *open the television	331 (58.1)	1818 (40.0)
Erroneous preposition after verb, e.g., *listen the radio	108 (18.9)	1714 (37.7)
Misuse of de-lexicalized verb, e.g., *do a mistake	15 (2.6)	267 (5.9)
Noun-based deviation		
Erroneous noun choice, e.g., *have a travel	51 (8.9)	432 (9.5)
Incomplete noun phrase, e.g., *buy the lottery	24 (4.2)	122 (2.7)
Other deviation		
Erroneous syntactic structure, e.g., *prevent the environment	10 (1.8)	52 (1.1)
Redundant repetition, e.g., *achieved many achievements	3 (0.5)	13 (0.3)
Erroneous use of idioms, e.g., *burn the night	2 (0.4)	14 (0.3)
Mixed deviation, e.g., *slow modernization	24 (4.2)	101 (2.2)
Unclear meanings, e.g., *presides the same truth	2 (0.4)	8 (0.2)
Total	570 (100)	4541 (100)

misuse in all three studies. This indicates that the semi-automatic method used in the current study generates similar results as manual retrieval methods and shows that the proposed method is a feasible alternative to efficiently retrieve second language learners' collocation errors in a large corpus. Moreover, most V-N collocation errors were attributed to the incorrect use of verbs in the current study, which corroborates the findings of previous corpus-based research (e.g., [16, 18, 19, 24, 26, 29–31, 41, 47, 48]).

Though the results of this study share some similarities with those of previous research, differences were also found. The first difference is the percentage of Verb-Preposition-Noun (V-P-N) errors. The current study, for example, found that 37.7% of miscollocations were due to misused prepositions after verbs, a much higher figure than that found by Zhang and Yang [48]. One possible reason might be the different data retrieval methods employed. Potential V-N miscollocations in Zhang and Yang's study were retrieved by locating errors tagged as 'CC3' from the error-tagged CLEC. Errors involving prepositions, however, were tagged as prepositional, and most of these errors

Table 2 Comparison of the three most common categories of misuse in previous studies and the current study

Top 3 common categories of misuse		
Nesselhauf [31]	Zhang and Yang [48]	Present study
1. Wrong choice of verb	1. Erroneous verb choice	1. Erroneous verb choice
2. Wrong choice of noun	2. Misuse of de-lexicalized verb	2. Erroneous preposition after the verb
3. Whole combination inappropriate	3. Erroneous noun choice	3. Erroneous noun choice

were thus not retrieved in their study. This was acknowledged by Zhang and Yang, who commented that the low percentage of Erroneous Preposition after Verb does not reflect the learners' better mastery of these V-N collocations [48]. In contrast, the *Sketch-Diff* function adopted in this study retrieved a great number of V-P-N miscollocations by comparing corpus data from native speakers with that from Chinese-speaking EFL learners, revealing that this kind of V-N collocation was problematic for the learners. This suggests that it may be important to include this construction in any future analysis and the teaching/learning of V-N collocations [48].

Another difference between the results of this study and those of Zhang and Yang's [48] was the number of V-N miscollocations resulting from the misuse of de-lexicalized verbs. While this error category ranked the second among the 12 error categories in Zhang and Yang, only 5.9% of V-N miscollocations were assigned to this category in the current study. One plausible explanation might be the stricter baselines for retrieving miscollocations in this study. As described in the method section, collocations appearing more than once in BNCCOCA were considered as acceptable combinations and therefore excluded from the analysis. It is possible that V-N miscollocations resulting from misused de-lexicalized verbs also appear, though less frequently, in the native reference corpus. Due to the stricter threshold of this study, some potential miscollocations might have been filtered out at the first stage. Future studies are thus suggested to include all potential miscollocates in the red area for further analysis.

Possible Causes of Chinese EFL Learners' V-N Miscollocations

Table 3 illustrates the distribution of V-N miscollocations across the four main causes of misuse. It should be noted that, under approximation, four sub-causes were identified in the current study, including misuse of a (near)-synonym, misuse of a hypernym/hyponym, misuse of an antonym, and misuse of lexeme with similar form/sound.

Negative L1 transfer refers to the negative influence of a learner's L1 on their production of the L2. Of the 1552 instances of L1 interfered miscollocation, 1410 were attributed to a direct Chinese-English translation of either the verb collocates (e.g., “*eat medicine” instead of “take medicine”) or the noun collocates (e.g., “*pay

Table 3 Distribution of V-N miscollocation types and tokens among the causes of misuse

Cause of misuse	Types	Tokens
	<i>N</i> (%)	<i>N</i> (%)
Negative L1 transfer	277 (48.6)	1552 (34.2)
Undergeneralization	120 (21.1)	1755 (38.6)
Approximation		
Misuse of a (near)-synonym	90 (15.8)	634 (14.0)
Misuse of a hypemym/hyponym	13 (2.3)	83 (1.8)
Misuse of an antonym	1 (0.2)	4 (0.1)
Misuse of lexeme with similar form/sound	13 (2.3)	60 (1.3)
Overgeneralization	28 (4.9)	334 (7.4)
Total	570 (100)	4541 (100)

strength” instead of “pay efforts”). This is illustrated in the concordance lines (1) and (2):

- (1) If I **eat the medicine* that it can let me live longly, I can know what it happens in the future.
- (2) It cannot be a lucrative job if they always have to **pay double strength* to take care but half of the harvest they earned.

In addition to directly translating L1 verb/noun collocates, Chinese-speaking EFL learners were also found to transfer L1 concepts of certain noun collocates to form incomplete noun phrases in the L2, such as “**want to learn computer well*” (instead of “*want to learn computer skills well*”). This is illustrated in the concordance line (3):

- (3) If you want to **learn computer well*, you can just go to play it.

In Chinese, the noun *diànnǎo* (i.e., *computer*) can denote both the tangible object (e.g., the machine) and intangible concepts (e.g., computer skills) that are related to computers, and it is thus acceptable to use the word *diànnǎo* to form collocations such as *xué diànnǎo* (i.e., to *learn computer skills*). In English, however, the word *computer* only denotes tangible object. It is likely that the learners negatively transferred the Chinese concept of *diànnǎo* and produced these miscollocations.

In this study, undergeneralization (i.e., a failure to obey the restrictions of an existing structure) is also influential in the formation of miscollocations. Of the 1755 occurrences of these miscollocations, 1700 instances resulted from a missing preposition after a verb collocate (e.g., “**adapt the society*” instead of “*adapt to society*” and “**agree this view*” instead of “*agree with this view*”). This is illustrated in concordance lines (4) and (5):

- (4) So children should foster the awareness of the competition to **adapt the society*.
- (5) According to the passage following, I **agree this view*.

Learners’ ignorance of prepositions in these V-N miscollocations may be because V-P-N collocations are rare in their L1. The inclusion of a preposition after a verb in English V-N collocations is relatively common, but this type of combination is rarely seen in Chinese. Most Chinese V-N collocations are formed with the noun directly following the verb. It is likely that the learners applied the rules of Chinese V-N collocations in their English production and thus caused a great number of these miscollocations.

The third most common cause of V-N miscollocations was approximation. A majority of the 781 occurrences of approximation errors were attributed to the misuse of a (near)-synonym (e.g., “**have a travel*” instead of “*have a trip*”). This is illustrated in the concordance line (6):

- (6) Several weeks ago, I **had a travel* to Kenting National Park.

Another example of learners’ misuse of near-synonyms is the use of *look*, *watch*, and *see*. The learners were frequently found to collocate the verb *look* with other noun collocates where *watch* and/or *see* were more appropriate (e.g., “**look some animal*”

instead of “see some animals” and “*look the news” instead of “watch the news”). This is illustrated in concordance lines (7) and (8):

- (7) Because I want to **look some animal* is Australia.
 (8) I had even **look the news* on TV.

In addition to misuse of near-synonyms, the learners were also found to occasionally misuse a hypernym/hyponym (e.g., “*own a degree” instead of “have a degree”), or antonym (e.g., “*oppose the fact” instead of “can’t accept the fact”) in forming collocations. This is illustrated in concordance lines (9) and (10):

- (9) However, many young people were misled and thought to **own a higher degree* is the only way out.
 (10) Because of the complex reason between these two families, both of them **opposed this fact*.

Furthermore, the learners were sometimes observed to form miscollocations by misusing a lexeme with similar sound/form (e.g., “*effect the world” instead of “affect the world” and “*release the pain/burden” instead of “relieve the pain/burden”), as illustrated in concordance lines (11) and (12):

- (11) Second, on the contrary, our life can **effect the media world*.
 (12) It can **release the pain* physically and mentally.

According to WordNet 3.1, the word *affect* is more frequently used as a verb to express the meaning “to have an effect upon something,” while the word *effect* often serves as a noun to mean “a phenomenon that follows and is caused by some previous phenomenon.” It is possible that the learners confused the two words and misused them interchangeably because their forms and meanings resemble each other. Similarly, the forms as well as the semantic meanings of *release* and *relieve* are also partially similar, and these similarities might thus confuse the learners.

The next most common cause of miscollocations was overgeneralization, referring to the incorrect application of a deviant structure instead of the appropriate one. One example is “*pay emphasis on modernization,” which might be attributed to the expression “pay attention to something.” This is illustrated in the concordance line (13):

- (13) Tradition seems outdatedly because someone **pay too much emphasis* on modernization.

In general, findings regarding the common causes of V-N miscollocations made by Chinese-speaking EFL learners are similar to those identified in previous studies, with negative L1 transfer and undergeneralization being the major causes. Based on the findings, some suggestions for the teaching/learning of English V-N miscollocations are offered. First, a complete listing of the 4541 miscollocations in this study is provided as supplementary material (see Online Resource 1–3). Teachers and material writers are suggested to include these miscollocations as well as the correct alternatives in their teaching and/or materials. By helping learners notice these common errors and

the correct usages, learners might be more aware of their production of V-N collocations and thus be less likely to make errors. In addition, when teaching intransitive verbs, teachers are suggested to introduce these verbs with their prepositions as complete units, which might reduce the possibility of making V-P-N miscollocations. Lastly, teachers and materials writers are also encouraged to introduce and differentiate some of the commonly misused synonymous verbs/nouns in their teaching and/or materials so as to enhance learners' understanding of the collocational restrictions of different synonyms.

Conclusion and Suggestions for Future Research

This study explored Chinese-speaking EFL learners' V-N miscollocations in a 7.4-million-word learner corpus through the use of the online corpus analysis tool *Sketch Engine*. With the help of the *Sketch-Diff* function, V-N collocations retrieved from the learner corpus were compared to data in a native reference corpus to semi-automatically reveal potential V-N miscollocations, and 4541 tokens (570 types) of V-N miscollocations were then identified with human inspection. Analysis of these miscollocations revealed that most errors were verb-based, such as inappropriate verb choice or missing prepositions after verb collocates. It was also found that many of the miscollocations were caused by negative transfer from the learners' L1 or undergeneralization of L2 syntactic rules, while some miscollocations were due to the misuse of semantically related words.

The research results show that using *Sketch-Diff* to semi-automatically retrieve potential V-N miscollocations from large corpora not only is feasible, but also would allow researchers to quickly extract potential miscollocations for further analysis. Compared to the traditional labor-intensive method of manual retrieval, this approach is much less time-consuming when studying common types of V-N miscollocations in larger learner corpora. Moreover, *Sketch-Diff* can also be used to analyze over ninety languages including French, German, and Spanish. The results could benefit L2 teachers who seek to better understand high-frequency collocation errors in many interlanguages. Nevertheless, one suggestion for the better use of *Sketch-Diff* is proposed here by the researchers. While many miscollocations caused by missing prepositions were identified, other types of prepositional error after the verb collocates (i.e., replacement or addition of a preposition) were not retrieved. This is due to the *SkE* platform's discrete categorization of Verb-Noun and Prep-Noun into two different groups. Because the current study only examined potential miscollocations under the category *_object of*, other potential types of V-P-N miscollocations might have been overlooked. It is thus suggested that future research should include an analysis of combinations under the category of *pp_* to identify these types of V-P-N miscollocation. Another suggestion for the better use of *Sketch-Diff* is to correct misspelled words in a given learner corpus before uploading the data unto *SkE*. Since some false alarms are caused by misspelling, correcting these spelling errors beforehand can thus reduce the chances of potential false alarms.

By adopting a computer-assisted analysis that integrates semi-automatic error retrieval and human analysis, this study explored common V-N miscollocations produced by Chinese-speaking EFL learners. The findings revealed that this

method can generate results at least on a par with those generated by manual retrieval in previous studies, suggesting that it could better retrieve potential collocation errors in large learner corpora. Future studies employing the proposed method to investigate L2 learners' collocation errors in much bigger corpora are encouraged.

Funding Information This work was financially supported by the “Chinese Language and Technology Center” of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

- Aitchison, J. (1987). *Words in the mind*. Hoboken: Blackwell Publishing.
- al-Hassnawi, H. (2017). Analysis of collocational errors in selected Iraqi published papers. *Adab Al-Kufa*, 31, 9–24.
- Altenberg, B. (1993). Recurrent verb-complement constructions in the London-Lund Corpus. *English language corpora: design, analysis and exploitation*, 227–245.
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101–114.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Brown, D. F. (1974). Advanced vocabulary teaching: the problem of collocation. *RELC Journal*, 5(2), 1–11.
- Chanel, J. (1981). Applying semantic theory to vocabulary teaching. *English Language Teaching Journal*, 35(2), 115–122.
- Chen, M. H. (2008). *A study of English collocation competence of college students in Taiwan*. Master's thesis, National Taiwan University of Science and Technology, Taiwan.
- Chen, H. J. H. (2011). Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer assisted language learning*, 24(1), 59–76.
- Chorbwhan, R., & McLellan, J. (2016). First language transfer and the acquisition of English collocations by Thai learners. *Southeast Asia: A Multidisciplinary Journal*, 16, 16–27.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Ellis, N. C. (1996). Sequencing in SLA. *Studies in second language acquisition*, 18(1), 91–126.
- Fan, M. (2009). An exploratory study of collocational use by ESL students—a task based approach. *System*, 37(1), 110–123.
- Gitsaki, C. (1999). *Second language lexical acquisition: a study of the development of collocational knowledge*. Maryland: International Scholars Publications.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and lexical phrases. In A. P. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 145–160). Oxford: OUP.
- Hong, A. L., Rahim, H. A., Hua, T. K., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: a corpus-based error analysis. *3L: language, linguistics, literature®*, 17(S1), 31–44.
- James, C. (1998). *Errors in language learning and use: exploring error analysis*. Abingdon: Routledge.
- Juknevičienė, R. (2008). Collocations with high-frequency verbs in learner English: Lithuanian learners vs native speakers. *Kalbotyra*, 59(3), 119–127.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Language learning*, 61(2), 647–672.
- Lewis, M. (2000). *Teaching collocation: further development in the lexical approach*. London: Language Teaching Publications.
- Li, C. C. (2005). *A study of collocational error types in ESL/EFL college learners' writing*. Master's thesis, Ming Chuan University, Taiwan.
- Lien, H. Y. (2003). *The effects of collocation instruction on the reading comprehension of Taiwanese college students*. PhD dissertation, Indiana University of Pennsylvania, US.
- Liu, C. P. (1999). An analysis of collocational errors in EFL writings. In *The proceedings of the eighth international symposium on English teaching* (pp. 483–494). Taipei: Crane Publishing Co., Ltd..

24. Liu, L. E. (2002). *A corpus-based semantic investigation of verb-noun miscolllocations in Taiwan learners' English*. Master's thesis, Tamkang University, Taiwan.
25. Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing* (Vol. 999). Cambridge: MIT Press.
26. Marco, M. J. L. (2011). Exploring atypical verb+noun combinations in learner technical writing. *International Journal of English Studies*, 11(2), 77–95.
27. Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320.
28. Matsuda, A., & Matsuda, P. K. (2010). World Englishes and the teaching of writing. *Tesol Quarterly*, 44(2), 369–374.
29. Men, H. (2016). *Vocabulary increase and collocation learning*. Shanghai: Shanghai Jiao Tong University Press.
30. Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.
31. Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
32. Nguyen, T. M. H., & Webb, W. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298–320.
33. Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149.
34. Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. *Language and communication*, 191, 225.
35. Rychlý, P. (2008). A lexicographer-friendly association score. In *RASLAN* (pp. 6–9).
36. Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.
37. Sinclair, J. (1991). *Corpus, concordance, collocation*. Hong Kong: Oxford University Press.
38. Stenson, N. (1983). Induced errors. In B. W. Robinett & J. Schachter (Eds.), *Second language learning: contrastive analysis, error analysis and related aspects* (pp. 256–271). Ann Arbor: University of Michigan Press.
39. Van Rooy, B., & Schäfer, L. (2002). The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African linguistics and applied language studies*, 20(4), 325–335.
40. Wang, C. J. (2001). *A study of the English collocational competence of English majors in Taiwan*. Master's thesis, Fu Jen Catholic University, Taiwan.
41. Wang, Y., & Shaw, P. (2008). Transfer and universality: collocation use in advanced Chinese and Swedish learner English. *ICAME journal*, 32, 201–232.
42. Woolard, G. (2000). Collocation-encouraging learner independence. In M. Lewis (Ed.), *Teaching collocation: further developments in the lexical approach* (pp. 28–46). Hove: Language Teaching Publications.
43. Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
44. Wray, A. (2008). *Formulaic language: pushing the boundaries*. Oxford: Oxford University Press.
45. Wu, W. S. (1996). Lexical collocations: one way to make passive vocabulary active. In *The Proceedings of the 11th Conference on English Teaching and Learning in the Republic of China* (pp. 461–480).
46. Zhang, X. (2017). Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. *System*, 66, 156–167.
47. Zhang, Y., & Gao, Y. (2006). A CLEC-based study of collocation acquisition by Chinese English language learners. *CELEA Journal*, 29(4), 28–35.
48. Zhang, W. Z., & Yang, S. (2009). An analysis of V-N collocation errors in CLEC. *Journal of PLA University of Foreign Languages*, 32(2), 39–44.

Affiliations

Christine Ting-Yu Yang¹ · Howard Hao-Jan Chen¹ · Chen-Yu Liu¹ · Yu-Cheng Liu²

✉ Howard Hao-Jan Chen
hjchen@ntnu.edu.tw

Christine Ting-Yu Yang
christiney37@gmail.com

Chen-Yu Liu
chenyuliu1019@gmail.com

Yu-Cheng Liu
liuyucheng.allen@gmail.com

¹ English Department, National Taiwan Normal University, 162, Section 1, Heping E. Rd, Taipei City 106, Taiwan

² Taipei First Girls High School, 165, Section 1, Chongqing S. Rd, Taipei City 100, Taiwan