COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2019.1607880

Using collocation clusters to detect and correct English L2 learners' collocation errors

Ping-Yu Huang[a] and Nai-Lung Tsao[b]

[a]General Education Center, Ming Chi University of Technology, New Taipei City, Taiwan, R.O.C.; [b]Office of Information Services, Tamkang University, New Taipei City, Taiwan, R.O.C.

ABSTRACT

In this article, we describe an online English collocation explorer developed to help English L2 learners produce correct and appropriate collocations. Our tool, which is able to visually represent relevant correct/incorrect collocations on a single webpage, was designed based on the notions of collocation clusters and intercollocability proposed by Cowie and Howarth. As they pointed out, in a collocation cluster L2 learners generally cannot distinguish true collocations (e.g., *tell truth*, *state truth*, and *state fact*) from impossible combinations (e.g., **say fact* and **say truth*). Accordingly, our tool applies natural language processing techniques to construct collocation clusters to enable learners to easily differentiate between correct and incorrect pairs. Relying on data from a reference corpus, our system instantaneously processes the collocability of users' target combination (verb–noun or adj–noun) and all other relevant words and presents true/false collocations that L2 learners should master/avoid. To assess our tool, we investigated its performance in detecting and correcting learners' V–N and A–N errors, with results comparable to those of most previous studies. Piloted using a sample of 13 intermediate- or upper-intermediate level English as a foreign language learners, our tool was found to help them self-correct their collocation errors effectively. Compared with similar tools or approaches, our tool requires much less data resources, but still demonstrates a remarkable capability to detect/correct errors and generate useful collocational knowledge in English.

KEYWORDS
Collocation cluster; collocation error detection; collocation error correction; digital reference tool

Introduction

A number of second language acquisition (SLA) studies have empirically demonstrated that collocations present a tough problem for L2 (second language) learners (e.g., Bahns & Eldaw, 1993; Granger, 1998;

CONTACT Ping-Yu Huang alanhuang25@hotmail.com General Education Centre, Ming Chi University of Technology, No. 84 Gungjuan Rd., Taishan Dist., New Taipei City 24301, Taiwan R.O.C.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ncal.

Nesselhauf, 2003). Such a learning problem, according to Laufer (2011), is commonly present in learners' output language but largely results from learners' inability to perceive collocations as formulaic language in input. L2 learners generally have little or no difficulty comprehending the meanings of collocations in speech or writing, but they also do not (consciously) realize or (subconsciously) learn which word combinations in input constitute collocations. As they attempt to produce a collocation, they thus do not know which words tend to co-occur and generate word pairs which are "semantically possible but collocationally impossible" (Nesselhauf, 2003, p. 239).

To help L2 learners overcome these difficulties, CALL researchers have thus focused on either learners' input or output, and during the past two decades have reported fruitful results. Concerning input, for example, CALL researchers utilize a data-driven learning (DDL) approach that encourages learners to discover linguistic patterns with corpus-based concordancers. Subjects in general were found to enjoy doing concordancing searches and to improve in their collocational knowledge (Cobb, 1999; Lee & Liou, 2003). Such improvements, however, did not last long and tended to regress as time went on (Chan & Liou, 2005). Moreover, the DDL approach usually takes teachers considerable time to prepare materials (Boulton, 2010), and when practiced in classrooms requires much effort from students to decide which combinations are proper (Chan & Liou, 2005). Concerning output, in addition to web-based retrieval tools developed to help learners generate collocations mentally (Chen, 2011; Chen, Wu, Yang, & Pan, 2016; Davies, 2008; Yeh, Liou, & Li, 2007), much CALL research addresses the detection and/or correction of L2 collocation errors. These detection/correction studies were conducted because it is expected that some reliable assisting tools can be provided to help learners produce correct and proper collocations. In this article, our main purpose is to report an online English collocation exploration tool that has been developed for intermediate- or higher-level English L2 learners to verify the word pairs that they use, and to easily differentiate correct from incorrect collocations. In the past 7–8 years, although various natural language processing (NLP) techniques have been proposed for collocation error detection/correction, stable tools or systems that are either freely or commercially available for learners have been rare. Our tool, which is readily available to all English learners, employs a resource-friendly technique, collocation clusters (Cowie & Howarth, 1996), to extract useful lexico-grammatical knowledge from corpora. Our technique basically requires little computational resources but demonstrates collocation error detection/correction performances comparable to those of most previous studies.

Literature review

In the literature, collocations do not have a precise, generally accepted definition. Whereas the phraseological approach to language (e.g., Cowie, 1981; Mel'čuk, 1998) identifies collocations on the basis of semantic properties, the frequency-based approach (e.g., Sinclair, 1991) defines collocations as combinations showing strong co-occurrence strengths in texts. A combination of the two approaches has been adopted by many researchers; Nation (2001), for instance, regards a collocation as a string of lexical items that frequently co-occur and "have some degree of semantic unpredictability" (p. 317). Based on their grammatical features, moreover, collocations have often been divided into two categories: lexical and grammatical collocations (Benson, Benson, & Ilson, 1986). The former refers to those composed of two open-class words, the latter those containing a content and a function word.

In addition to semantic unpredictability, collocations are characterized by the restricted co-occurrence relationships of their components. Such restrictedness, however, seriously confuses L2 learners. L2 learners usually do not know which words constitute collocations whose components cannot be freely replaced by semantically similar words. Collocational knowledge thus reflects knowledge of a language (Nation, 2001), and applied linguists, especially those working on corpus investigations, strive to develop techniques that can effectively detect L2 collocation errors and/or further suggest appropriate corrections.

Collocation error detection

Detection of L2 learners' content word errors has long been considered a more challenging task than the identification of function word errors (Kochmar & Briscoe, 2014, Kochmar & Briscoe, 2015). As function words, such as articles and prepositions, are closed classes of words, an error involving them can be easily detected if an alternative is found to be more appropriate in its context. Content word errors, such as wrong verb–noun pairs (e.g., **eat medicine*), are difficult to detect, as the alternatives making up a confusion set might be infinite.

Wible, Kuo, Tsao, Liu, and Lin (2003) focused on verb–noun errors that were found to cover a substantial proportion of English as a foreign language (EFL) learners' collocation problems. Some of the errors being investigated were collected when their producers (learners) submitted essays to teachers that the teachers marked and provided comment tags on. The teachers' tags bootstrapped, so that based on those marks the researchers could identify more errors and conduct further error extractions. To detect errors, Wible et al. (2003) targeted nine nouns, each paired with a list of previously identified misused verbs. A grammar checker was designed to retrieve errors from a learner corpus whenever one of the nine nouns was accompanied by a misused verb. The precision rate of the grammar checker achieved 95.5%. Similarly, Chang, Chang, Chen, and Liou (2008) also concentrated on verb–noun errors made by EFL learners. However, unlike the former, which made use of teachers' tags, the latter adopted a comparison approach. The standard comparison data used by Chang et al. (2008) were collocations retrieved from the British National Corpus ([BNC], 2007). Chang et al. (2008) employed two NLP techniques, chunk processing and clause parsing, for collocation identification. In chunk processing, a sentence is segmented into smaller syntactic patterns or chunks such as noun and verb phrases. Verb–noun collocations could mostly be identified in verb phrases. For certain syntactic structures like relative clauses (e.g., *… the email he sent to the mayor was …*), however, chunk processing was inadequate. The authors thus further conducted clause parsing to correctly identify clauses and collect more collocations. Chang et al. (2008) used the gathered BNC collocations to examine EFL learners' verb–noun errors listed in Liu (2002), and found that they achieved a detection precision of 90.7%. The high precision rates reported in these two studies indicated that both teachers' comments and standard reference collocations are useful resources for detecting collocation errors.

These two approaches, though enjoying high precision rates, may be limited in that they require either a large number of previously collected teacher comments or highly reliable standard data. Rare but true collocations, for example, would be wrongly detected by the comparison approach. To avoid such limitations, in the recent decade some innovative approaches have been proposed. The precision values of those new attempts might not be high (e.g., Futagi, Deane, Chodorow, & Tetreault, 2008's rank-ratio approach), but they do provide more possibilities for future investigations. An important recent approach was introduced by Kochmar and Briscoe (2014). Specifically, in this study, Kochmar and Briscoe (2014) examined compositional distributional semantics approaches to develop a supervised classifier involving 13 semantic features. Among these, a key feature being examined was adjective identity, which, based on a list of common adjective–noun errors, was useful in deciding whether an adjective had a greater chance of being an error (e.g., *big* in **big wind*). The supervised classifier discovered that the mixed-feature model that produced the best performance used adjective identity, ranked density in close proximity, and distance between model-generated adjective–noun pairs and input nouns as main features, and their accuracy rate reached 75%. It is worth mentioning that an important contribution of Kochmar and Briscoe (2014) was to further investigate combinations that are incorrect in most contexts but acceptable in particular contexts (e.g., *classic dance* as an error in "… *a rock'n'-roll dance and a classic dance* …," p. 1741, and as a correct usage in "*classic Ceilidh dance*," p. 1743). Kochmar and Briscoe's (2014) classifier demonstrated a satisfying rate of 62% in the task of judging such pairs as correct usages. This study highlighted the uniqueness of such usages and pointed out the risk that certain acceptable collocations might be misidentified as errors by previous approaches.

Collocation error correction

Compared with error detection research, studies addressing the automatic correction of collocation errors in general showed lower precision values. The latter issue is challenging because good corrections not only have to be true and common collocations, but semantically should be closely similar to a target error.

Unlike detection research, which mostly utilizes native speakers' productions as standard references for comparison, most correction studies rely on learners' L1 as the main resource. The reason for this is straightforward. L2 learners very often have little idea which words constitute frequent word pairs; when they intend to create one, they tend to translate an L1 expression directly into L2. Such translated strings undoubtedly are likely to be errors. A classic example offered by Liu (2002) was **write homework*. When Chinese learners of English, especially lower-level ones, would like to say *do homework*, they sometimes translate the Chinese collocation *xie3 gong1 ke4* into English. Whereas in Chinese it is fine to *write homework* or *compose assignment*, in English both are unacceptable combinations. Chang et al. (2008) made use of this sort of translation equivalent to correct English L2 errors. They examined the verb errors in verb–noun pairs reported by Liu (2002), and their system searched in bilingual dictionaries for alternative verbs that shared translation equivalents with the errors. Searched alternatives were then put back into original contexts, and the frequency and Log Likelihood Ratio (Jian, Chang, & Chang, 2004) in the BNC were used to determine whether good collocations were formed. The authors evaluated their approach based on two types of data. First, they observed whether the gold answers suggested by English instructors and reported in Liu (2002) were satisfactorily covered by their system-generated alternatives. Second, they looked over their top ten alternatives for each error, attempting to acquire good collocates that ranked higher than those gold answers. Chang et al. (2008) approach achieved impressively high precision rates of 84.4 and 94.1% for both data sets. The quality of Chang et al. (2008) corrections, additionally, was also impressive as the MRR (mean reciprocal rank[1]) of their corrections reached 0.66, a value hardly surpassed by similar studies in the literature.

Dahlmeier and Ng (2011) and Kochmar and Briscoe (2015) also utilized L2 learners' native languages as major resources in error correction. The precision values both studies reported, however, were somewhat lower than Chang et al. (2008). In Dahlmeier and Ng (2011), the authors observed that an error (either a verb or a noun in V–N pairs) might be caused by a similar spelling, the same pronunciation, similar word meanings, or shared L1 equivalents. They thus investigated the effects of editing distance (for spelling), homophones, WordNet[2] resources, and words with shared L1 translations on suggested alternative collocates. However, unlike Chang et al. (2008), who employed bilingual dictionaries, Dahlmeier and Ng (2011) utilized a parallel corpus for L1–L2 translation resources. Their reported precision rate (25%) does not seem very impressive at first sight. The main reason for such a low figure might be that their system only checked whether a single gold answer was included in ten system-generated alternatives. Their precision rate was sure to improve if all top ten answers could be manually evaluated as correct/incorrect alternatives.[3] Kochmar and Briscoe (2015) similarly adopted a hybrid approach. Specifically, Kochmar and Briscoe's (2015) algorithm searched for orthographically or semantically similar words as well as annotators' corrections as alternative collocates. An innovative idea that the researchers experimented with for annotator corrections was to rank the corrections based on learner corpus data. The learner dataset that they consulted was the Cambridge Learner Corpus. In the dataset, all errors in adjective–noun combinations had been annotated with certain alternatives (e.g., *big, high,* and *loud* as confusion words for *strong* in the word pair **strong noise*). To determine their appropriateness as good corrections, Kochmar and Briscoe (2015) not only considered their frequency and association strengths in a standard corpus, but further calculated their probabilities in learner productions in comparison with the replaced words. This approach, as the researchers explained, took into account both native and learner languages to determine whether a correction was appropriate. Kochmar and Briscoe's (2015) system worked satisfactorily with the errors from the Cambridge Learner Corpus. Among the errors, 71.2% of them were found to be corrected by alternative collocates, and the MRR was 0.506. These values were indeed impressive considering that the alternatives were gold answers, rather than correct replacements judged manually. This approach performed less effectively for other learner corpus data, with around 50% of errors being covered and corrected.

Correction research using L1–L2 bilingual dictionaries, like detection research using teacher annotations, demonstrated high precision rates. However, if we intend ESL or EFL learners to actually benefit from these approaches, we have to first collect abundant comment tags from teachers as well as rich learners' L1 resources. The latter is hardly available considering that there are 60 countries in which English is used as an important L2 (Seargeant, 2016) and there are over two billion people learning English worldwide. What researchers should develop, then, is a tool or an approach that is inexpensive in terms of the resources it consumes, and can consistently provide useful collocational knowledge. Such tools ideally also need to be suitable for all English L2 learners to gain collocational knowledge no matter their L1 background.

Collocation clusters

The notions of collocation clusters and intercollocability were first proposed by Cowie and Howarth (1996). According to them, the traditional and common view that collocations are fixed word strings that recur frequently in language is misleading. Collocations, more specifically restricted ones, mostly are not fixed but admit of limited variations that are arbitrary and exist for no particular reason. In text frequency, most collocations are not as frequent as previous studies had assumed. Cowie and Howarth's (1996) observation was that collocations repeatedly appear to a limited degree across texts, which is not denying the significance of collocations; rather, they sought to formulate a more precise definition or description of collocations for fellow researchers. Based on their analysis, collocations should be described as familiar and memorized lexical combinations that vary to a limited extent. Another common but not necessary characteristic of collocations is semantic opaqueness, which is optional as not all collocations are semantically opaque.

On the basis of these definitions and clarifications, Cowie and Howarth (1996) divided restricted collocations into four categories: invariable collocations, collocations with limited choice at one point, collocations with limited choice at both points, and overlapping collocations. Invariable collocations (e.g., *make amends*), as the name suggests, are completely fixed pairs showing no variations. The second and third types, collocations allowing limited variations at one or both ends, include instances such as *provide/gain/obtain access to noun phrases* and *encounter/experience—difficulty/problem*. The first three kinds of collocations, which allow either no or very limited variation, generally pose little difficulty for English L2 learners. What indeed confuses L2 learners is overlapping collocations. Overlapping collocations are problematic and confusing because in language use they show the characteristics of both openness and restrictions. An example that Cowie and Howarth (1996) discussed is the cluster *convey/express/*communicate/*get across—regret/condolence*. While all the four verbs in this cluster can co-occur with neutral message nouns (e.g., *point*, *opinion*, *message*, etc.), the noun position is restricted when the messages to be conveyed are expressions of sad feelings. This cluster is difficult for English L2 learners to grasp or master because most learners cannot discern that *convey/express* collocate with *regrets/condolences* whereas *communicate/get across* do not. The intercollocability of this cluster is limited and it seems almost impossible for L2 learners to figure this out based on only their oral or written input. The phenomenon that *communicate/get across* cannot collocate with *regrets/condolences* is idiosyncratic, and the successful mastery of such idiosyncratic usages clearly differentiates native speakers from L2 learners, even advanced ones.

To our knowledge, Liu, Wible, and Tsao (2009) is the only study that applied Cowie and Howarth's (1996) collocation clusters to analyze L2 learners' collocation errors. Their purposes specifically were to propose a probabilistic model that could correct verb errors in L2 verb–noun pairs. Liu et al.'s (2009) model involved three main features: word association strength, semantic similarity, and intercollocability. The association strength and semantic similarity that the authors evaluated were based on mutual information (MI, Church & Hanks, 1990) and WordNet lexical relations, respectively. Concerning intercollocability, Liu et al. (2009) extended Cowie and Howarth's (1996) concepts of collocation clusters to error correction. An example offered by Liu et al. (2009) was *reach purpose*. In English, it is fine to say *fulfil goal*, *fulfil purpose*, *achieve goal*, *achieve purpose*, or *reach goal*, but *reach purpose* sounds odd and is basically an error. The verbs and nouns involved in this example thus form a restricted collocation cluster. Liu et al.'s (2009) assumption was that a good substitute would have many collocates that overlap those of a verb error. The researchers experimented with these three features and obtained results that were rather encouraging. Specifically, they found that intercollocability was as effective in error correction as word association strengths and semantic similarities, and the combination of all three yielded the best performance. Among the top ten corrections automatically gathered, the three features taken together demonstrated a high precision rate of 94%, almost as good as that reported by Chang et al. (2008). A limitation of Liu et al.'s (2009) study was that their method was tested on only one-third of rather than all the errors in Liu's (2002) data pool, which made a complete comparison between the two experiments unlikely.

Cowie and Howarth's (1996) notions of collocation clusters, as discussed in detail here, present a persuasive and precise account of L2 learners' problems with collocations. L2 learners very often do not know the restrictions of clusters and intercollocability of words, and consequently produce combinations that sound awkward to native speakers. The error correction approach proposed by Liu et al. (2009) further highlights the usefulness of correcting collocation errors based on collocation clusters. Considered together, then, these two studies motivated us to develop an effective reference tool that can consistently produce useful collocation clusters for L2 learners. Within the clusters generated, L2 learners should be able to easily distinguish true from false collocations. Such clusters enable L2 learners not only to test whether a combination is correct, but also to obtain all related combinations that they should use/avoid in their own productions.

Tool design

The web-based exploration tool that we developed is named *NetCollo*.[4] As the name suggests, NetCollo is meant to visually represent collocational networks or clusters; within the clusters, related correct/incorrect combinations are clearly listed together. On the NetCollo interface, when learners key in an incorrect collocation (e.g., **say truth*), the tool shows that this combination is not likely in English, with some possible alternatives provided (e.g., *state truth* and *tell truth*). The two sorts of information are not very difficult to acquire if learners are familiar with collocation dictionaries (e.g., *Macmillan Collocation Dictionary*, Rundell & Fox, 2010) and corpus concordancing tools (e.g., the COCA online tool, Davies, 2008). Developed based on the concepts of intercollocability, NetCollo is capable of offering more useful knowledge. NetCollo can further show: (1) other related combinations that also should be avoided by learners (e.g., **say fact* and **tell fact*) and (2) other related and possible collocations suggested to L2 learners (e.g., *state fact*). Figure 1 illustrates a collocation cluster generated by NetCollo.

In Figure 1, the word pair being keyed in and exemplified is the error **say truth*. The combinations in grey shading (e.g., **say truth* and **say fact*) or missing (e.g., **state reality*) are identified by NetCollo as incorrect or uncommon collocations. The pairs in white shading are those suggested to users as candidate collocations in English. The main standard corpus underlying the current version of NetCollo is the BNC. As Figure 1 shows, NetCollo judges **say truth* as a collocation error although there are many sentences containing it in the BNC. Specifically,

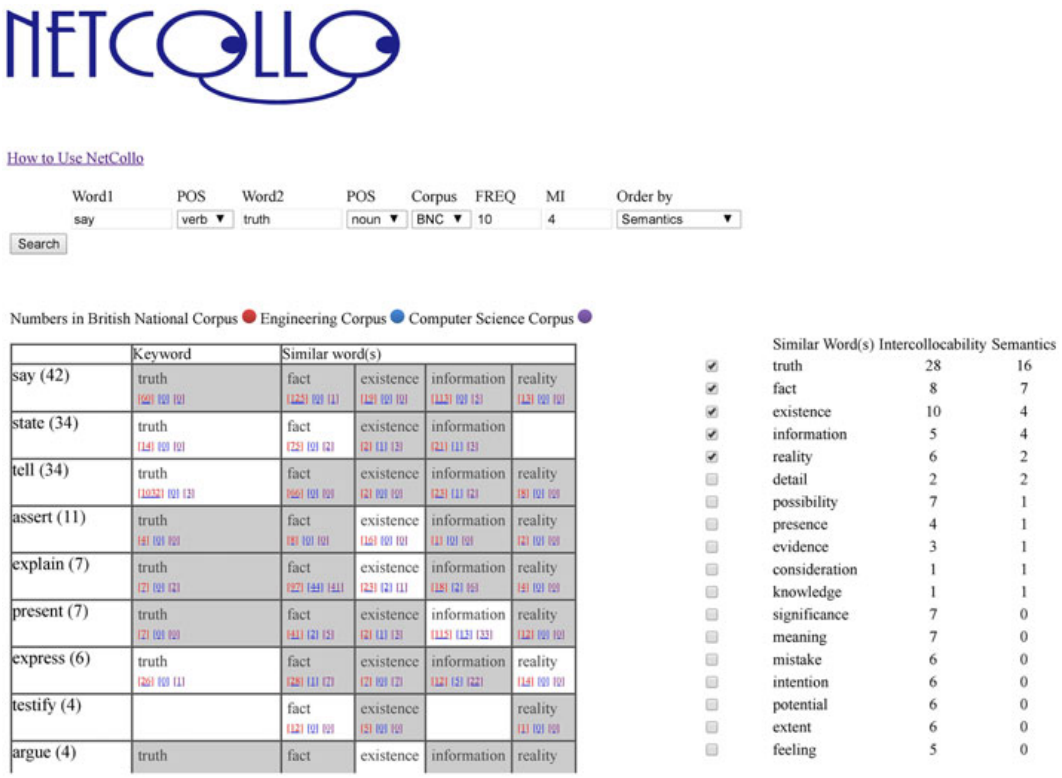

Figure 1. Search results for **say truth* on NetCollo.

the icon [60] in the **say truth* box indicates that 60 instances of it occur in the BNC. However, as users click on the [60] icon, they will find that most of the instances are in fact *say with truth* or are parts of other longer sequences (e.g., *He said I want the truth*). With such a cluster, learners can learn that they should avoid **say truth* but use *tell truth* or *state truth* instead. If learners look at the cluster more closely, furthermore, they will discover that they also should stop using odd combinations such as **say information* and **tell fact*, whereas *state fact* is a fine collocation in English.

NetCollo, which operates on corpus resources, utilizes a hybrid approach that considers both notions of intercollocability and semantic information to create collocation clusters. NetCollo generates lexical networks as follows. First, as two words (e.g., the wrong collocation **say truth*) are chosen by users and keyed in, NetCollo begins to automatically collect the collocates for each of them. The measures that we adopt to determine association strengths and find collocates are raw frequency and mutual information. The default scores for the two measures are set at 10 and 4.0, respectively. Next, NetCollo searches in the BNC for the words that share the most collocates with the second keyword. This is the crucial step of forming clusters because while doing so NetCollo attempts to find overlapping collocations based on users' key word pair(s). However, though at this step complete clusters are constructed, we found that the knowledge represented in those clusters is not pedagogically useful enough. That is, based solely on overlapping collocations, it was observed that some words that tend to share collocates with the key word pair(s) may be highly frequent words in English (e.g., *mistake* and *possibility*, two frequent nouns that semantically are not related to *truth* but share many collocates with it). To offer more useful clusters, among the generated networks we deemed it necessary to rank higher those words that semantically are closer to the searched keywords. We thus applied WordNet to obtain the necessary semantic information. We give words higher scores if they have synonym (2 points), hypernym (1 point), or hyponym (1 point) relationships in WordNet with the keywords.[5] By consulting WordNet, we found that the words scoring higher are indeed those we intended to acquire (e.g., *tell* for *say* and *fact* for *truth*). Consequently, we decided to rank the collected words within the formed clusters based mainly on semantic information or scores. Words showing no WordNet relationship with the keywords are then ordered based on how many collocates they and the keywords share.

The algorithms that we employ to generate collocation clusters are more sophisticated than those applied by Liu et al. (2009) and, as a result, the information NetCollo can provide is richer and more useful. In addition to identifying words sharing collocates with a wrong verb as Liu et al. (2009) did, NetCollo further processes all the collocabilities of the words embedded in a cluster. Using our algorithms, NetCollo can not only suggest alternatives to a wrong verb but show all the related correct and incorrect word pairs to learners. In the current version of NetCollo, we provide searches of both verb–noun and adjective–noun clusters. Furthermore, to meet the needs of some ESP (English for specific purposes) learners, on NetCollo we offer explorations in two domain-specific corpora: engineering and computer science. The sizes of these two technical corpora are around 10 and 12 million tokens, respectively. In compiling the two databases, we used journal articles coming from high-quality academic journals, with unwanted sections such as references and authors' affiliations removed. With the resources of the BNC as well as the two technical corpora, we expect that ESP learners can discover not only which pairs are correct/incorrect in English, but which collocations are specialized in certain domains and which are rather common in general use.

Tool evaluation and discussion

To evaluate the usefulness of the results of NetCollo, we assessed how NetCollo detected and corrected verb or adjective errors in verb–noun or adjective–noun combinations. First, regarding V–N pairs, we examined EFL learners' productions offered by Liu (2002). Among the 219 errors that she listed, however, it was found that there were several pairs that needed to be removed from analysis. Those pairs included, for instance, duplicate errors (e.g., **write homework*, appearing three times), noun-verb structures in which the nouns served as subjects (e.g., … *my heart couldn't move* …), noun collocates that were reflexives or pronouns (e.g., *do myself*), and combinations that were in fact acceptable and correct (e.g., *develop friendship*). Additionally, some of Liu's pairs were in fact noun problems rather than verb errors. In the wrong collocation *meet imagination*, for example, what its producer wanted to say was *expectation* rather than *imagination*. To collect a more reliable list of verb–noun errors for examination, we hired two experienced English instructors to judge Liu's (2002) data. In total, the two judges helped identify 31 duplicates, 6 noun errors, 3 correct usages, 3 noun–verb structures, and 16 non-collocation errors. Examples of the later four types are shown in Table 1. We removed all 59 of these items from our analysis, and the final V–N error list thus included 160 combinations. For each of those combinations, both judges confirmed that there was indeed an error in the verb position. For adjective–noun errors, we examined the CLC-FCE (Cambridge Learner Corpus-First Certificate in English) dataset, which serves as a resource for content word error analysis (Yannakoudakis, Briscoe, & Medlock, 2011). The CLC-FCE dataset, which contains 1,244 learners' exam scripts extracted from the Cambridge Learner Corpus, has been manually tagged for both adjective and noun errors (Nicholls, 2003). As with the V–N data, we focused on only the CLC-FCE combinations in which the errors appear in the adjective positions. These incorrect combinations, which mostly reflect similar orthography (e.g., **economical activity*, which should be *economic activity*) or similar word meanings (e.g., **full fun*, which should be *great fun*), collectively formed a list of 82 A–N combinations. The adjectives involved, as Kochmar and Briscoe (2014) analyzed, represented typical English L2 errors, as they were found to be extremely problematic for learners.

As explained earlier, NetCollo utilizes raw frequency and MI as main measures to extract collocates and, with the collocates collected, NetCollo employs intercollocability and semantic properties of words to construct collocation clusters. In this evaluation, the two measures were set at 10 and 4.0, respectively. In other words, a verb–noun or adjective–noun pair was regarded as a candidate collocation if its MI score was 4.0 or greater and it appeared at least 10 times in the BNC. For a few pairs in which the components appeared rarely in the BNC (e.g., **inspirit creativity*, **catch vocabulary*, and **conquer difficulty*), we set the MI/frequency measures at 4.0/5.[6] Basically, we adopted a rather strict threshold for MI and an easy one for frequency. In Ackermann and Chen's (2013) compilation of the Academic Collocation List, for example, for a word pair to

Table 1. Examples of removed verb–noun pairs.

Types	Examples	Correct usages
Noun errors	**meet imagination*	*meet expectation*
	**solve trouble*	*solve problem*
Non-collocation errors	**invent gravity*	*discover gravity*
	**remain him*	*remain himself*
	**do myself*	*be myself*
Noun-verb structures	**movie tell*	*movie shows*
	**aftershocks caused*	*aftershocks occurred*
Correct usages	*broaden knowledge*	–
	establish friendship	–

be considered a candidate academic collocation, it had to show up at least one time per million and hold an MI score higher than 3.0 in the authors' academic corpus. The reason that our frequency requirement was set at around one appearance per 10 million[7] was we observed that many true collocations were not very common in plain texts (e.g., *compose song*: 18 instances, and *endure pain*: 12 instances in the BNC). However, to avoid an extremely high recall resulting from such low frequency thresholds, we decided to use a higher MI. The two measures were tested on a series of verb–noun and adjective–noun pairs and were found to yield more reliable clusters than other MI-frequency combinations.[8] Furthermore, to state more clearly how NetCollo detected and corrected errors for this study, let us refer back to the **say truth* example. A complete cluster generated with this word pair by NetCollo is displayed in Figure 2. As Figure 2 illustrates, **say truth* was shown as an error, as it was in grey shading. Likewise, word pairs missing in the cluster were identified as wrong collocations as well. For error correction, the verb–noun pairs in white shading were suggested as alternative correct usages. The verbs were ranked by semantic proximity to the key verb and, if two (or more) verbs had the same proximity values (e.g., *state* and *tell*), they were further ranked by how many collocates they shared with the searched verb. NetCollo basically offers 18-verb × 5-noun (or 18-adjective × 5-noun) clusters. In such small clusters, users very often are provided with fewer than ten verb/adjective substitutes that appropriately collocate with the searched noun (e.g., only eight in the **say truth* example). Although larger networks are sure to offer more correct alternatives, we decide not to enlarge our default network size. The clusters or the collocational knowledge that learning or reference tools offer should be manageable and digestible for learners; large networks might take learners too much time to choose between several possible collocates.

Performance of collocation error detection

We fed the 160 verb–noun combinations extracted from Liu (2002) and the 82 adjective–noun combinations from the CLC-FCE dataset into

	Keyword	Similar word(s)			
say (42)	truth [60] [0] [0]	fact [125] [0] [1]	existence [19] [0] [0]	information [113] [0] [5]	reality [13] [0] [0]
state (34)	truth [14] [0] [0]	fact [75] [0] [2]	existence [2] [1] [3]	information [21] [1] [3]	
tell (34)	truth [1032] [0] [3]	fact [66] [0] [0]	existence [2] [0] [0]	information [23] [1] [2]	reality [8] [0] [0]
assert (11)	truth [4] [0] [0]	fact [8] [0] [0]	existence [16] [0] [0]	information [1] [0] [0]	reality [2] [0] [0]
explain (7)	truth [7] [0] [2]	fact [97] [44] [41]	existence [23] [2] [1]	information [18] [2] [6]	reality [4] [0] [0]
present (7)	truth [7] [0] [0]	fact [41] [2] [5]	existence [2] [1] [3]	information [115] [13] [33]	reality [12] [0] [0]
express (6)	truth [26] [0] [1]	fact [28] [1] [7]	existence [7] [0] [7]	information [12] [5] [22]	reality [14] [0] [0]
testify (4)		fact [12] [0] [0]	existence [5] [0] [0]		reality [1] [0] [0]
argue (4)	truth [6] [0] [0]	fact [33] [0] [1]	existence [15] [0] [1]	information [21] [0] [5]	reality [5] [0] [2]
suggest (4)	truth [4] [0] [0]	fact [53] [1] [3]	existence [31] [3] [8]	information [41] [4] [17]	reality [5] [0] [0]
acknowledge (0)	truth [15] [0] [0]	fact [34] [0] [2]	existence [42] [0] [1]	information [7] [2] [0]	reality [10] [0] [0]
recognise (0)	truth [13] [0] [0]	fact [55] [0] [0]	existence [27] [0] [0]	information [12] [0] [0]	reality [18] [0] [0]
reveal (0)	truth [50] [0] [0]	fact [18] [2] [2]	existence [30] [11] [2]	information [61] [11] [35]	reality [6] [0] [0]
ignore (0)	truth [5] [0] [0]	fact [167] [1] [3]	existence [20] [2] [0]	information [8] [4] [13]	reality [31] [0] [0]
convey (0)	truth [10] [0] [1]	fact [16] [0] [0]	existence [1] [0] [0]	information [112] [19] [26]	reality [7] [0] [0]
grasp (0)	truth [10] [0] [0]	fact [21] [0] [0]		information [1] [0] [2]	reality [11] [0] [0]
demonstrate (0)	truth [7] [0] [0]	fact [47] [5] [0]	existence [26] [4] [1]	information [13] [3] [4]	reality [11] [0] [1]
accept (0)	truth [23] [0] [1]	fact [162] [2] [1]	existence [24] [0] [0]	information [21] [0] [3]	reality [37] [0] [0]

Figure 2. Collocation cluster for *say truth* generated by NetCollo.

NetCollo to assess how well our system detected incorrect English collocations. Overall, NetCollo identified 149 out of the 160 V–N combinations as incorrect usages, for a precision rate of 93.1%. Of the 82 A–N pairs, only two of them were wrongly regarded as correct usages by NetCollo, yielding a precision rate of 97.5%. The precision performances of NetCollo are impressive since they are greater than those reported in several previous studies (e.g., Futagi et al., 2008: 28% and Kochmar & Briscoe, 2014: 75%) and comparable to those generated by Wible et al. (2003: 95.5%) and Chang et al. (2008: 90.7%). Turning to the unidentified errors, we found that NetCollo might not effectively deal with errors involving polysemous nouns. In **gain appreciation*, for instance, its producer used the noun *appreciation* to refer to a feeling of gratitude, in which case the verb should be *earn* rather than *gain*. However, in English, it is fine to say *gain appreciation* for having a better or deeper understanding of something. There are several examples of the latter in the BNC that prevented NetCollo from detecting our target error. Another limitation of NetCollo was that our tool appeared not to identify two non-collocable words that happened to frequently show up together in longer strings. Pairs such as **remain memory* (which should be *retain memory* instead) were judged by NetCollo as correct usages because they were often embedded in longer expressions such as *remain fixed in someone's memory* or *remain vivid in someone's memory*. The two A–N pairs that were not correctly detected, **full enthusiasm* (correct usage: *complete enthusiasm*) and **full fun* (correct usage: *great fun*), were also components of common longer multi-word sequences: *full of enthusiasm* and *full of fun*.

Performance of collocation error correction

As in previous research, we evaluated our system by examining whether it was able to (1) find gold answer alternatives to target wrong collocates and (2) provide more alternatives that were also acceptable and correct. First, considering gold answers, Liu (2002), in addition to listing incorrect V–N collocations, also offered suggestions for each verb error that appropriately corrected the errors in their contexts. For the A–N errors that we examined, in the CLC-FCE dataset each error was also manually coded, with the most appropriate correction being suggested (iLexIR, 2016). These verb and adjective corrections were adopted here as the gold answers. An error in either Liu's (2002) or the CLC-FCE data pool sometimes was tagged with more than a suggestion. We thus input all the suggestions into NetCollo and chose the one which showed the best performance. The collocation error **solve obstacle*, for example, was given two corrections: *overcome* and *clear*. Both verbs were found in the NetCollo cluster generated by **solve obstacle*, with *overcome* ranked the 1st suggestion and *clear* the 4th. We included *overcome* in our analysis only to ensure the best performance of NetCollo. In addition to the gold answers, some other verbs or adjectives suggested by NetCollo were also found to be appropriate alternatives for the wrong verbs/adjectives. Take the incorrect pair **heavy disease* as an example. Although NetCollo failed to find its gold answer *serious*, it nevertheless acquired another adjective, *severe*, that properly corrected this error. Similar examples included *achieve* for **gain aim* (gold answer: *reach*) and *vast* for **large knowledge* (gold answer: *broad*). The two English instructors who helped us examine Liu's (2002) verb–noun list were invited to assess all the corrections automatically produced by NetCollo. We counted only those that both judges agreed to be good corrections as true positives.

The mean reciprocal rank, MRR, implemented by previous studies such as Chang et al. (2008) and Dahlmeier and Ng (2011) as a measure of correction quality, was adopted in our analysis to assess how well our corrections performed. In information retrieval, MRR (Voorhees & Tice, 1999) has been frequently employed to evaluate the quality of correct responses. It is calculated by adding up all the reciprocal ranks at which good responses are found and dividing the sum by the total number of queries. In collocation correction research, the reciprocal rank of a correction is $1/r$ when the rth response is confirmed to be a good alternative collocate. The MRR then is the average ranking score.

We show the performances of NetCollo in correcting collocation errors in Table 2, with the MRR values listed as quality indicators. Overall, our results demonstrate that NetCollo performed rather poorly in finding gold answers but effectively acquired many other correct verb/adjective collocates. Considering the former, the precision rates that our system yielded (V–N corrections: 55.6%; A–N corrections: 45.1%) were better than Dahlmeier and Ng's (2011: 25.2%) but considerably lower than those reported by Chang et al. (2008: 84.4%) and Kochmar and Briscoe (2015: 71.2%). The MRR values that NetCollo achieved, however, were far better than in either of the latter two studies (0.50 and 0.51, respectively). NetCollo's high MRR scores were not surprising to us. As stated earlier, we offer very small clusters on NetCollo lest too much information overwhelms L2 learners. In an 18×5 verb–noun or adjective–noun cluster, in addition to verbs or adjectives frequently collocating with a target noun, users are also presented with several other verbs/adjectives that do not collocate with it. In the NetCollo cluster generated by the error **enlarge knowledge*, for example, whereas users obtain many alternative verbs such as *extend*, *increase*, and *broaden*, they also acquire wrong collocates (e.g., *augment* and *affect*). Suggested verbs/adjectives for errors consequently are mostly high-ranking corrections. Among the manually examined suggestions, NetCollo showed a high precision of

Table 2. NetCollo's performance in correcting wrong verb/adjective collocates.

POS	Type	Precision	MRR
Verb	Gold answers	55.6%	0.705
	All correct alternatives	86.2%	0.734
Adjective	Gold answers	45.1%	0.756
	All correct alternatives	75.6%	0.755

86.2% for verbs and a slightly lower one of 75.6% for adjectives, with the MRR values again producing impressive performances (0.734 and 0.755). Some of the good examples that NetCollo extracted are shown in Table 3. The comparatively lower performance of the adjective suggestions might be attributed to the unavailability of WordNet hypernym/hyponym resources for adjectives. We expect that NetCollo will more effectively correct adjective errors once fuller semantic resources for adjectives are available.

As our results confirm, NetCollo is a useful tool for correcting most English L2 learners' collocation errors. Concerning our slightly lower correction performances than in some previous studies, one important reason might be that NetCollo does not utilize any learners' L1 information as main computation resources. Chang et al.'s (2008) approach, as described earlier, made use of Chinese(L1)-English(L2) translation equivalents and effectively corrected Chinese EFL learners' errors. The tool developed based on their approach (Chang, n.d.), consequently, would be most useful for Chinese learners of English. However, NetCollo relies on the computation and processing of only target language resources and will thus be suitable for all English L2 learners to verify the collocations that they produce. NetCollo and the results that we report here indicate that collocation clusters are an inexpensive technique that can be successfully applied to the detection and correction of L2 learners' content word errors.

Limitations

Although it has been shown that NetCollo can correct collocation errors using limited resources, its effectiveness and value to users nevertheless are restricted by the resources that it depends on. Below we discuss three limitations of NetCollo that must be solved in future improvements.

First, concerning detecting incorrect collocates, while NetCollo successfully identified over 90% of target errors, our approach may run the risk of wrongly tagging acceptable combinations as errors. Error detection studies using comparison approaches, such as Chang et al. (2008), generally compare learners' word pairs to combinations stored in a standard reference corpus, with pairs being tagged as errors if they are

Table 3. Examples of verb/adjective collocates suggested by NetCollo.

Incorrect Collocations	Teachers' Suggestions	NetCollo's Suggestions
**appear colour*	*show*	*display*
**frustrate confidence*	*shatter*	*undermine*
**gain aim*	*reach*	*achieve*
**loosen pressure*	*lessen*	*reduce*
**strong hunger*	*intense*	*desperate*
**nice friendship*	*good*	*close*
**near area*	*nearby*	*neighbouring*

found to appear rarely in the reference dataset. As noted earlier, comparison approaches, if the thresholds that they adopt for collocation validation are set at high levels, can mostly produce high performances. However, they inevitably suffer from the problem that rare but acceptable combinations may be misidentified as collocation errors (e.g., the example collocation *appropriate concern* discussed in Kochmar & Briscoe, 2014). Although the techniques utilized by NetCollo are not traditional comparison approaches (i.e., NetCollo regards a pair as a candidate collocation or not based on whether it shows up in an automatically generated cluster), it relies mainly on a reference corpus and may face similar problems. Standard reference corpora of enormous sizes would be a feasible but not optimal solution to the problem.

Second, NetCollo cannot construct collocation clusters effectively if the target verbs, adjectives, or nouns are scarce in our reference corpus, the BNC. In the tool evaluation presented above, we lowered our collocation frequency threshold from 10 to 5 for low-frequency keywords, and in this way NetCollo did successfully generate clusters for most of them. Some words, however, are extremely scarce in the BNC, so that cluster generation was almost impossible. For example, in the error **present politeness* tested in our evaluation, the target noun *politeness* appeared only four times in the BNC. NetCollo thus could not find verb collocates for *politeness* and the following shared collocates searches and cluster generation could not be completed successfully. A plausible but not ideal solution to this is replacing the target noun with a semantically similar word with higher frequency counts. We tested this idea by replacing **present politeness* with **present courtesy*, whereupon NetCollo did find a good collocation: *show courtesy*. Such good results, however, were not observed when we examined other errors (e.g., the V–N error **lower motivation*).

Finally, like most collocation error detection/correction studies, our system evaluates errors without taking their surrounding contexts into consideration. As Kochmar and Briscoe (2014) criticized, systems that do not consider contexts cannot find that a combination that is unacceptable in most contexts is appropriate and correct in particular contexts. In order not to error-tag these combinations in their correct contexts researchers must annotate the combinations and their specific contexts first. Though labor-intensive, the annotation work helps reduce the risk of wrongly tagging students' correct productions.

Use of NetCollo in English L2 learning

NetCollo, with its capability to represent lexico-grammatical knowledge in collocation clusters, is expected to be a useful tool for both lexicographers and textbook writers. English L2 learners, especially those in EFL environments with little or no access to native speakers who can help verify true/false collocations, can also directly benefit from NetCollo when learning collocations. Like corpus concordancers, NetCollo is best used as a reference tool in EFL writing classes. EFL learners, especially higher-level ones, are suggested to employ NetCollo to improve their use of collocations in writing and learn from the collocational knowledge provided by our tool. To understand whether learners could efficiently locate proper collocations using NetCollo, we conducted a pilot study to address this question. More specifically, in the pilot study we recruited 15 master's students who majored in TEFL (teaching of EFL) in Taiwan and investigated the V–N and A–N pairs that they produced. According to their earlier English proficiency tests, eight of them were rated upper-intermediate (i.e., CEFR B2) and seven were at intermediate (CEFR B1) levels. They were randomly assigned one of two Chinese thesis abstracts to translate into English. Both abstracts were about 500 words in length, and their topics were both focused on TEFL. The subjects were given an hour to complete the translation work and, after a 10-minute break, were led to assess all the V–N and A–N combinations that they used by consulting NetCollo. They were given a short introduction to the main functions of NetCollo and were asked to check the correctness and appropriateness of their collocations. If there were any incorrect combinations identified by NetCollo, the students were advised to replace them with NetCollo's suggested alternatives. The subjects spent an average of 25 minutes doing the second task. The data of two B1-level students, however, were removed from further analysis because they did not complete the translation task. Table 4 provides the total numbers of errors of the subjects before and after they consulted NetCollo, with several errors and corrections shown as examples.

Overall, as Table 4 indicates, the 13 subjects corrected about 88% of the errors that they made in the translation task. This self-correction rate was markedly better than that reported by Chen (2011), where the researcher found that his web-based collocation retrieval tool helped EFL learners self-correct around 50% of their collocation errors. Such a

Table 4. Numbers and examples of errors found in pilot study.

	Numbers	Examples
Made in translation task	33	–
Corrected by subjects	29	**inevitable role* → *crucial role* **apply strategy* → *employ strategy* **fit student* → *suit student* **deep interview* → *in-depth interview*
Not corrected by subjects	4	**conduct questionnaire* **turn out finding*

difference did not surprise us because unlike NetCollo, which aims to generate semantically related alternatives to a wrong collocate, Chen's tool retrieved a keyword's candidate collocates without taking meaning into consideration. Consequently, it would be easier for learners to self-correct collocation errors by consulting NetCollo. Furthermore, the 88% rate was even higher than the error correction percentages that we reported earlier in the tool evaluation sections. This finding thus suggests that NetCollo would perform better in correcting errors in academic texts than in non-academic texts. This possibility, however, needs to be examined further for a better understanding of NetCollo's real effectiveness in constructing academic collocation clusters. The pilot study discussed here nevertheless did provide us with some promising evidence that NetCollo can effectively enable English L2 learners to use collocations more accurately and appropriately in writing.

For EFL learners with lower L2 proficiency levels, we recommend a task that teachers can easily undertake in classrooms to help their students use collocations more correctly. Specifically, teachers are suggested to lead students to notice correct and incorrect collocations shown on NetCollo. For example, if in class the error **say truth* is found, teachers can encourage students to answer the following questions by consulting NetCollo:

1. Is **say truth* a correct collocation according to NetCollo?
2. What are the verbs that share meanings similar to *say* and collocate well with *truth*?
3. In the cluster generated by **say truth*, do you get any other similar combinations that are also incorrect in English?
4. In the cluster generated by **say truth*, besides *state truth* and *tell truth*, do you get any other similar combinations that are also correct in English?

Students who show a high interest in analyzing lexical patterns should be led to explore the example sentences containing **say truth* in the BNC. In this way, they can further discover the useful pattern *say with* *truth*. Answering these questions will be more difficult if learners consult collocation dictionaries or corpus concordancers. Using dictionaries such as the *Macmillan Collocation Dictionary* (Rundell & Fox, 2010), for example, learners will discover that it is fine to say *speak truth*, *tell truth*, or *reveal truth* in English (p. 852). However, EFL students still cannot be sure whether they can use **say truth* in their writings since the fact that information is not listed in dictionaries does not show or prove its incorrectness.

Corpus concordancing tools also have shortcomings. Searching for **say truth* in English L1 corpora such as COCA (Davies, 2008) can prevent some learners from using this wrong expression. Correct alternative verb usages, however, will not be made available to learners until they find semantically similar words to *say* by themselves. Concordancing tools that are made less complicated and designed especially for language learners provide more learner-friendly information. SKELL (Sketch Engine for Language Learning, n.d.), for instance, enables English learners to check word usages and look up collocations in a 'non-scary version of Sketch Engine' (Kilgarriff, Marcowitz, Smith, & Thomas, 2015, p. 66). Specifically, on SKELL, learners can get example sentences for a keyword, patterns such as adjectives or modifiers for the keyword, and semantically similar words. Learners can also key in a word combination to evaluate whether it is a true collocation in English. Compared with traditional concordancing tools, SKELL thus helps English learners more efficiently verify incorrect collocations and obtain correct verb alternatives. However, to collect all the knowledge that NetCollo shows, including verb alternatives (e.g., *tell* and *state* for **say truth*), noun alternatives (e.g., *fact* for **say truth*), other semantically relevant correct usages (e.g., *state fact*), and other semantically relevant incorrect combinations (e.g., **say fact*), learners have to make many searches on SKELL which takes considerable time. NetCollo, furthermore, allows users to search for V–N or A–N examples that co-occur in five-word windows, whereas SKELL offers searches of only n-grams. The limitations of dictionaries and concordancing tools are part of the reasons that we developed NetCollo. Using NetCollo, learners can gain the collocation information and knowledge that they need quickly, and further explore and discover the other relevant combinations that they had better avoid and other acceptable pairs that they can consider using in writing to add variety.[9]

In short, as a collocation exploration tool, it is suggested that EFL learners use NetCollo as a reference tool when writing English articles. Learners are also urged to explore the example sentences that NetCollo offers by clicking on icons shown below each word combination to learn how to use a collocation correctly in real contexts. ESP learners, specifically those in the computer science or engineering fields, are encouraged to use NetCollo to compare collocations in general use (i.e., the BNC) and in their field of study. We provide domain-specific collocational knowledge for ESP learners because, as Hyland and Tse (2007, pp. 246–247) indicate, 'stable word combinations are an important part of a discipline's discoursal resources' that ESP learners must notice and learn.

Conclusions

In this article, we report an online English collocation exploration tool capable of detecting and correcting collocation errors made by EFL learners. The approach that our tool adopts to address learners' collocation problems is innovative; NetCollo automatically constructs collocation clusters based on the lexico-grammatical information extracted from a standard reference corpus, with the clusters clearly showing which combinations are possible in English and which are not. To establish clusters, NetCollo relies only on the processing and formation of lexico-grammatical patterns in a single corpus. In this regard, NetCollo is extremely inexpensive in terms of the resources that it requires. We deem it necessary to take amounts of computational resources into consideration when evaluating error detection/correction systems because teacher annotations or learners' L1 resources, which many previous studies employed to detect/correct errors, take considerable time and effort to collect. In this article, we point out the importance of designing digital reference tools based on restricted resources, and accordingly use inexpensive techniques to develop a tool that is readily available to both English L2 instructors and learners.

Perhaps more importantly, we offer a lengthy discussion of Cowie and Howarth's (1996) notions of intercollocability and collocation clusters to highlight their significance in the research field of L2 collocation learning. Not only does Cowie and Howarth's (1996) work provide a viable and useful foundation for devising digital reference tools, as we have demonstrated in this article, but it presents a systematic and adequate account of ESL/EFL learners' difficulties in grasping English collocations. Numerous researchers in SLA since the early 1990s have noticed such difficulties and applied a variety of approaches to better understand the causes of such difficulties. Common methodologies included elicitation tasks, reception tests, and corpus data analysis. The research techniques employed varied and their findings differed. Elicitation studies such as translation or fill-in tasks (e.g., Bahns & Eldaw, 1993), for example, pointed out that L2 learners' productive knowledge of collocations was considerably below that of single words. Specifically, Bahns and Eldaw's (1993) results showed that L2 learners made many more collocation errors in translations than with single words. Reception tasks, like recognizing errors or differentiating correct from incorrect combinations (e.g., Gyllstad, 2005), revealed strong correlations between learners' collocation and single-word knowledge.

These production- and reception-based studies taken together suggest that learners' real problem lies not in identifying correct collocations, but in producing and using them appropriately. Learner corpus studies furthermore indicate that L2 learners frequently use collocations as native speakers do. What indeed troubles L2 learners is they are over-dependent on familiar ones (Hasselgren, 1994), and their errors were negatively influenced by their first language (Nesselhauf, 2003). Whereas all previous studies present pieces of a general picture of L2 collocation knowledge and use, we nevertheless need systematic accounts like Cowie and Howarth (1996) that adequately explain why L2 learners cannot use collocations entirely correctly and appropriately. Cowie and Howarth's (1996) work is not only applicable in practice to error analysis and digital tool design, but is theoretically useful for understanding L2 learners' learning problems.

In our future improvements of NetCollo, we plan to investigate and adopt much larger textual datasets, such as ENCOW (Schäfer, 2015) or ukWaC (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009). It is expected that our precision in finding gold answers will significantly improve and the clusters that NetCollo automatically generates will be even more informative if NetCollo relies on a giga-token corpus when extracting collocations and forming clusters.

Notes

1. The mean reciprocal rank is a measure frequently adopted in information retrieval to evaluate the quality of collected responses. We discuss this measure in more detail later when we report the correction performances of our own tool.
2. Wordnet is an enormous English lexical database. In the database, words are organized into sets of synonyms, with each describing an individual concept. These sets, which are called synsets, are further structured and inter-connected by specifications of a variety of lexical relations, including hypernymy, hyponymy, meronymy, holonymy, troponymy, entailment, and coordinate terms. Wordnet can be effectively used as a dictionary or thesaurus. With its abundant knowledge and structures, it has been considered a valuable E-tool for computational linguistics and NLP. Wordnet is accessible via its web-browser: http://wordnetweb.princeton.edu/perl/webwn (Princeton University, 2010).
3. In fact, Dahlmeier and Ng (2011) did consult two judges who manually evaluated the candidate collocates found by their system. However, the judges were only

required to check the top three suggestions. The researchers found that the features mixed together could achieve precisions of 38.2, 32.87, and 29.3% at ranks 1–3, respectively.

4. NetCollo is available at: http://www.netcollo.info/
5. It should be noted that Wordnet provides different lexical relations for different parts of speech. Synsets of adjectives, for example, are not inter-connected by hypernym and hyponym relations. On NetCollo, we thus calculate and decide the semantic similarities of adjectives based solely on the synonym information of WordNet. The unavailability of the relation information, however, somewhat restricts the construction and usefulness of NetCollo clusters involving adjectives. We discuss this further in the next section when we present and evaluate NetCollo's performance in correcting A-N errors.
6. The 5 (frequency) × 4.0 (MI) combination was used on 34 of the overall 242 tested items. In each of the 34 combinations, one or both components were low-frequency words (e.g., the word *vocabulary* in the wrong collocation **catch vocabulary*). As noted, on NetCollo, a cluster is generated and developed by shared collocates. Accordingly, we had to lower the frequency threshold for the low-frequency words to obtain at least some collocates first, after which our tool could find other words that share collocates with them. In this study, if one or two components of a tested combination held a frequency number lower than 2,000 in the BNC, we utilized the low-frequency measure. On the NetCollo interface we also suggest that users lower the frequency numbers to 5 or 3 if they would like to input low-frequency keywords.
7. The XML edition of the BNC comprises around 100 million running tokens.
8. In our evaluation, reliable clusters refer to those clusters capable of suggesting correct alternatives to collocation errors. To examine which MI and frequency combination produced the best performances, we tested MI scores ranging from 1.0 to 7.5 and frequency thresholds ranging from 5 to 60 and found that the 4.0 (MI) × 10 (frequency) combination collected the most gold answers to our V–N and A–N errors. We present the evaluation results in the Performance of collocation error correction section.
9. For advanced learners of English, we suggest that they choose semantically similar words by themselves to construct more useful clusters. When users search the wrong collocation **make income* on NetCollo, for example, they will see that similar nouns for *income* include *profit*, *earnings*, *dividend*, and *amount*, and NetCollo does not automatically provide appropriate corrections for *make*. Users, however, can use the Reselect function on NetCollo to choose intuitively similar words *earnings* and *wage* for *income*, and in this way they can get a good verb alternative: *earn*.

Notes on contributors

Ping-Yu Huang is an assistant professor at the General Education Center, the Ming Chi University of Technology. He focuses his research works on corpus linguistics, digital language learning, and second language acquisition.

Nai-Lung Tsao is an assistant researcher at the Office of Information Services of Tamkang University. His main research interests include information retrieval and natural language processing.

ORCID

Ping-Yu Huang http://orcid.org/0000-0001-9538-5625
Nai-Lung Tsao http://orcid.org/0000-0003-1199-6803

References

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL)–A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235–247. doi:10.1016/j.jeap.2013.08.002

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations?. *System*, *21*(1), 101–114. doi:10.1016/0346-251X(93)90010-E

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–226. doi:10.1007/s10579-009-9081-4

Benson, M., Benson, E., & Ilson, R. F. (1986). *Lexicographic description of English*. Amsterdam: John Benjamins Publishing.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572. doi:10.1111/j.1467-9922.2010.00566.x

British National Corpus, version 3 (BNC XML Edition) (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning*, *18*(3), 231–251. doi:10.1080/09588220500185769

Chang, Y. C. (n.d.). Collocation Checker. Retrieved April 14, 2016 from http://miscollocation.blogspot.com/.

Chang, Y. C., Chang, J. S., Chen, H. J., & Liou, H. C. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning*, *21*(3), 283–299. doi:10.1080/09588220802090337

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. doi:10.1080/09588221.2010.526945

Chen, H. J. H., Wu, J. C., Yang, C. T. Y., & Pan, I. (2016). Developing and evaluating a Chinese collocation retrieval tool for CFL students and teachers. *Computer Assisted Language Learning*, *29*(1), 21–39. doi:10.1080/09588221.2014.889711

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, *12*(4), 345–360. doi:10.1076/call.12.4.345.5699

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, *2*(3), 223–235. doi:10.1093/applin/2.3.223

Cowie, A. P., & Howarth, P. (1996). Phraseological competence and written proficiency. *British Studies in Applied Linguistics*, 11, 80–93.

Dahlmeier, D., & Ng, H. T. (2011). Correcting semantic collocation errors with L1-induced paraphrases. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 107–117).

Davies, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990–present. Retrieved March 17, 2017 from http://corpus.byu.edu/coca/.

Futagi, Y., Deane, P., Chodorow, M., & Tetreault, J. (2008). A computational approach to detecting collocation errors in the writing of non-native speakers of English. *Computer Assisted Language Learning*, *21*(4), 353–367. doi:10.1080/09588220802343561

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*, (pp. 145–160). New York: Oxford University Press.

Gyllstad, H. (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. *The Department of English in Lund: Working Papers in Linguistics*, *5*, 1–31.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, *4*(2), 237–258. doi:10.1111/j.1473-4192.1994.tb00065.x

Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, *41*(2), 235–253. doi:10.1002/j.1545-7249.2007.tb00058.x

iLexIR. (2016). Error annotation in adjective–noun (AN) combinations. Retrieved June 28, 2018 from https://ilexir.co.uk/datasets/index.html.

Jian, J. Y., Chang, Y. C., & Chang, J. S. (2004). Collocational translation memory extraction based on statistical and linguistic information. In Proceedings of the 16th Conference on Computational Linguistics and Speech Processing (pp. 257–264).

Kilgarriff, A., Marcowitz, F., Smith, S., & Thomas, J. (2015). Corpora and language learning with the Sketch Engine and SKELL. *Revue Française de Linguistique Appliquée*, *20*(1), 61–80.

Kochmar, E., & Briscoe, T. (2014). Detecting learner errors in the choice of content words using compositional distributional semantics. In Proceedings of COLING 2014 (pp. 1740–1751).

Kochmar, E., & Briscoe, T. (2015). Using learner data to improve error correction in adjective–noun combinations. In Proceedings of the 10th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 233–242).

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, *24*(1), 29–49. doi:10.1093/ijl/ecq039

Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching & Learning*, *27*(3), 35–56.

Liu, L. E. (2002). A corpus-based lexical semantic investigation of verb-noun miscollocations in Taiwan learners’ English. Master’s thesis, Tamkang University, Taipei.

Liu, A. L. E., Wible, D., & Tsao, N. L. (2009). Automated suggestions for miscollocations. In Proceedings of the 4th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 47–50).

Mel’čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.) *Phraseology: Theory, analysis, and applications* (pp. 23–53). New York: Oxford University Press.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: CUP.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223–242. doi:10.1093/applin/24.2.223

Nicholls, D. (2003). The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT. In *Proceedings of the corpus linguistics 2003 conference* (Vol. 16, pp. 572–581).

Princeton University (2010). *About WordNet. WordNet*. Princeton, NJ: Princeton University.

Rundell, M., & Fox, G. (2010). *Macmillan collocations dictionary*. Oxford: Macmillan.

Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In Proceedings of Challenges in the Management of Large Corpora (CMLC-3) (IDS Publication Server) (pp. 28–34).

Seargeant, P. (2016). World Englishes and English as a lingua franca. In G. Hall (Ed.), The Routledge *handbook* of English *language teaching* (pp. 13–25). New York: Routledge.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sketch Engine for Language Learning. (n.d.). Retrieved September 21, 2018 from https://skell.sketchengine.co.uk/run.cgi/skell.

Voorhees, E. M., & Tice, D. M. (1999). The TREC-8 question answering track evaluation. In Proceedings of the 8th Text REtrieval Conference (pp. 83–105).

Wible, D., Kuo, C. H., Tsao, N. L., Liu, A., & Lin, H. L. (2003). Bootstrapping in a language learning environment. *Journal of Computer Assisted Learning*, *19*(1), 90–102. doi:10.1046/j.0266-4909.2002.00009.x

Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011). A new dataset and method for automatically grading ESOL texts. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (vol. 1, pp. 180–189).

Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, *20*(2), 131–152. doi: 10.1080/09588220701331451