Student self-editing of written work using online concordance

Stephen Jennings

Introduction

Self-editing of student writing may be thought of as one of the goals of advanced writers in a foreign language\(^1\). However, a question that needs to be posed here is “how are teachers to help students achieve a better understanding of the gap between what they want to write in English and what they actually write?” Previous studies have shown that learners expect accurate correction of writing assignments by teachers. Nevertheless, the best way to go about correcting student errors in writing (see also Desrosiers and Fujieda in this issue) is still a moot point. It is true to say, though, the majority of scholarly papers have shown the need for effective responses to student errors in their writing (see Fujieda 2006 for more information).

This essay will show how an online concordance program (The Collins Cobuild Corpus\(^2\) Concordance Sampler) combined with some amount of teacher intervention, can help students become able to self-edit their written work. Further evidence will demonstrate that this set of materials is useful in pre-empting miscommunication and is, as such, a step toward enabling lesson content to be of use outside the classroom and after graduation.

The Collins Cobuild Corpus Concordance Sampler\(^3\) is a very useful tool. Though useful in allowing higher level learners to search for appropriate word partners, it is of most practical use, I believe, in allowing students to edit their own mistakes. The appropriate teacher feedback allows students to be a “detective”, searching for the most likely solution to the “crime” that was committed. The “clues” to the solution need to be pointed out, and a probable answer deduced. This deduction contributes to awareness-raising of general language patterns and a greater ability to put the more appropriate language form into long term memory.

In sum, this essay, it is hoped, is a readable and engaging piece of work, it will encourage readers to think and ask themselves questions as they “work through” the pages. As a first step toward understanding the process by which students are shown how they can self-edit their written work; it is important to accurately define some of the important background knowledge of the general topic. The three main notions are Collocation, Online Corpus and Editing written work by using online concordance.
1. Collocation

Collocation is an important word for this essay. It is defined by Halliday, Teubert, Yallop and Cermakova (2004) as ‘the habitual meaningful co-occurrence of two or more words in close proximity to each other; as a lexical relationship’. In other words, as Lennon and McCartney of the Beatles wrote in their song *Michelle* “… words that go together well”. In addition to words going together well, it will be shown below that some words go together better than others.

If the premise is taken that some words go together better than others, look at the common Japanese and English proverbs in the box below and try to finish the sentences:

<table>
<thead>
<tr>
<th>Japanese proverb</th>
<th>English proverb</th>
</tr>
</thead>
<tbody>
<tr>
<td>猿も木…</td>
<td>It is better to be safe…</td>
</tr>
</tbody>
</table>

English, like all languages, is built on phrases, often called “chunks of language”. These chunks are made up of partnerships of words. They are “word partners” because they are used together more often than just by chance.

- The Japanese proverb has the ending 「から落ちる」
- The English proverb has the ending “than sorry”

How do we know these answers? It is probably because we learned them as children in our mother tongues, or it could be that they were learned in a foreign language as a youth or an adult.

However, if you were unable to finish the sentence, why were you unable? Let us look at the English proverb “It is better to be safe …”. A native English speaker reader would be able to correctly finish the sentence with the accepted phrase without problem. Some readers, though, may be able to finish off the sentence with *a grammatically correct phrase* but not know *the accepted phrase*. This inability is the heart of the matter to which this essay pertains.

As a way to summarise the above; do the activity below:
Activity 1
Which words may go in the blanks?

- I watched a great ________ _______ last night

Which word may go in the blank?

- I watched a great horror ________ last night

Explanation of Activity 1
With the phrase “I watched a great ________ _______ last night” there are many words that can be put into the blanks. For example “TV program”, “baseball game” … etc. However, with the phrase “I watched a great horror ________ last night” the words “movie” or “film” are by far the most likely to be put in the gap. The reason for this is that “watch” and “great” are relatively weakly linked together. On the other hand, “watch”, “great” and “horror” used together as a part of an utterance are strongly linked to the words “movie”, “film”.

Set phrases
Proverbs are a type of set phrase, but it is not only proverbs that can be agreed patterns of words. Set phrases are groups of words that become habitually grouped together. According to Widdowson (1996), these types of phrases are compounds of words; not combinations of words, i.e. they cannot be grammaticalised; as a result the whole pattern, or chunk of language, must be noticed if it is to be learned.

Often these chunks of language show markers in an utterance, for example, the phrase “As I was saying” shows that the speaker will continue with a speech act that had been interrupted. This is what is meant by a compound of words. Activity 2, below, attempts to show the reader how to become more aware of the set phrases or chunks of language they use.
Activity 2

1. Write a word or phrases that are commonly said before the common verb “drive”; then draw an arrow to the box.

2. Draw other arrows leaving the box, for words or phrases that are commonly said after it.

Two examples have been done for you. Try to get around five words or phrases.

1. 2. 3. 4. 5.

Now try to rank the words and phrases from strongest connection to drive (1) to weaker (5).

1. 2. 3. 4. 5.

Concluding remarks about collocation

There are three important pieces of information to keep from the above discussion and activities concerning Collocation; 1. Collocations are set phrases of
which proverbs are a good, but not the only, illustration; 2. Set phrases are compounds of words that are very useful in marking the intension of an utterance. 3. Language users know their first language well and their second language less well.

To illustrate, mother tongue speakers of any language can make chunks of language, using nouns and verbs, with ease but it is more difficult to do in a second language as all language learners are aware. Collocation is the term used to show the links between words in a sentence and is, therefore, one of the notions that learners need to be made aware of because they will then be better able to rationalise sentence structure.

It is also important to acknowledge that collocations (multi-word chunks) can be ranked to show that certain words have stronger links between them than others. These words with stronger links occur more often together, these are words that go together well; but are not known by language learners, this notion is most widely known as Interlanguage. The term Interlanguage was coined by Selinker (1972); Interlanguage is defined as “the interim grammars constructed by the second-language learners on their way to the target language”. This notion has also been known as approximative system (Nemser, 1971); idiosyncratic or transitional dialect (Corder, 1971); and Developing System (van Patten, 1996). The notion of second language learners having difficulty in assuming native-like fluency in another language is a part of the notion of Communicative Competence (cf. Hymes, 1966, Canale and Swain, 1980; Bachman, 1990).

The ability to use words that are expected to go together (that is set phrases, or patternings of nouns, verbs and prepositions) is the ability to not only have grammatically sound sentences but also to smoothly impart the information sought. When native speakers of any language do this, the communication is often effortless and does not cause misunderstanding. However, if one party is a learner of the language being used to communicate, we have seen that the imperfect knowledge of accepted chunks of language holds back timely, flawless communication because the language used by the learner is not what is expected by the native-speaker.

2. Online corpus

Background

Online, here refers to the Internet; an on-line corpus is an internet-based database containing millions of words uttered in context. Annotated corpora are a
valuable source of information because they can be searched by learners interested in finding out instances of how native-speakers of a particular language speak or write.

There has been much written about student errors in writing; teacher feedback to students; students’ perceptions of this process; and teachers’ feelings about the process of marking students’ work.

The thesis of this essay is that, in the field of English Writing, learners or near-native speakers will have the chance to;

- Fix their own errors
- Respond in an appropriate way to errors
- Become more independent
- Become aware of language in such a way as to become careful of error making

How the Collins Cobuild Corpus Concordance Sampler website can be used so as to enable learners to successfully edit their own work is discussed below.

**Is an Online Corpus really necessary to improve student editing?**

Many readers may be asking “Why use the Internet when there are very good dictionaries?” The answer may be found in a previous work cited in Jennings (2007); i.e. one of the main reasons for using corpora is that bilingual dictionaries do not record … complex units of meaning in a methodical way … that is why they do not help us to translate into a language we are not familiar with

(Teubert, and Cermakova in Halliday et al. p.123)

**Explanation**

For example, if the word “ubiquitous” is to be used in the sentence but the user is not quite sure how; concordances of collocations are very useful tool. This is because they give a list of sentences already used by native speakers (see the Data section) – more than you would find in a dictionary – guiding you to which pattern of words sound more natural with “ubiquitous”8. Another example is the word “Search”. When looking for an appropriate negative adverb, a quick search of the corpus will show a likely word-partner.
Activity 3: Which adjective sounds better with the word “search”?

- Fruitless search
- Bottomless search

Are you able to choose one? Why do you choose that one? 9

Activity 4
Many teachers give writing homework:

- What are the main errors students make in writing?

<table>
<thead>
<tr>
<th>Errors students make</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low level</td>
</tr>
</tbody>
</table>

- How is feedback given in your experience?

- What do students do after the feedback has been given?

Limitations to corpus informed feedback on student writing
At this point, a caveat to the collocation sampler’s efficacy must be explored, as some errors are simply not covered by collocation research. Below are two samples of student writing, where errors are too numerous, or the need for face to face intervention is too
great for effective use of the online concordance sampler. In such cases, simply using a corpus to find lists of collocations will not help the student enough.

**Problem one**
Look at this introduction to an essay:

I would like to write about living by myself. Living by myself has a lot of things to learn. For example, cooking, doing the laundry, spending time and money and so on.

**Problem two:**

In fact, in Japan, there are a lot of the big music festivals that are held outdoor, example in Sapporo, Ibaraki and more. People can feel nature.

(A discussion of these problems can be found in Appendix C)

There remains, though, the problem of finding the correct data from the results shown in the corpus. This problem is dealt with in some detail in the next section.

3. **Editing written work by using online concordance**

Samples of concordance are not only useful for wrong prepositions and inaccurate single vocabulary items but are also good for certain grammatically correct, but a phrasing which makes the reader unsure as to the meaning intended.

**Example of error**

In one example of writing, a student wrote:

Big cities play an important part of politics and economy.

The phrase ‘play an important part’ is used slightly incorrectly.
What should a teacher do?

As a teacher giving feedback to this student, the best course of action would be to look at the sentence and mark the “of” in the phrase “Big cities play an important part of politics and economy.” indirectly, with an underline. Bologna, Jennings and McCrostie (2006) write that learners have a variety of preferences for error correction methods but higher levels tend towards a mixed style of explicit and implicit (indirect) error marking.

The student would then use the concordance website to search similar phrases to check how English speakers usually use ‘play an important part’. Subsequent to this search, the student would then compare the list of phrases he found searching the concordance sampler, with the phrase that was marked with an underline by the teacher. Students are then often able to deduce an answer, or ask for further help. The asking for further help, I would argue, is mixture of implicit and explicit marking.

This process of learners trying to figure out their own mistakes helps them to become more aware of the process of learning i.e. it facilitates restructuring of language and further remembrance of grammar or word patterning. Students also perceive a sense that this type of language awareness activity can help pre-empt further error (Falkus and Jennings, 2004).

Survey results

In a survey carried out on the five participating fourth year students of a class in Advanced Writing at Kyoai Maebashi College, whom had undergone training in using the Collins Cobuild Concordance Sampler, four out of five answered “yes” to the question: “Is the teacher underlining one word enough to show your error enough to help you?”. An illustrative written answer follows; “When the answer is written it is hard to remember. Finding the answer yourself is kind of めんどくさい [bothering] but it helps me learn.” One learner answered; “sometimes”, this learner explained that “sometimes I can’t understand why it’s [sic] error”.

Another question which provides insight to the noticing of error is; answers to the question “Did using the Collins Cobuild Concordance Sampler in class help you become more aware of your errors?” The answers were three for “yes”; two for “sometimes” and one for “no”. In discussion, I think that if four learners were helped to see their errors in writing more clearly, the object of awareness raising of error through on-line concordance can, anecdotally at least be deemed to have been a success.
In-depth research into the error sentence “Play an important part of politics and economy”

A search of the concordance sampler can be seen in figure 1 (see Data section). If we only enter ‘important part’ into the search function this is what we will find: If you only look at the preposition which follows ‘important part’, you would think that either ‘in’ or ‘of’ would be okay. A better solution is to experiment with possible combinations of words with ‘play’ and ‘part’.

There is a more advanced way of inputting data but it does make the searching more complicated. The concordancer can recognize different forms of verbs and strings of words can be put together using the ‘@’ ‘+’ and ‘2’ symbols.

Example
- ‘@’ after ‘play’ searches will include plays, played and playing.
- ‘+’ means we want to find phrases with play and part.
- ‘2’ finds phrases with play and part with one or two words between

(For a full list of search strings see Appendix A)
(See figure 2 in the data section for the string ‘play@+2part’)

How would you correct this sentence?
“Big cities play an important part of politics and economy.”

For the next example look at this example of student writing

Most of students want to go abroad.

The problem is the preposition ‘of’. We know the solution to this problem but the student obviously doesn’t. How can we inform the student in a pedagogically appropriate way, which prepositions (or none) commonly go with “most” and “students”.

This search string would be a good way to analyse similar sentences:

‘most+1students’
This search looks for strings of words with “most” + “any one word between “most and “students” (See figure 3 in the Data section for results). The difference between the Japanese students’ phrase and that of the native speakers’ can clearly be seen.

**Rewrite this sentence**

“Most of students want to go abroad”

Below, the final activity, Activity 5, allows the reader to search the concordancer with possible strings of words. Numbers 1 - 8 are useful as examples of how online concordance may be used as a quick aide memoir when planning sentence structure for non- or near-native speakers. Numbers 4 – 8 deal with non-standard forms of writing by intermediate level students. The reader is asked to try to enter possible search strings and deduce a solution and correct the non-standard form.

**Activity 5**

Look at the following problems. What search could you enter to find an answer to these questions? (See Appendix D for answers)

1. What prepositions often follows the verb influence?

2. Which two prepositions commonly follow the question pattern “What do you think~?”

3. What article commonly occurs before the word ‘bride’?

4. Which verbs can go with the noun ‘karate’?

**Now, let’s look at some errors from student writing**
Write down the search you would perform on this to find a more standard form

5. In business situations, employees seek to learn a foreign language because some companies have moved to abroad or some foreign-affiliated companies have moved to Japan, too.

6. During my stay, I could not understand what my host family said and I could not tell my feelings and opinions very well, then I totally lost confidence.

7. Finally, I want to tell you what you should wear when you go to the live music performances. Almost of the fans’ fashion is very casual.

8. I think that is a useful communication tool that you can communicate with foreign people and you can know the different culture each other.

Conclusions

To paraphrase the introduction to this essay, using an online corpus; is a very good way to help intermediate to advanced learners to notice their own mistakes. After students have been shown the place in the sentence the teacher thinks a miscommunication occurs, using an online concordancer, students analyse the difference between what they have written and how native speakers form similar utterances.

It is the analysis and subsequent restructuring by the learners of their mental lexicon, which is the key to the usefulness of the Collins Cobuild Online Concordance Sampler. The learners are not being taught grammar per se, as they have already reached a point at which they can write reasonably freely but they are at a stage of interlanguage building where miscommunication is reasonably frequent. The miscommunication becomes the start of a puzzle to be solved.

The teaching involved in the analysis of concordance is totally student centred and tailored to each individual learner because the learner himself analyses his own use of English. Rather than having been corrected at the earliest stage by the teacher, the
teacher points out an area obliquely and the learner uses this clue to search for the solution to a more native-like English phrasing.

Results indicate that students perceive that this approach has helped them learn patterns of English more thoroughly than before and that this use of an online concordancer is a useful tool which encourages them to think more carefully as they write and lower the amount of simple mistakes they make.

Notes
(1) In this case, English
(2) Corpus is defined here as “a collection of texts or parts of texts upon which some general linguistic analysis can be conducted” (Meyer, 2002)
(3) See this URL: http://www.collins.co.uk/corpus/CorpusSearch.aspx
(4) Or, indeed, to a lesser degree, “flick”
(5) See Appendix B for an example of 5 words and phrases (with the central word “drive”) listed in order of their collocation strength
(6) Communicative Competence is based on research into illocutionary force of utterances (Searle, 1969), which is part of the notion of pragmatic intent. A learner with an as yet imperfect grasp of an L2 will be using an imperfect form of English. As a result, the pragmatic intent of a learner may be misunderstood by his interlocutor. If the learner is unable to communicate competently he does not have sufficient control over the way in which what he wants utter a speech act (Bialystok, 1993). This often leads to inefficacious exchanges between learner and native speaker, especially at the higher levels of competency (Takahashi and Beebe, 1987).
(7) Annotated corpora are lists of utterances in context that have been tagged or marked to show different parts of speech. Learners of foreign languages are then able to search these marks for instances of particular parts of speech. The Collins Cobuild Concordance Sampler makes use of annotated lists of utterances by native speakers of English.
(8) According to the Collins Corpus “ubiquitous” is most common in the form: “… the ubiquitous + NOUN …”
(9) The answer is “fruitless search”. “Bottomless” is a strong collocation of the word “pit”
References


Cambridge University Press.

**Data**

**Concordance results for ‘important part’**

![Figure 1](image1.png)

**Concordance results for ‘play@+2part’**

![Figure 2](image2.png)
Concordance results for ‘most+1students’

<table>
<thead>
<tr>
<th>Concordance details</th>
<th>Example text</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is possible for an exceptional student to climb to the next level after just two weeks, <strong>most students</strong> take at least four to six weeks.</td>
<td>[p] Learning a language is very much a partnership. While Cultural Survey [/h] By the beginning of the second session, <strong>most students</strong> will have reached at least an 'intermediate' level of language proficiency. At this point the <strong>In most cases, students</strong> are required to meet the majority of the cost. A common feature of student exchange, and a <strong>Most students</strong> consider taking time out only before or after a course. But here are some of the more common <strong>Although most students</strong> entering the General Engineering degree programme opt to follow one of the specific <strong>Most students</strong> choose a selection from the following lists. [p] [p] [h] FIRST YEAR [/h] 1 Mathematics students construct their own programme of studies, based on their tastes and career plans. <strong>Most students</strong> select topics from two or more of the main MORSE areas, but it is possible to specialise <strong>Experience</strong> <strong>most postgraduate students</strong> prefer self-catering accommodation where they are not</td>
</tr>
</tbody>
</table>

**Figure 3**
Appendix A

Using the Concordance Sampler

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN</td>
<td>a macro tag: stands for any noun tag</td>
</tr>
<tr>
<td>VERB</td>
<td>a macro tag: stands for any verb tag</td>
</tr>
<tr>
<td>NN</td>
<td>common noun</td>
</tr>
<tr>
<td>NNS</td>
<td>noun plural</td>
</tr>
<tr>
<td>JJ</td>
<td>adjective</td>
</tr>
<tr>
<td>DT</td>
<td>definite and indefinite article</td>
</tr>
<tr>
<td>IN</td>
<td>preposition</td>
</tr>
<tr>
<td>RB</td>
<td>adverb</td>
</tr>
<tr>
<td>VB</td>
<td>base-form verb</td>
</tr>
<tr>
<td>VBN</td>
<td>past participle verb</td>
</tr>
<tr>
<td>VBG</td>
<td>-ing form verb</td>
</tr>
<tr>
<td>VBD</td>
<td>past tense verb</td>
</tr>
<tr>
<td>CC</td>
<td>coordinating conjunction (e.g. &quot;and&quot; or &quot;but&quot;)</td>
</tr>
<tr>
<td>CS</td>
<td>subordinating conjunction (e.g. &quot;while&quot;, &quot;because&quot;)</td>
</tr>
<tr>
<td>PPS</td>
<td>personal pronoun subject case (e.g. &quot;she&quot;, &quot;I&quot;)</td>
</tr>
<tr>
<td>PPO</td>
<td>personal pronoun object case (e.g. &quot;her&quot;, &quot;me&quot;)</td>
</tr>
<tr>
<td>PPP</td>
<td>possessive pronoun (e.g. &quot;hers&quot;, &quot;mine&quot;)</td>
</tr>
<tr>
<td>DTG</td>
<td>determiner-pronoun (&quot;many&quot;, &quot;all&quot;, &quot;both&quot;, &quot;some&quot; etc.)</td>
</tr>
</tbody>
</table>

Use the ‘+’ to connect words and the symbols above.

If you want to include word(s) between any other words, you can use a number (maximum of 4).

Appendix B

A list of examples of 5 words and phrases (with the central word “drive”) listed in order of their collocation strength (from strongest connection to “drive” 1, to weaker, 5 is as follows:

1. Four wheel –
2. Drink and –
3. – thru  
4. Hard -  
5. Take a –

The main point here is that without the missing word a person with a good or very good command of English is able to guess the word “drive” with only the input in the list 1-5. Taking a closer look at number 5, we can see this is a week collocation because the head part of the phrase “Take a” by itself does not necessarily cause the reader to think of the word “drive”. There are a multitude of possible words that could be placed after this word.

Appendix C

Problem One
It is not “living by myself” that learns; it should be made clear that “a person who lives by themselves learns a lot” through living alone. Also “and so on” has the connotation that there are more things the writer wants to say but will not. This is probably a mistranslation of the Japanese など which has the connotation of inclusion. It would be better for the writer to delete “and so on”.

Problem Two
Here “outdoor” should be “outdoors”. Searching the concordancer for a word that doesn’t exist may be of use; but may also cause confusion. Another example of error is the negative transfer of the idiom “feel nature”. English does not have this idiom, the student would have to be shown a more standard way to write how he feels, for example “get in touch with nature”. This could be searched for in a corpus but a greater amount of teacher intervention would be needed here.

Appendix D

1. influence+IN 5. move@+1abroad  
2. what+do+you+think+IN 6. VERB+1my+feelings  
3. DT+1bride 7. almost+2fans  
4. VERB+karate 8. can+1VERB+the+different
要旨
オンライン・コンコーダンスを用いた英語ライティングの自己添削
ジェニングズ スティーブン

本研究は、EFL環境においてオンライン・コンコーダンス（ウェブサイト: Collins Cobuild Corpus Concordance Sampler）を用いた学習者による自己添削についてのアクション・リサーチである。Collins Cobuild Corpus Concordance Samplerは、英語知識が中級から上級の学習者にとり、とくに良いツールである。教師は、フィードバックの段階において、学習者自身の作文中の誤りを間接的に指摘する。

このフィードバック行為により、学習者は教師の示す手がかりに気付き、自らがおかしな人たちの型を知るための演繹的技術を使用できるようになる。そして、学習者は、コンコーダンスのオンライン・リストを用いて、不正確な文章の一節を、ネイティブ・スピーカーの発する英語の型へと自ら添削するのである。教師によって与えられた手がかりから、誤りを正すという作業は、学習者が自らの心的辞書を再構築する助けとなる。初期的な本研究では、学習者の、自らがどのようなことを書いているのかということに対する意識が高まり、こうした意識の高まりにより、その後の誤りがなくなるということが明らかになった。