Using Educational Technology in the English Language Classroom: Week 3
Grammar: Using a Corpus to Explore Grammatical Examples

Coca: The Collocates Function

https://youtu.be/mo1LlAYEmm0?list=PLRCcuy8rgTjdQUcgJi3fpLbkruWECbasn

Joe Geluso
Iowa State University

The next function that I want to look at is Collocates. This is one of my favorite functions in COCA.

A collocate or collocation is when two words co-occur or occur together frequently or have a really strong association with each other. To illustrate this, let’s look at our word “run” again. We put “run” in the word or phrase field, and for the collocates that we want to look for, we’ll make that a wild card or an asterisk, so we can see all the different words that collocate or co-occur with “run” in our corpus.

Figure 1. COCA Collocates function
Figure 2. Find collocates for "run"

Down here in this bar, we can specify how many words away from “run” we want to allow this wild card to occur. Let’s say that we want to capture all words that happen around three words after run and two words before. We can click on Find Collocates, and then, we’ll see what we get.

Figure 3. Number of words away from "run" for wild card to occur

Here, we have some really strange results. I don’t really know what these mean, and I’m actually glad this happened because a lot of times, you get strange results, and you’ll need to know how to deal with that. What we can see over here is an MI. We have really high MI. Ten is a really high MI. We kind of want to stick to MI around two or three or four.
Let’s go back to Search and change some of the settings. If you have something like this, where you get just a few words at a really high MI, or you get lots of words like prepositions and articles like “the,” “a,” and “at,” and they have a really low MI, then you want to change some things. The way to do that is to go down here to Sort/Limit, and you can click on this.

Now, we’ll see that Mutual Information (MI) is set at ten. That is too high. If you look over here on the side, you can see that it should’ve automatically been set to three, but that doesn’t always happen, so sometimes, you have to go in here manually and change it. MI is a measure of association between...
words, so the higher the MI score, the stronger that association, but when you get really high scores, then you’re just going to get low-frequency words that pretty much only occur with your target word, so I would suggest a MI score of two or three like COCA suggests.

Figure 6. MI should not be set at ten

We’ll change this now, and we’ll click on “Find Collocates” again to see if our results look a little different.

Figure 7. Find collocates for “run,” with MI of three

Good! These look much nicer, so we have things like “long,” “home,” “away,” and “short.” All these words collocate or co-occur with “run.” We’ll see here that our MI scores are much more reasonable.
### Figure 8. Collocates for “run”

Let’s try “long.” Let’s click on this and see what kind of context we get with “run.” Here, we have the long run,” “in the long run...” Look! We have lots of occurrences of this phrase “in the long run.” That’s good. Now, we can see that “long” co-occurs a lot with “run,” and we can see some typical phrases like “in the long run,” so this is a really interesting thing.

### Figure 9. Wider context of “long” as collocate of “run”

We can use Collocates to find a phrase like “in the long run,” and then, we can actually take it back to the List function and then just search the phrase “in the long run” that we just found through the Collocates function and see what we find in List.
This is a program of the U.S. Department of State administered by FHI 360 and delivered by Iowa State University

![Image of a program's logo]

---

<table>
<thead>
<tr>
<th>SEARCH</th>
<th>FREQUENCY</th>
<th>CONTEXT</th>
<th>ACCOUNT</th>
</tr>
</thead>
</table>

**Figure 10. Find matching strings for "in the long run"**

We have “in the long run” occurring 3,580 times. Interesting—so, it is a frequent phrase.

![Image of a chart showing frequency]

**Figure 11. Frequency information for "in the long run"**

Let me see how it’s used in its wider context. Now, we have all the instances of “in the long run.”
Figure 12. Authentic examples of usage of “in the long run”

You can have your students find frequent collocates of “run,” find phrases that those collocates are associated with, and then look them up in List to see the wider context.