Using Corpora in English Language Teaching: A Teacher's Experience

Adam Brod*
Notre Dame Seishin University in Okayama, Japan

Abstract
This paper recounts the process of using corpus data from on-line corpora (COCA and Google n-gram viewer) to answer questions asked by actual students of English in Japan. My own successes and challenges in these experiences can provide insights for other English language teachers to use corpora and data driven learning with their students.

Ivor Timmis concluded his book *Corpus Linguistics for ELT* with an observation that "some ELT practitioners are resistant to corpus use in a pedagogical context" (2006, p. 202). This, he suggested, is because "It is a big step from being made aware of what corpora can do to 'hands-on' use of corpora" (p. 202). I sympathize with the resistance language teachers (especially those who are teaching their native language) may feel towards the use of corpora for ELT. For a native English speaker, it is easy to convince myself that I can teach my native language without the use of a database of naturally occurring English. Furthermore, as a teacher following the communicative approach, I am wary of spending time using a machine to point out minute and possibly useless differences in the target language.

However, a student recently asked me to explain the difference between *electric* and *electrical.* Before learning how to answer this question meaningfully with corpus data, I might have given this student the same answer he got from his bilingual dictionary. His bilingual dictionary defined both words with the same definition, and I too might have said they are almost the same. But, knowing I could compare collocations of these words in *The Corpus of Contemporary American English* (COCA, Davies, 2008) to investigate if the difference between these two words was actually meaningless or not, I told the student I would look it up. On my way to the computer, I formed my own hypothesis of what the difference might be from my own lexicon. For example, I asked myself "What is the meaningful difference between *electric lightbulb* and *electrical appliance*?." I guessed *electrical* goes with abstract category words while *electric* goes with physical objects. By comparing the collocations of these words 1 space to the right in the COCA corpus, I found I was correct, and I had data or evidence to support my conclusion (Figure 1).

Corpus of Contemporary American English

SEARCH	FREQUENCY	CONTEXT	OVERVIEW

WORD 1 (W1): **ELECTRIC** (2.19)

	WORD	W1	W2	W1/W2	SCORE
1	MIXER	937	0	1,874.0	854.5
2	GUITAR	384	0	768.0	350.2
3	BOWL	271	0	542.0	247.2
4	MEDIUM	334	1	334.0	152.3
5	GENERAL	1633	6	272.2	124.1
6	GUITARS	98	0	196.0	89.4
7	CREAMY	93	0	186.0	84.8
8	HYBRID	174	1	174.0	79.3
9	RAZOR	82	0	164.0	74.8
10	STOVE	80	0	160.0	73.0
11	DRILL	76	0	152.0	69.3
12	BLANKET	74	0	148.0	67.5
13	BASS	73	0	146.0	66.6
14	CREAM	58	0	116.0	52.9
15	DRESS	56	0	112.0	51.1
16	TYPEWRITER	55	0	110.0	50.2
17	CORPORATION	49	0	98.0	44.7
18	HYBRIDS	49	0	98.0	44.7
19	TOKYO	48	0	96.0	43.8
20	STAND	47	0	94.0	42.9
21	TOOTHBRUSH	45	0	90.0	41.0
22	WELCH	45	0	90.0	41.0
23	PHILADELPHIA	43	0	86.0	39.2
24	NBC	41	0	82.0	37.4
25	BUTTER	40	0	80.0	36.5

WORD 2 (W2): **ELECTRICAL** (0.46)

	WORD	W2	W1	W2/W1	SCORE
1	DEGREE	77	0	154.0	337.7
2	PROFESSOR	94	1	94.0	206.1
3	BROTHERHOOD	34	0	68.0	149.1
4	CONNECTORS	29	0	58.0	127.2
5	ACTIVITY	238	5	47.6	104.4
6	ENGINEERING	591	13	45.5	99.7
7	CONDUIT	20	0	40.0	87.7
8	CAD	18	0	36.0	78.9
9	CONDUCTION	18	0	36.0	78.9
10	BACHELOR	17	0	34.0	74.6
11	CONDUCTIVITY	66	2	33.0	72.4
12	IMPEDANCE	16	0	32.0	70.2
13	RESISTIVITY	16	0	32.0	70.2
14	CONNECTIONS	61	2	30.5	66.9
15	CONTRACTING	15	0	30.0	65.8
16	GROUNDING	15	0	30.0	65.8
17	MALFUNCTION	15	0	30.0	65.8
18	PHYSICS	15	0	30.0	65.8
19	STANFORD	15	0	30.0	65.8
20	TENS	28	1	28.0	61.4
21	WRAPPED	14	0	28.0	61.4
22	SCIENCE	53	2	26.5	58.1
23	ABNORMAL	13	0	26.0	57.0
24	AUDITORY	13	0	26.0	57.0
25	CODES	13	0	26.0	57.0

Figure 1. Top 25 collocations for *electric* (left) and *electrical* (right) 1 space to the right

For the purposes of teaching this student to learn in an active way, instead of explaining my rational to him, I printed the list of the top 25 collocations for each term. I gave it to the student and asked him to guess what the difference might be. Although he did not have the extensive English vocabulary that I do, he was able to form the same hypothesis I came to.

After successfully using corpora in one actual teaching situation, I became much more inclined to use it for other students' questions. Another student asked what the difference is between *what do you think about* and *what do you think of*. Once again, I formed my own hypothesis first: *what do you think of* sounded older to me; *what do you think about* more contemporary. I was able to confirm my hypothesis by putting both into Google N-Gram Viewer (Figure 2).

Indeed, *what do you think of* has been declining since the year 1900, while *what do you think about* has been ascending in use since the year 1900, at least in books. Further, comparing what collocates with each phrase 2 spaces to the right in COCA (Figure 3), I could also see *what do you think of* collocates with more specific words—like first names, e.g., "What do you think of John?", whereas *what do you think about* collocates with bigger and more abstract ideas—like "What do you think about abortion" (Davis, 2008).
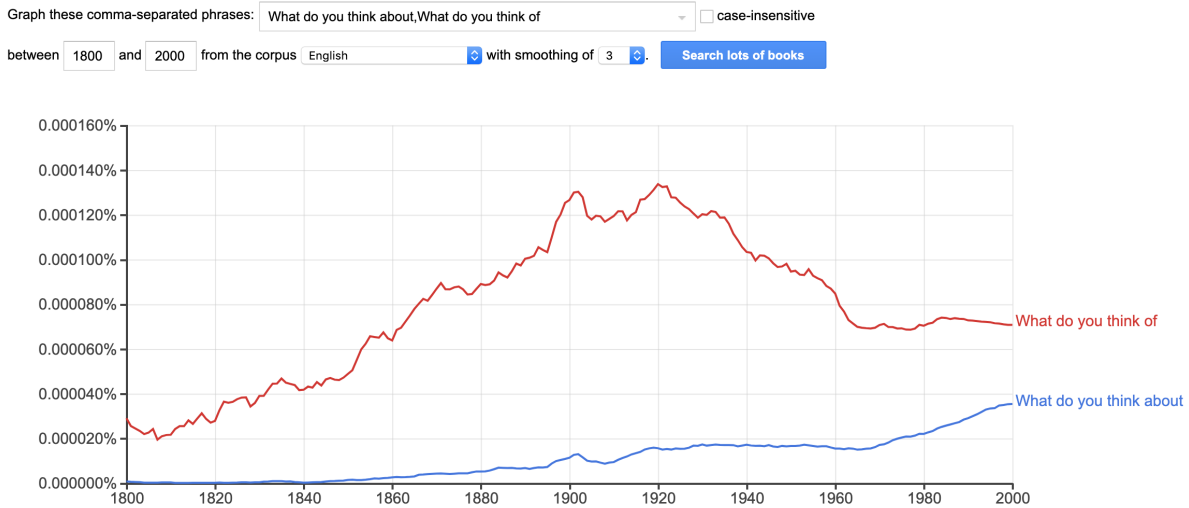
Figure 2. Google N-Gram Viewer for *What do you think of* (red) and *What do you think about* (blue)

	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	PUTTING	7	0	14.0	15.9	1	OF	2169	0	4,338.0	3,816.0
2	ABORTION	4	0	8.0	9.1	2	COMMENTS	27	0	54.0	47.5
3	USING	3	0	6.0	6.8	3	OUR	21	2	10.5	9.2
4	ROE	3	0	6.0	6.8	4	RANDY	5	0	10.0	8.8
5	GOING	5	1	5.0	5.7	5	JOHN	5	0	10.0	8.8
6	,	6	2	3.0	3.4	6	MOM	4	0	8.0	7.0
7	BEING	3	1	3.0	3.4	7	YOUNG	4	0	8.0	7.0
8	"	3	1	3.0	3.4	8	SEN	3	0	6.0	5.3
9	?	29	11	2.6	3.0	9	SADDAM	3	0	6.0	5.3
10	WHEN	23	11	2.1	2.4	10	TED	3	0	6.0	5.3
11	ALL	55	27	2.0	2.3	11	AS	3	0	6.0	5.3
12	HAVING	6	3	2.0	2.3	12	EARTH	3	0	6.0	5.3
13	TAKING	4	2	2.0	2.3	13	JACK	3	0	6.0	5.3
14	..	5	3	1.7	1.9	14	MY	52	10	5.2	4.6
15	PEOPLE	5	3	1.7	1.9	15	HOWARD	5	1	5.0	4.4
16	WHERE	3	2	1.5	1.7	16	JOE	5	1	5.0	4.4
17	TRUMP	3	2	1.5	1.7	17	MR	13	3	4.3	3.8
18	IT	84	65	1.3	1.5	18	THEM	21	5	4.2	3.7

Figure 3. Top collocations for *What do you think about* (left) and
What do you think of (right) 2 space to the right

Once again, I was able to print all of this out and discuss it with the student instead of merely telling her what I think and to take my word for it.

Finally, after making a habit of checking my own intuitions against real world data from corpora, I caught myself in a lie. A student asked me which is correct: *difficult to me* or *difficult for me.* I unequivocally told the student "*difficult for me* is correct and *difficult to me* is incorrect." However, in thinking about my advice, I told the student I would research the question and get back to her. Although the student had only asked me about those two language chunks, I had subconsciously assumed she wanted to say "English is difficult for me." This is why I answered "(English is) *difficult for me*" is absolutely correct and "(English is) *difficult to me*" is incorrect, but when I searched *difficult to me* in the list function of COCA and quickly scanned the concordance lines, I realized the chunk *difficult to me*, though less common, is equally acceptable. For example, *English seems difficult to me* is just as correct as *English is difficult for me*.

Furthermore, the language environment and the underlying meaning of any adjective (not just *difficult*) + *to me* versus an adjective + *for me* intrigued me. I wanted to be able to show the student the meaning difference between the two constructions. By reading a handful of concordance lines containing each language chunk, the pattern became apparent. In sentences containing the language chunk *difficult for me*, the focus of the sentence is *me*, with *me* being the actor undergoing or experiencing the action. For example,

"My camera guy passed out. It was certainly very difficult *for me.*"
"It's hard *for me* to remember right now."
"It's tough *for me* to stay in one place too long" (Davis, 2008).

By contrast, in sentences containing the language chunk *difficult to me*, the focus of the sentence is on the subject or agent, and *me* serves as an observer whose opinion in being expressed. For example,

"*Her job* sounded difficult *to me*.
"I think *the two year marker* was much more *difficult to me*.
"*The masters* seems more *difficult to me* than beating seven people in a grand slam"
(Davis, 2008).

This nuanced difference can be seen in this example. If I was looking at a yellowish-green picture, I might say "It's yellow *to me*." Another person might say, "It's green *to me*." But, if a friend and I were looking together at a picture that was always changing colors, I might say "It's yellow *for me*, but now it's black *for you*." It is worth noting how this investigation with corpora opened my mind to both possibilities.

However, thinking back to why I was so fixated on how *difficult* must be followed by *for me*, I wondered if some adjectives tend to be followed more frequently by *for me* while others tend to be followed more frequently by *to me*. This is the kind of search which might only be possible with corpora, and it is as quick and easy as any other key word search. By simply conducting two different list searches, one for any adjective preceding "to me" [_j* to me] and one for any adjective preceding "for me" [_j* for me] in COCA, I compiled two separate lists of the top ten adjectives which precede each language chunk (Figure 4).

10 most common adjectives with "TO ME"	10 most common adjectives with "FOR ME"
1. Important to me	1. Hard for me
2. Clear to me	2. Difficult for me
3. Good to me	3. Good for me
4. Interesting to me	4. Easy for me
5. Amazing to me	5. Important for me
6. Nice to me	6. Easier for me
7. New to me	7. Impossible for me
8. Obvious to me	8. Sorry for me
9. Familiar to me	9. Possible for me
10. Strange to me	10. Better for me

Figure 4. Top ten adjectives preceding *to me* (left) and *for me* (right)

By noticing the kinds of words which frequently collocate with each language chunk, I realized adjectives with *to me* tend to be evaluative adjectives which express a personal opinion, and adjectives with *for me* tend to be factually descriptive adjectives which express a self-judgment. Furthermore, with this list, students would be able to learn these common adjectives with the prepositional phrase *to me* or *for me* as a meaningful chunk of language instead of learning the adjectives separately from how they are commonly used, and then having to figure out which language chunk they might fit with. Both the adjective and the language chunk might thus become more meaningful.

As these examples show, corpus analysis has helped me to answer very specific questions from my students in a more concrete and effective way, and it has once corrected my intuition. I acknowledge that I have only used a few corpora functions regularly to help me with English language teaching. Nonetheless, even using a limited set of corpora functions has really helped me to answer my students' questions and become a better teacher. I firmly believe that any hands-on use of corpora by teachers will lead to more and more hands-on use, and will move teachers away from the view that corpora is only for computer scientists and linguists.

References

Davies, Mark. (2008-). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present* [Website]. Available online at https://www.english-corpora.org/coca/.

Google n gram viewer. Retrieved September 13, 2019 from http://books.google.com/ngrams

Timmis, Ivor. (2006). *Corpus Linguistics for ELT: Research and Practice* [Kindle for Mac]. Retrieved from https://www.amazon.com/.

About the Author:
Adam Brod's research and academic interests include translation in language teaching, building metaphorical awareness in a second language and the use of corpus linguistics in English Language teaching. He is generally interested in trauma theory and the pedagogical, psychological and cognitive linguistic benefits of bilingualism.