

Frankenberg-Garcia, A. (forthcoming) "How Language Learners Can Benefit From Corpora, Or Not". *Recherches en didactique des langues et des cultures: les cahiers de l'Acedle*. pre-publication version.

How language learners can benefit from corpora, or not

Ana Frankenberg-Garcia
University of Surrey

This paper discusses how learners can come into contact with corpora and the different ways in which corpora can be useful to them. There are three main ways in which this can be done, and rather than competing against each other, I see these three different gateways to corpora as being complementary. I also wish to show that, as with any kind of materials and activities developed under the aegis of a particular methodology and approach to learning, some uses of corpora in the classroom can be very effective, while others may not work at all.

Introduction

Most people will have come into contact with corpus-based language aids developed by experts without even realizing that the tools and resources they are using were built thanks to corpora. That is the first way in which the general public can benefit from corpora. Next, language learners can come into contact with corpora when their teachers have received some kind of training in corpus linguistics and decide to prepare corpus-based materials for classroom use. And finally, with guidance from their teachers, language learners can also have a go at using corpora themselves. In this paper I will discuss these three ways in which corpora can be useful for language learners, drawing particular attention to the many ways in which corpora can be directly exploited by teachers and learners and to some of the common pitfalls that should be avoided.

Published corpus-based tools and resources

People are not generally aware that computational linguists use corpora to develop all sorts of language tools that have become commonplace in our everyday lives, from simple spell checkers, to autocorrect options in word processors and web browsers, to sophisticated machine translation programs. This is the first way in which learners can experience the benefits of corpora is without realizing it. Likewise, people resorting to dictionaries may not know that a number of new-generation dictionaries have been compiled on the basis of corpus data. Lexicographers use corpora to find out how frequently words are used and then utilize this information to select which headwords are important to include in learners' dictionaries, which senses of polysemous words to present first, which words to use in the definitions, and which grammatical properties and collocations of words to draw attention to. Modern learners' dictionaries also complement basic information on senses with corpus-based examples that can help learners see how words are used in context, and with data from learner corpora in order to draw attention to certain aspects of language that can be particularly problematic.¹

¹ See <http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/>, Sinclair (1987) and Rundell and Atkins (2008) for information on how modern lexicography makes use of corpus data.

Apart from dictionaries, at least in the ELT market, there is a growing body of grammars and textbooks with pedagogical explanations and exercises that draw on corpus data². Figure 1 lists a few corpus-based dictionaries, grammars and textbooks for learners of English.

Figure 1. Examples of corpus-based ELT publications

Cambridge Dictionary of American English
Cambridge International Dictionary of English
Cambridge Grammar of English
Collins COBUILD English Dictionary for Advanced Learners
Collins COBUILD English Usage
Collins COBUILD Intermediate English Grammar
Longman Dictionary of Common Errors
Longman Dictionary of Contemporary English
Longman Grammar of Spoken and Written English
Macmillan English Dictionary
Macmillan Collocations Dictionary
Natural Grammar (Oxford)
Oxford Advanced Learner's Dictionary
Oxford Collocations Dictionary for Students of English
Practical English Usage (Oxford)
Touchstone series (Cambridge)
Vocabulary in Use series (Cambridge)

Materials like the ones outlined in figure 1 contain language that has been selected and edited by experts, so its users are not required to understand what a corpus is or to have direct contact with corpora. In the first book of the Touchtone Series (McCarthy et al 2005), for example, students get titbits of corpus information right from the start: in conversation, “*phone* is six times more common than *telephone*” (p.6), “*yeah* is ten times more common than *yes*” (p.9), “People use *'s not* and *'re not* after pronouns [...] *Isn't* and *aren't* often follow nouns” (p.25), “People say *Sometimes I* seven times more often than *I sometimes*” (p.46) and so on. On the one hand, this is excellent, since a lot of the information that can be extracted from raw corpus data has already been processed by experienced professionals, which means teachers and learners can benefit from the final product without having to worry about how it was obtained. On the other hand, however, the amount of language that can be presented in this polished format is limited, simply because language is infinitely bigger and more complex than what can be summarized in a book or any other language learning aid. Language learners (and their teachers) often have questions for which there are no answers or which are not treated in sufficient detail in dictionaries, grammar books and other publications. Corpora can provide more language, and can disclose solutions to language queries that have not been dealt with in edited linguistic resources. This is why another option is for teachers and learners to use corpora directly, as shall be seen below.

Direct uses of corpora

When Tim Johns and his students at the University of Birmingham began exploring corpora directly in one-to-one student-teacher consultations in the late eighties, in an approach to learning that came to be known as data-driven learning (Johns 1991), the use of computers was not widespread and corpora were not accessible to the general

² See O’Keeffe et al (2007) for an account of how corpus data can be harnessed to produce textbooks for language learners.

public. More than twenty years have gone by, and nowadays there are a number of corpora in various languages that can be easily accessed online by anyone with a connection to the Internet. It is no longer necessary to be a linguist, a lexicographer or any other kind of language expert working at a research centre or for a commercial publisher to be able to access a corpus. Although according to Mukherjee (2004) and Frankenberg-Garcia (2012a) not many language teachers, let alone learners, will have heard of corpora, at least in Britain there now seem to be a number of MA in English Language Teaching degrees that include modules or module components about corpora. There are also books about using corpora in the classroom Reppen (2010) and Bennet (2010), and several online tutorials too.³

Language teachers who have received some training in corpus linguistics can resort to corpora to complement their teaching in all sorts of ways. This means learners can come into contact with corpora indirectly, by means of corpus-based materials and activities prepared by their teacher, or directly, when they are prompted by their teacher to use corpora themselves. Gabrielatos (2005) has referred to this as the ‘soft’ and the ‘hard’ approach to corpus use, while Boulton (2010a) termed this same distinction the ‘hands-off’ and the ‘hands-on’ approach. In this paper I shall refer to the latter, simply because it is more mnemonic. As observed in Frankenberg-Garcia (2012b), the hands-off and the hands-on approaches are not mutually exclusive, and I hope it will become clear in this paper that there is a time and place for both.

Using corpora hands-off

When resorting to corpora hands-off, learners do not need to learn how to use concordancing software, and as pointed out by Boulton (2010a), this can be a great advantage, since learners may find it difficult to understand how to work with the software and may not know what exactly to do with corpora or how to interpret the results they get. There are countless different types of hands-off corpus-based materials and activities that teachers can prepare for their students, but it is crucial that teachers are able to use them judiciously.

The mere fact that a handout or a classroom activity is based on corpora is no guarantee of quality or success. Let me give an example. Imagine you are learning English and your teacher has given you a printout with the concordances in figure 2 and asked you to try and discover what *grimple* means. By reading them, you can deduce that *grimple* is a noun, that people wear *grimps* (probably on their heads) and that a *grimple* can hide your face. Putting it all together, you can get a vague idea of what *grimple* might mean, but maybe you are not 100% sure.

From a purely linguistic perspective, or from the viewpoint of language awareness, this is a very interesting exercise, as it helps learners to realize that meaning can be derived from context. But why would learners bother with exercises such as this one, when it would have been far simpler and a lot faster to look up the word in a dictionary? In the present case, the concordances for *grimple* are in actual fact concordances for *helmet*. According to the Macmillan English Dictionary Online⁴, a helmet is *a hard hat that you wear to protect your head*, and according to the English-French dictionary at Oxford

³ For example, see http://calper.la.psu.edu/corpus_portal/tutorial_overview.php, http://www.ict4lt.org/en/en_mod3-4.htm and <https://eltadvantage.ed2go.com/>, <http://corpuscall.eu/> [27/10/2013]

⁴ <http://www.macmillandictionary.com/dictionary/british/helmet> [27/10/2013]

Dictionaries Online⁵, a helmet is a *casque*. If you did not know the meaning of *helmet*, it would have been a lot quicker to find out what it meant by looking it up in a dictionary instead of being asked to analyse corpus data like a trained lexicographer or linguist. Learners may very well perceive it is a waste of time and effort to spend so much time analysing concordances when they can find out what words mean more efficiently and effectively using a dictionary. Indeed, this kind of corpus-based exercise can be downright frustrating, especially if learners only have a few hours a week to study a language.

Figure 2. Random concordances for *grimple*⁶ from COCA⁷

A smooth cheek, nearly hidden beneath **grimple** and sunglasses.
Protected by a **grimple** of bone, the brain was an organ of mind-boggling complexity.
Shoshana noticed her hair was so perfectly sleek it looked like a blond **grimple**.
Megan flopped down next to me and yanked off her **grimple**.
I heard my own laughter, echoing inside the **grimple**.
a woman with a ponytail sneaking out from under her **grimple**.
The balaclava can also help counteract chills from your vented **grimple**.
You can be pretty anonymous with a **grimple** on and I wanted to keep it that way.
She said he wore a **grimple** with the visor down.
During rehab, she will probably wear a **grimple**.

A very different situation is to use concordances to reinforce the meaning of words that learners have heard for the first time in class or elsewhere, and to which they have not had the opportunity to be exposed more fully. In Frankenberg-Garcia (2012b), I gave the example of learners not understanding the meaning of the word *aisle*, which they had encountered for the first time in a dialogue in their coursebooks. The word appeared in the context of air travel, and its meaning in that sense was briefly explained immediately after the learners had asked what it meant. In the following lesson, the learners were then given the hands-off exercise in figure 3. Although the concordances used in the *aisle* exercise are essentially similar to the ones used in the *grimple* exercise in that they both focus on a given word and draw attention to the contexts in which that word is used, the two exercises are in actual fact totally different. Rather than focussing on a random word selected by the teacher, the concordances for *aisle* focus on a word that the learners were already curious about because they had not understood it when it was presented to them in an authentic classroom situation. Rather than asking the learners to use concordances figure out the meaning of a word at random as if they were linguists or lexicographers, the concordances for *aisle* were used to expand the learners' previous one-off contact with the word in the classroom. With the *aisle* exercise, the learners' exposure to the new input was exponentially enhanced with concentrated doses of the word in context. The concordances also enabled the learners to figure out that aisles exist not just on aeroplanes, which was the original context in which they had seen the word, but also in places like trains, shops, churches and supermarkets. Additionally, the concordances served to help the learners notice that there is a

⁵ <http://www.oxfordlanguagedictionaries.com> [27/10/2013]

⁶ As explained in the main text, *grimple* is an invented word used in these concordance lines to replace a real English word.

⁷ Corpus of Contemporary American English, available at <http://corpus.byu.edu/coca/> [27/10/2013]. See also Davies (2008).

distinction between aisles and corridors, which does not exist in the learners' native Portuguese.

Figure 3. Handout with selected BNC⁸ concordances for aisle (Frankenberg-Garcia 2012b:40)

A. Read the sentences below and make a list of the sort of place where aisles are found.

B. Does aisle translate into Portuguese always in the same way?

1. The air hostesses inquired what I was making and a man passing in the aisle quite genuinely complimented me on my work.
 2. I arrived at Salisbury Cathedral, just as the bride was about to go up the aisle.
 3. As she looked around she felt a twinge of sadness that in a carriage where 70 per cent of the commuters were men there were five women forced to stand in the aisle.
 4. They looked at the passports and then started to walk down the aisle, pointing their guns at the passengers.
 5. He hurried up the aisle of the church.
 6. She picked up her suitcase and made her way along the aisle.
 7. The layout of the store, with wide aisles, gives customers room to move around.
 8. I spend much of my time at the shops; wandering through the aisles, faltering, never knowing what to buy.
-

In addition to enhancing language comprehension, concordances can also be very useful in language production. Sometimes learners already know the meaning of words but have problems using them correctly. Concordances illustrating how words are used in context can help learners overcome these difficulties. For example, a frequent error by Portuguese learners of English is to say **congratulations for something*. Figure 4 shows how a handout with concordances for *congratulations* can help these learners notice what prepositions to use after *congratulations*.

Figure 4. Handout with selected concordances from enTenTen⁹ for *congratulations*.

A. Underline the prepositions used after *congratulations*.

B. What sorts of things come after each preposition?

- Our congratulations to all the candidates and to the winners.
Congratulations on a job well done.
Congratulations from us all.
Congratulations on your impending marriage!
I do not expect congratulations from you.
Congratulations on your 100th birthday!
Congratulations on your decision.
Congratulations to whoever made that call!
Congratulations to the artist.
So congratulations to all the authors.
We got lots of congratulations from passers-by.
-

Concordance-based activities can also help learners expand their vocabulary and use the target language more idiomatically. As exemplified in figure 5, Frankenberg-Garcia (2012b) suggests an activity with concordances for *the bus*, used to help learners employ more idiomatic English when writing a composition about something that happened on a bus.

⁸ <http://www.natcorp.ox.ac.uk/index.xml> [27/10/2013]

⁹ The enTenTen corpus is a large web-based corpus available from the Sketch Engine at <https://the.sketchengine.co.uk/> [1/11/2013]. See also Kilgarriff et al (2004) and Jakubíček et al (2013)

Figure 5. Handout with selected BNC and COCA concordances for *the bus* (Frankenberg-Garcia 2012b:43)

Notice the words that go with 'the bus' in the sentences below. Can some of the expressions in bold be useful in your composition?

1. Back **at the bus stop**, the other people were furious.
 2. He bought this old factory the other day, down **by the bus station**.
 3. You **missed the bus**, " she said.
 4. When **the bus pulled up to our school** on Lincoln Street, I stood up.
 5. I saw a man trying **to catch the bus** into Manhattan as it pulled out of the station.
 6. Once I did that I was able **to get on the bus**.
 7. She reached the gate just as **the bus driver was collecting** the last of **the tickets**.
 8. Blind as well?'; **the conductor** asked, and **rang the bell to stop the bus**.
 9. The people **in the bus queue** are going to work.
 10. They'd all **got off the bus** together.
 11. He walked to **the bus shelter** at the roadside, and waited.
 12. Benny knew they would come **to meet her off the bus**.
 13. I'm usually **late for the bus**.
 14. I always **miss the bus**.'
 15. I used **to ride the bus** by myself when I was your age.
-

Hands-off activities need not be endless lines of concordances printed on paper. When concordances are presented to groups of students in slide format, it is easier for the whole class to discuss them together. Figure 6, shows a screenshot of a concordance-based slide presentation devised to help learners understand which prepositions to use after *pay*. It was developed to complement a unit about payment methods in the regular textbook adopted in class, as a reaction to the observation that the students were making many errors regarding the use of the prepositions used in the context of *pay*. In the first few slides, the learners were led to figure out the differences between paying people (no preposition), paying *for* things in general, paying bills/taxes/rents/etc.(no preposition), paying *in* different currencies and cash, and paying *by* cheque/credit card/etc. In the excerpt shown, the learners practised using the right preposition where appropriate in a conventional fill-in-the-gap format. What was different, however, was that instead of the bland content that is usually given in traditional exercises for practising grammar, the concordance lines used (extracted from COCA) prompted the students to discuss all sorts of issues related to them, such as why Americans living in Europe and being paid in dollars feel like they have taken a pay cut (concordance 1), why someone in Russia would get paid in dollars (concordance 3) and whether they were paid to do odd jobs at home when they were little (concordance 10). Thus in addition to being exposed to massive doses of the word *pay* in context, the students also had stimulating conversation topics to discuss as a group, in what may have otherwise been just another solitary and boring grammar practice exercise.

Figure 6. Excerpt of classroom slides with a concordance-based exercise to practise the prepositions used after pay¹⁰.

The figure consists of four separate slides arranged in a 2x2 grid. Each slide has a dark green background with white text. The top-left slide contains the instruction 'Fill in the gaps with a preposition when necessary' and two numbered sentences. The top-right slide contains two numbered sentences. The bottom-left slide contains three numbered sentences. The bottom-right slide contains two numbered sentences.

Slide 1 (Top Left):

Fill in the gaps with a preposition when necessary

1. Americans living in Europe but being paid dollars feel as if they've taken a big pay cut.
2. Users and sellers would have to be persuaded to pay the tax.

Slide 2 (Top Right):

3. Anton's mother had been working as the cleaning lady at his summer house for almost a year now, and he paid her dollars, not rubles.
4. Brenna's grandmother paid the trip with what would have been Brenna's college money.

Slide 3 (Bottom Left):

5. He foolishly agreed to pay Daniel the outrageous sum of fifteen thousand dollars to represent him in court.
6. How do we pay the meal?
7. I never paid the book. I stole it.

Slide 4 (Bottom Right):

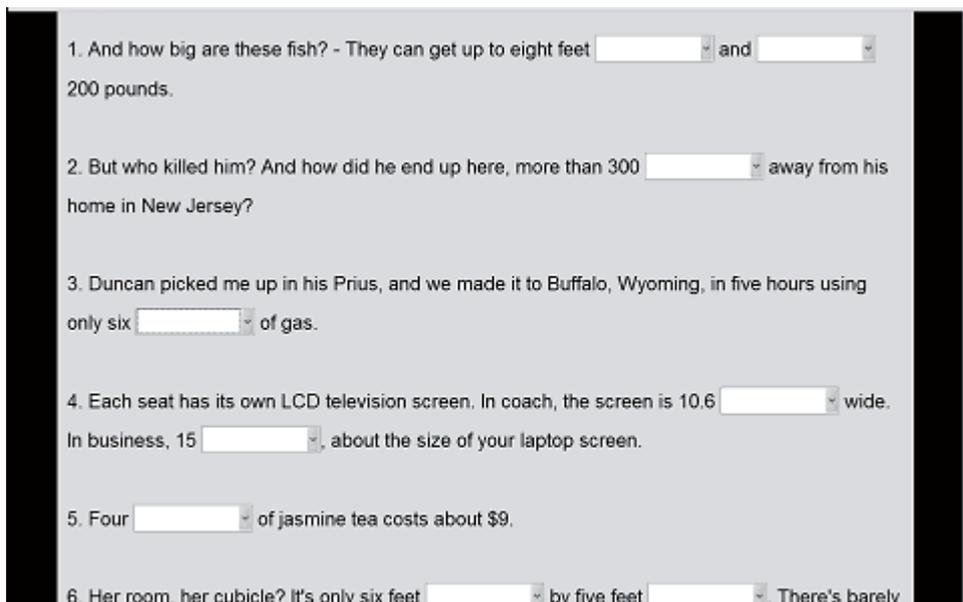
8. Owen paid the bill and drove her home.
9. I paid my latte and dropped a \$5 bill into the tip cup.
10. I paid her \$5 to cut the grass.

Another way of working with hands-off concordances without paper is to transpose them to interactive electronic exercises for the Web. With the help of user-friendly freeware like HotPotatoes¹¹, teachers can prepare concordance-based interactive exercises which learners can then work on at their own pace. Figure 7, taken from Frankenberg-Garcia (2012b), shows a screenshot of a computer-assisted language learning (CALL) exercise with concordances from COCA used to help Portuguese learners of English studying English for Tourism become more familiar with non-metric units of measure utilized in the USA. In this exercise, the students choose from a dropdown list of options words like *feet*, *inches*, *gallons*, *ounces*, *wide*, *long*, *high*, *weigh* and so on to fill in the gapped concordances.

¹⁰ The entire slideshow can be downloaded from <http://anafrankenberg.synthasite.com/resources/Pay%2BPreposition.pps> [30/10/2013]

¹¹ HotPotatoes was developed by Stewart Arneil and Martin Holmes at the University of Victoria, Canada, and can be downloaded from <http://hotpot.uvic.ca/> [31/10/2013]

Figure 7. Non-metric units concordance-based CALL exercise (Frankenberg-Garcia (2012b:42)¹²)



1. And how big are these fish? - They can get up to eight feet and 200 pounds.

2. But who killed him? And how did he end up here, more than 300 away from his home in New Jersey?

3. Duncan picked me up in his Prius, and we made it to Buffalo, Wyoming, in five hours using only six of gas.

4. Each seat has its own LCD television screen. In coach, the screen is 10.6 wide. In business, 15 , about the size of your laptop screen.

5. Four of jasmine tea costs about \$9.

6. Her room, her cubicle? It's only six feet by five feet . There's barely

Although concordance-based exercises like the ones outlined above can be very stimulating, it is very important to bear in mind that the learners using them will not have received any training in corpus linguistics. Note that in all concordance-based activities exemplified so far, not once was the word concordance mentioned. There is no reason why teachers should confuse learners with corpus-linguistics terminology when it is perfectly possible to give instructions using general words like sentence or sentence extract instead of concordance.

Printing out concordances and asking learners to observe them and analyse what they notice is also probably asking too much. Learners should receive very specific guidelines on what to look out for in the concordances or they will probably not understand what the purpose of the exercise is. In the examples given, learners are guided along what they are supposed to do with the data and are led to interpreting it in the ways the teacher had planned them to interpret it. Although language is not black and white and it is not possible to predict all that is going happen, i.e., learners will also notice things which the teacher had not anticipated, the point I want to make here is that learners will not be lost in a sea of data and the essential aspect of language that the teacher had intended to present will have been covered.

Another recommendation is for teachers to select which concordances they want to transpose to a handout rather than printing them out indiscriminately. There is no harm in editing the concordances presented to learners so as to correct languages mistakes¹³, leave out language that is considered too difficult for the level of proficiency of the learners, and exclude language that is offensive or sensitive or for any other reason

¹² The full exercise can be downloaded from <http://anafrankenberg.synthasite.com/non-metric-units.php> [31/10/2013]

¹³ The last concordance line in figure 4 is an edited version of what was originally **We got lots of congratulations from passer-bys.*

inappropriate for classroom use. The fact that the teacher can shelter learners from many of the problems of dealing with raw corpus data is an added advantage of presenting learners with hands-off concordances.

Having said this, a learner's hands-off contact with corpora need not be based on concordances alone. As demonstrated in Frankenberg-Garcia (2012b), collocation data can be used to boost vocabulary related to practically any topic. Figure 8, for example, shows a word sketch – i.e., a one-page, automatic corpus-based accounts of a word's grammatical and collocational behaviour (Kilgarriff et al 2004) – for the noun *guitar* in the enTenTen corpus (Jakubíček et al 2013). A teacher can instantly select from this screenshot words that learners might want to use when talking about guitars, such as what people can do with guitars (strum, play, tune, etc.), what types of guitars there are (acoustic, electric, bass, classical, etc.), what sort of nouns can be modified by guitar (riff, solo, chord, etc.), and so on. These can be presented to learners in a brainstorming session or in any other format.

Figure 8. A word sketch for *guitar*¹⁴

guitar (noun) enTenTen12 freq = 475021 (36.6 per million)									
object of 113493 0.3		subject of 52189 0.2		adj subject of 7089 0.2		modifier 183252 0.2		modifies 183390 0.2	
strum	1504 8.59	strum	629 8.28	left-handed	23 5.15	acoustic	19677 11.04	riff	5001 9.3
play	36699 7.38	solo	405 7.83	instrumental	80 3.99	electric	16954 9.42	solo	7519 9.07
tune	1722 7.34	riff	311 7.46	acoustic	59 3.89	bass	8148 8.85	chord	4438 8.61
pedal	350 5.95	pedal	265 6.12	peripheral	29 3.54	classical	3701 8.21	playing	4531 8.21
chime	204 5.57	shred	96 5.29	reminiscent	28 3.18	lead	2989 7.82	amp	2359 7.88
harp	198 5.57	tune	316 5.17	loud	81 3.18	rhythm	2355 7.27	bass	3394 7.59
jangle	141 5.29	sound	1163 5.09	prominent	38 1.92	Fender	763 6.97	lick	1269 7.53
master	387 5.29	wail	72 4.93	versatile	20 1.67	distorted	816 6.95	string	4510 7.38
plug	321 5.18	jangle	51 4.89	online	486 1.11	-string	692 6.93	lesson	7484 7.33
wail	151 5.16	fret	79 4.71			vocal	1465 6.87	neck	3808 6.95
pluck	169 5.1	hook	191 4.64			beginner	1556 6.87	drum	1288 6.86
smash	219 5.07	pluck	61 4.31			Gibson	986 6.72	tab	2150 6.81
and/or 99710 0.2		pp_obj of-i 25393 0.2		pp_obj on-i 20627 0.6		pp_obj with-i 10786 0.3		pp in-i 6910 0.1	
vocal	6987 9.46	strum	187 7.27	strum	157 7.2	arm	153 4.98	band	552 2.93
drum	4960 9.18	neck	3337 7.01	chord	336 5.72	ballad	28 3.48	lap	60 2.62
bass	7567 8.97	headstock	43 5.57	wail	61 5.46	experiment	28 2.95	tune	84 2.54
mandolin	1538 8.71	twang	47 5.51	strap	124 5.39	vocal	41 2.5	hand	544 1.29
banjo	1616 8.61	playability	37 5.15	fret	59 4.84	solo	49 2.49	studio	52 1.01
piano	4928 8.59	fretboard	33 4.91	lick	52 4.76	synonymous	18 2.42		
harmonica	945 8.04	fret	45 4.35	improvise	51 4.75	melody	30 2.28		
violin	1486 7.95	accompaniment	53 4.32	jamming	37 4.72	amp	22 2.21		
percussion	1035 7.73	string	431 4.27	riff	72 4.6	dude	28 2.16		
keyboard	3336 7.54	mastery	61 4.13	Clapton	27 4.48	stool	19 1.71		
fiddle	770 7.43	fret	25 4.04	shred	33 4.46	pair	18 1.69		
flute	941 7.41	Neck	27 4.01	string	450 4.34	sing	62 1.58		

Using corpora hands-on

In addition to teachers preparing hands-off activities and materials suitable for their particular groups of learners, there is also the option of learners using corpora hands-on. This has often been reported as being more problematic than using corpora hands-off (for example, see Gabrielatos 2005 and Boulton 2010a), for in addition to computers

¹⁴ See footnote 8.

being needed in the classroom, learners will also have to grapple with new technology, raw corpus data and a new approach to learning, all at the same time. From technological viewpoint, language classrooms are not usually equipped with computers, language learners may not be technology lovers, computers can crash, the online connection can fail, and concordancing software does not tend to be particularly user-friendly. From the perspective of the corpus itself, learners have to decide which corpus to use, understand the strengths and limitations of the chosen corpus and interpret the results they get in the light of the first two of these factors. And from the viewpoint of learning, learners may feel overburdened with having to decide what aspect of language to explore, with having to read unedited text, and with not knowing what to look out for when scrolling down countless lines of concordances.

As if all this were not enough to completely discourage the hands-on use of corpora by language learners, another problem is that many of the hands-on activities proposed in the literature encourage learners to use corpora as if they were linguists. Textbooks about using corpora with language learners are full of ideas for hands-on activities that can be carried out in the classroom, but while interesting from the viewpoint of language research, many of the suggestions given tend to be dissociated from the reality of everyday language learning. Reppen (2010:44), for example, suggests that learners can look up in MICASE¹⁵ which texts “have the highest normed counts for *well* and then go to those texts and see if *well* is being used as a filler [...], as a discourse marker [...] or as an adverb”. Another activity proposed by Reppen (2010:62) is to ask learners to create frequency lists for newspaper texts and scan the resulting word lists for difficult words in order to determine the level of difficulty of each text. Exercises such as these can be interesting for a linguist, but I would argue that a language learner simply wants to be able to use the language rather than analyse it. Thus in the same way as the hands-off *grimple* exercise described earlier is relevant to the linguist and the lexicographer but not necessarily to the language learner, hands-on exercises too must focus on the interests and needs of learners.

As I argued in Frankenberg-Garcia (2012b), hands-on corpus consultation seems to be particularly suited to answer individual questions by learners that emerge when they are writing or when they are receiving feedback to their writing, such as ‘how do you say this?’, ‘is this right?’, ‘which word sounds better, this or that?’, ‘what verb/adjective/preposition/etc. should I use here?’, ‘why is this wrong?’ and so on. If all learners are asking the same questions, then perhaps it would be worth preparing hands-off activities for the group as a whole. However, there will be many questions that are not relevant to the entire class, but only to individual learners. This is when it might be useful to teach learners to use corpora to look up the answers they need by themselves. Of course, if the answers can be found more readily in dictionaries, learners should use dictionaries first for the simple reason that it will be quicker. However, not even the best learners’ dictionaries will have the answers to all questions that emerge when a person is learning a new language.

Learners do not need to be taught what corpora are or be trained to carry out linguistic analyses as if they were linguists. All that is really necessary is to show learners how to look up the answers to their specific questions. In Frankenberg-Garcia (2012b:42) I gave the example of a learner wanting to know which preposition to use in the context

¹⁵ Available at <http://quod.lib.umich.edu/m/micase/> [2/11/2013]

of ‘*I received a message ___ my cell phone*’. As the information provided in a learners’ dictionary was not enough, the teacher demonstrated how the learner could look up concordances for *cell phone* in the context of *message* in COCA so as to retrieve the concordance lines in figure 9, from which it became instantly clear to the learner which preposition to use.

Figure 9. Concordances for *cell phone* in the context of *message* in COCA (Frankenberg-Garcia 2012a:42)



In another situation, during a session where learners were looking at different ways of ending business letters, one of the students questioned the use of *I look forward to hearing from you*. She said her former tutor (a native speaker of English) had told her that the right way of saying this was using the present continuous: *I am looking forward to hearing from you*. The teacher (a non-native speaker) felt both forms were correct, but for reassurance she showed the student how to look up the strings *look forward to hearing/seeing* and *looking forward to hearing/seeing* in the Business Letter Corpus¹⁶, which contains around 1 million words of sample UK and US business letters. The results summarized in figure 10 showed both the student and the teacher that not only it was perfectly acceptable to end a letter with *look forward to hearing/seeing*, but also that it seemed in fact to be more conventional than *looking forward to hearing/seeing*.

Figure 10. Distribution of *looking/look forward to seeing/hearing* in the Business Letter Corpus

Search string	Corpus Frequency
look forward to hearing	212
looking forward to hearing	19
look forward to seeing	156
looking forward to seeing	15

¹⁶ Available at <http://www.someya-net.com/concordancer/index.html> [26/10/13]. See also

Demonstrating to learners that they can use ‘this online tool’ (there is no need to say it is a corpus) ‘like this’ (showing how to carry out the relevant query) to find out answers to questions about language that are in the forefront of their minds can help them see the utility of corpora immediately. If students are not all asking questions at the same time, a single computer in class used like a shared dictionary will often suffice. Little by little, with more questions and more demonstrations on how corpora can help to fulfil their language needs, learners are bound begin to use corpora autonomously, without the teacher’s support. It is unlikely that they will ever become power users of corpora capable of carrying out sophisticated queries and analyses, but I do not think this is what the average language learner aims for. It will already be a great achievement if learners manage to use corpora hands-on to answer their own immediate questions about language.

Conclusion

I hope to have shown that learners can benefit from corpora in three ways: via published corpus-based language learning resources developed by experts for the general public, via tailor-made hands-off exercises and activities prepared by their teachers, and via hands-on queries they carry out themselves. I also hope to have made it clear that there is a time and place for each way of using corpora, and that the fact that a resource or activity is corpus-based does not automatically mean that it is good. In particular, there are numerous hands-off exercises and activities that can be created, but when preparing them teachers must make sure that they are relevant, useful and accessible to their particular group of learners. Likewise, there are hundreds of ways in which learners can explore corpora hands-on, but it is easy for learners not to know what queries to carry out and what to do with the data retrieved. Most importantly, learners cannot be expected to be captivated by analysing corpus data, just because it is fascinating to linguists and language teachers. Unfortunately, quite a few of the corpus activities for language learning exemplified in the literature have more to do with the linguist’s interest in language than with language learners’ actual needs. On the other hand, if teachers bring to the classroom corpus-based materials and activities conceived to complement the regular materials adopted in class with extra input that will help learners develop their language skills more fully, and if learners are taught how corpora can be used hands-on to help them find the specific information about language that they cannot find in elsewhere, then I believe the direct use of corpora by language learners can be made to work. To conclude, I should like to point out that when examining different studies about learners’ reactions to data-driven learning¹⁷, one should be extra careful when comparing one study with another for the simple reason that there different ways in which data-driven learning can be implemented in the classroom, some being much better than others.

References

- Bennett, G. (2010) *Using Corpora in the Language Learning Classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.
- Boulton, A. (2010a) Data-driven learning: Taking the computer out of the equation. *Language Learning* 60(3): 534-572.

¹⁷ See Boulton (2010b) for a comprehensive account of such studies until 2010.

- Boulton, A. (2010b) Learning outcomes from corpus consultation. In Moreno Jaén, M., Serrano Valverde, F. & Calzada Pérez, M. (eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching*. London: Equinox, 129-144
- Davies, M. (2008) The Corpus of Contemporary American English (COCA): 410+ Million Words, 1990–Present. <<http://www.americancorpus.org>>.
- Frankenberg-Garcia, A. (2012a) "Raising Teacher's awareness of Corpora". *Language Teaching*, vol 45, 4, pp 475-489.
- Frankenberg-Garcia (2012b) "Integrating corpora with everyday language teaching". In Thomas, J. and Boulton, A. (eds.) *Input, Process and Product: developments in teaching and language corpora*. Brno: Masaryk University Press, pp 36-53.
- Gabrielatos, C. (2005) Corpora and language teaching: Just a fling or wedding bells? *Teaching English as a Second Language Electronic Journal* 8(4): 1-35. Available at <http://tesl-ej.org/ej32/a1.html> [17/09/11].
- Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlý, V. Suchomel (2013) *The TenTen Corpus Family*. Proceedings of *Corpus Linguistics*, Lancaster, UK.
- Johns, T. (1991) Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (eds), *Classroom Concordancing*. *English Language Research Journal* 4: 1-16.
- Kilgarriff, A., Rychlý, P., Smrz, P., and Tugwell, D. (2004) *The Sketch Engine*. Proceedings of EURALEX. Lorient, France
- McCarthy, M., McCarten, J. and Sandiford, H. (2005) *Touchstone*. Student book 1. Cambridge: Cambridge University Press.
- Mukherjee, J. (2004) Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor (ed.), *Applied Corpus Linguistics: A multidimensional perspective*. Amsterdam: Rodopi, p. 239-250.
- O'Keeffe, McCarthy, M. and Carter, R. (2007) *From Corpus to Classroom*. Cambridge: Cambridge University Press.
- Reppen, R. (2010) *Using Corpora in the Language Classroom*. New York: Cambridge University Press.
- Rundell, M. and Atkins, S. (2008) *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Sinclair, J. (ed.) (1987) *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London and Glasgow: Collins ELT.