


1 Corpora and language learning

In this chapter you will get answers to the following questions:


- What is a corpus?
- Why use a corpus with language learners?
- What are some ways to use a corpus with language learners?
- What will corpus-based activities look like?

In recent years there has been an increased interest in corpus linguistics along with an increased interest in using corpora for language instruction. The goal of this book is to provide teachers and teachers-in-training with the background and information needed to use corpora for language teaching. You will learn how to use corpora as a resource for developing materials and activities for a variety of classroom language-teaching situations.



A different use of corpus linguistics, covered in other books, is to discover patterns of language use, which takes advanced research skills and often involves computational skills. This book will not explain the methods needed to carry out such corpus research but builds on that research. You will discover that it is relatively easy to use existing corpus-research findings and corpora to enhance your teaching. If you are interested in the research aspect of corpus linguistics, there are several books listed in Additional Reading at the end of this chapter which provide a solid introduction to corpus-linguistic research. In addition, there is an extensive list of both online and print resources provided in Appendix B.

Let's start by addressing some questions that will help to provide a foundation for the topics covered in the following chapters. These questions include:

- What is a corpus?
 - Is one corpus as good as the next?
 - Should I use a corpus to teach my students English?
 - How can I use a corpus to teach my students English?
 - How can I adapt and develop materials from corpora for use in my classroom?
- 

A logical place to begin to answer these and other related questions is with an explanation and overview of what a corpus is.



What is a corpus?


In the world of corpus linguistics, a corpus is a large, principled collection of naturally occurring texts (written or spoken) stored electronically. Let's look more closely at this definition by answering the following questions: What is meant by "naturally occurring texts"? What about spoken language? What is "a principled collection"? How big is "large"?

What is meant by "naturally occurring texts"?

Naturally occurring texts is language that is from actual language situations, such as friends chatting, meetings, letters, class assignments, and books, rather than from surveys, questionnaires, or just made-up language.


What about spoken language?


Collecting a corpus of spoken language involves recording and then tran-



scribing the spoken language. Creating written transcripts of spoken language can be quite time-consuming and involves a series of choices based on the interests of the corpus compilers. For example, if researchers are interested in how pauses are used, they may time the pauses between words and also between speaker turns. If this is not a primary concern of the researchers, then they may only note long pauses (e.g., those over five seconds) or not note any pauses. Transcribing a spoken corpus with prosodic information (rising and falling intonation) is a major undertaking and will often be accomplished in several stages. The first stage is a rough transcription; next the transcript is reviewed to mark the rising and falling patterns of the words. Even the supposedly simple task of just getting spoken words into written form requires several decisions. How will spoken contractions be transcribed? For example, if a speaker says, “Are you gonna call Sam tonight?” do you transcribe *gonna* as it was spoken (*gonna*), or as the conventional written version *going to*? Other examples include *kinda* instead of *kind of*, *gotta* instead of *got to*, and the reduced forms of *cuz* or *coz* for *because*. If you are interested in exploring whether these contracted or reduced forms have a pattern, or if they tend to occur with certain words and not others, then transcribing them true to the form uttered (e.g., writing *kinda* instead of *kind of*) would be essential.

Even with written texts, corpus compilers often have to make decisions about spelling conventions, punctuation, and errors such as word







omissions or grammatical errors. The intended uses of the corpus shape the decisions made during the compilation of a corpus. For example, a corpus of essays written by language learners may prove a useful resource for teaching editing strategies. In this case, it would not be a good idea to correct spelling or other errors, since a class activity could involve editing the essays and discussing the types of changes that were made.

What is “a principled collection”?

The design of the corpus must be principled: The goals of the researcher or teacher shape the design of the corpus and guide the collection of texts. The texts in a corpus need to represent the type of language that the corpus is intending to capture. For example, if a corpus is to be representative of written language, then the corpus designer would need to make a comprehensive list of the different written language situations (e.g., fiction, academic prose, personal letters, office memos) and then create a plan to collect these various texts.

The task of collecting a general representative corpus is enormous. Fortunately, it is not necessary for interested teachers to build their own corpora. Several general corpora are readily available (see Appendix B), e.g., Brown; Lancaster, Oslo, Bergen corpus (LOB); British National






Corpus (BNC); the Corpus of Contemporary American English (COCA), and the International Corpus of English (ICE), and provide valuable resources for information on how spoken and written language are used in a range of settings. However, in addition to corpora that represent written and spoken language in general, sometimes teachers need specialized corpora that represent a particular type of language use, such as EFL student compositions, university introductory chemistry lectures, lab reports, or business memos. Chapter 4 will provide some guidelines and ideas for how you can create your own specific corpora for classroom use.

How big is “large”?

In addition to being a principled collection of naturally occurring texts, another defining characteristic of a corpus is that it is a large collection of texts. However, *large* is an extremely relative term. As technology has advanced, corpus size has grown. In the 1960s, when some of the first electronic corpora were being built (e.g., LOB, Brown), one million words was considered large for a general corpus. Now, just over 40 years later, general corpora, the BNC for example, are often 100 million words, and COCA is over 400 million words! General corpora are often larger than specialized corpora since specialized corpora represent a smaller




slice of language. So, although the notion of size is rather fluid, it is important to realize that size is a reflection of the type of corpus (general or specialized) and the purpose of the corpus. Though earlier corpora may seem small by today's standards, they continue to be used. Studies have shown that one million words is sufficient to obtain reliable, generalizable results for many, though not all, research questions (Biber 1993; Reppen & Simpson 2002). A one-million-word general corpus will be adequate to address linguistic patterns of use and grammatical co-occurrence patterns, but not for lexical investigations. For lexical investigations corpora need to be very large to ensure that all the senses of a word are represented.


Why use a corpus with language learners?


Corpus-based investigations can identify linguistic and situational co-occurrence patterns. Most fluent speakers of a language have strong and fairly accurate intuitions about whether a form is grammatical or not. For example, if you hear someone say, *He don't like apples*; you know that the correct form is *doesn't* – *He doesn't like apples*. However, when asked to comment on patterns of use (e.g., Which verb tense is most frequent in conversation? What are the 10 most frequent verbs in conversation?), native speakers' intuitions are often ill-informed (Biber & Reppen 2002). Native speakers often notice the marked, or unusual, rather than the unmarked, or typical, uses of language. It is in this area that corpus linguistics can make the greatest contribution to language teaching. Since corpus linguistics can provide descriptions of actual language use, this information can then be used to shape and develop language-teaching materials, and even be used to develop language tests.

English as a Second Language/Foreign Language (ESL/EFL) professionals, from teachers to testing specialists, repeatedly make decisions about language, including which linguistic features and vocabulary to teach and/or test. In recent years, most ESL/EFL professionals have adopted a preference for “authentic” materials,





presenting language from natural texts rather than made-up examples (Byrd 1995; McDonough & Shaw 1993). Corpora provide a ready resource of natural, or authentic, texts for language learning. In addition to the preference for authentic texts, studies of second language learning have shown that when learners are engaged in meaningful activities (e.g., hands-on activities) that involve them in manipulating language, they learn more information and retain that information longer. Corpus activities directly address both of these areas by meaningfully engaging learners.






presenting language from natural texts rather than made-up examples (Byrd 1995; McDonough & Shaw 1993). Corpora provide a ready resource of natural, or authentic, texts for language learning. In addition to the preference for authentic texts, studies of second language learning have shown that when learners are engaged in meaningful activities (e.g., hands-on activities) that involve them in manipulating language, they learn more information and retain that information longer. Corpus activities directly address both of these areas by meaningfully engaging learners.





Chapter 3). In a reading class, for example, word lists can be used to identify words students will encounter in a reading. The teacher can then use the word lists to make certain the students control the vocabulary needed to read the text without too much difficulty.

Concordance programs create word lists that can be arranged in either alphabetical order or in order of word frequency (i.e., with the most frequent words appearing first). Knowing which words are infrequent in a text can also be important and give insight as to the specialized nature of the reading. Infrequent words most likely have specialized meanings that are specific to a particular area of study, which is especially true in scientific texts. Figure 1.1 shows a word frequency list from a subcorpus of 30 *New York Times* articles in the American National Corpus, generated using the concordance program MonoConc Pro 2.2 (Barlow 2002).



MonoConc Pro - [Corpus Frequency List]

File Concordance Frequency Display Window Info

Count	Pct	Word
1455	6.0491%	the
666	2.7689%	to
651	2.7065%	a
594	2.4695%	and
480	1.9956%	of
462	1.9208%	in
239	0.9936%	for
211	0.8772%	said
207	0.8606%	that
199	0.8273%	he
197	0.8190%	was
196	0.8149%	with
169	0.7026%	is
163	0.6777%	it
158	0.6569%	on
158	0.6569%	his
154	0.6403%	by
140	0.5820%	as
130	0.5405%	but
119	0.4947%	who
112	0.4656%	at
110	0.4573%	be
101	0.4199%	have
95	0.3950%	had
95	0.3950%	i
93	0.3866%	are
93	0.3866%	not
89	0.3700%	from
84	0.3492%	-
84	0.3492%	this
81	0.3368%	they

30 files in current corpus 24,053 words, 5,110 types

Figure 1.1: A word frequency list in MonoConc Pro 2.2, from approx. 25,000 words from *New York Times* articles in the American National Corpus.

Table 1.1: Frequency and alphabetical order word lists from 30 New York Times articles (approx. 25,000 words) created using MonoConc Pro 2.2

Frequency order		Alphabetical order	
frequency	word	frequency	word
1455	the	1	abandoned
666	to	1	abandonment
651	a	1	ability
594	and	3	able
480	of	1	abortion
462	in	56	about
239	for	2	absent
211	said	2	absolutely
207	that	1	absorbing
199	he	1	abundant
197	was	1	abusing
196	with	1	accept
169	is	1	accepted
163	it	1	access
158	on	1	accidental
158	his	1	accidents
154	by	1	acclimated
140	as	1	accolades
130	but	1	accompanied
119	who	8	according
112	at	2	accounting
110	be	1	accreditation
101	have	2	accused
95	had	2	accustomed
95	i	3	achieved
93	are	2	achievement

Table 1.1 shows the first 25 lines of the word list from this same small subcorpus of *New York Times* articles (approximately 25,000 words) in two different orders. The list on the left is in frequency order and the list on the right is in alphabetical order.

Your turn

Look at the word lists in Table 1.1 and think of two activities to do with students. Did you come up with any of the following ideas?

The information from the lists can be used as a starting point for several class activities:

- Discuss how the two lists are arranged (frequency vs. alphabetic). What are some of the differences in the types of words in the two lists?
- Find content words (i.e., nouns, adjectives, verbs, and adverbs) vs. function words (e.g., articles, pronouns, prepositions). Then answer the following questions: How many content words do you find in each list? How many function words? Why do you think there is a

difference between the two lists?

- Find related word forms (*abandoned, abandonment; achieved, achievement*) and examine the role of prefixes and suffixes. How do prefixes or suffixes change the core meaning of a word? How do prefixes or suffixes change a word from a noun to a verb or vice versa?
- Explore which words in the alphabetical list can go with words in the frequency list (e.g., *ability to*), or use the words in the two lists as the basis for a sentence scramble activity.
- Ask students to scan the lists and mark unfamiliar words. Then use those words as a basis for a vocabulary lesson.

Even from something as simple as a word list, several meaningful learning activities can be developed.

Using concordance lines

As a learner, knowing which words go together – and which words do not go together – is often a puzzle. Teachers can spend many classroom hours trying to provide students with meaningful input on which words can go together and on how certain words occur in some situations of language use and not in others (e.g., chatting with friends vs. writing

difference between the two lists?

- Find related word forms (*abandoned, abandonment; achieved, achievement*) and examine the role of prefixes and suffixes. How do prefixes or suffixes change the core meaning of a word? How do prefixes or suffixes change a word from a noun to a verb or vice versa?
- Explore which words in the alphabetical list can go with words in the frequency list (e.g., *ability to*), or use the words in the two lists as the basis for a sentence scramble activity.
- Ask students to scan the lists and mark unfamiliar words. Then use those words as a basis for a vocabulary lesson.

Even from something as simple as a word list, several meaningful learning activities can be developed.

Using concordance lines

As a learner, knowing which words go together – and which words do not go together – is often a puzzle. Teachers can spend many classroom hours trying to provide students with meaningful input on which words can go together and on how certain words occur in some situations of language use and not in others (e.g., chatting with friends vs. writing

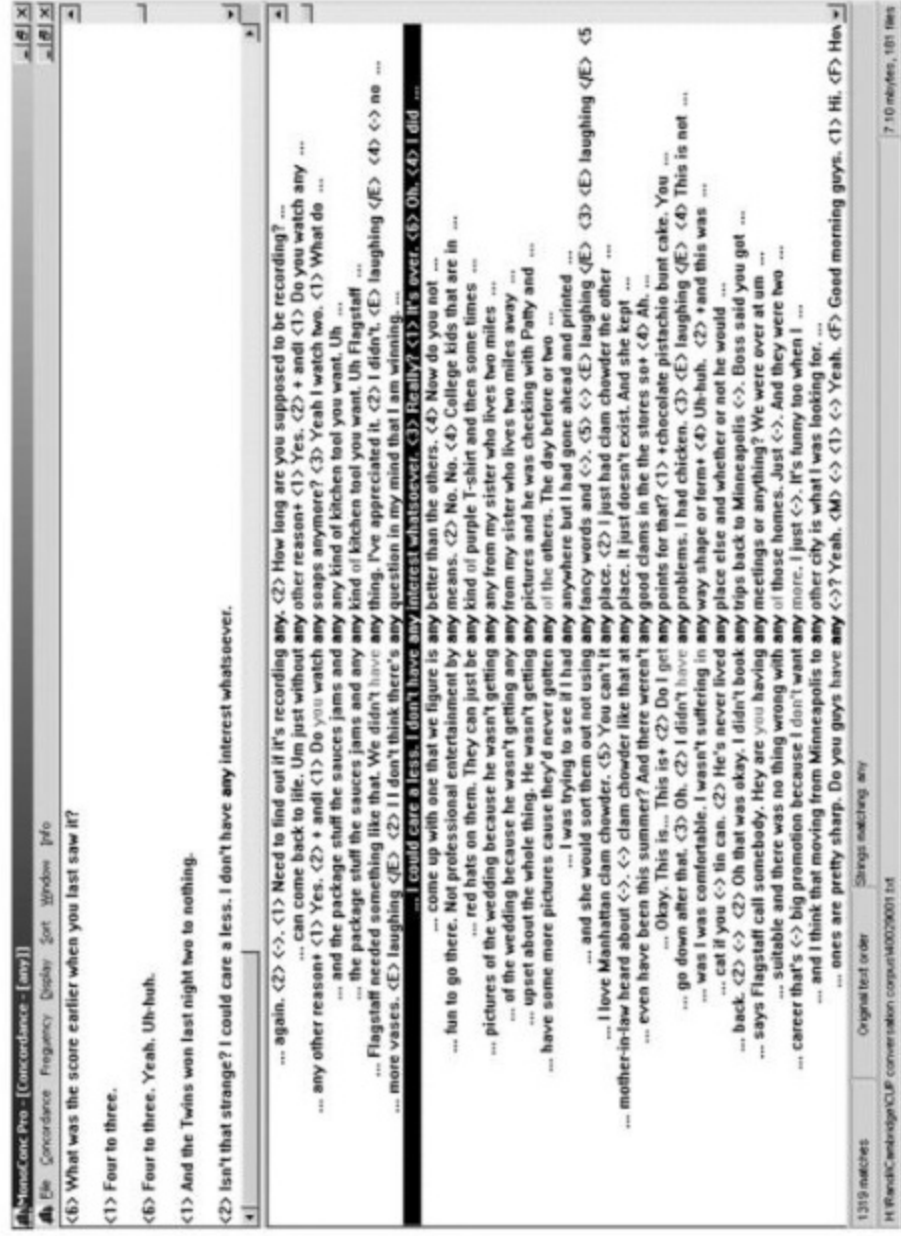


Figure 1.2: KWICs of the target word any in MonoConc Pro 2.2, from a corpus of spoken conversation.



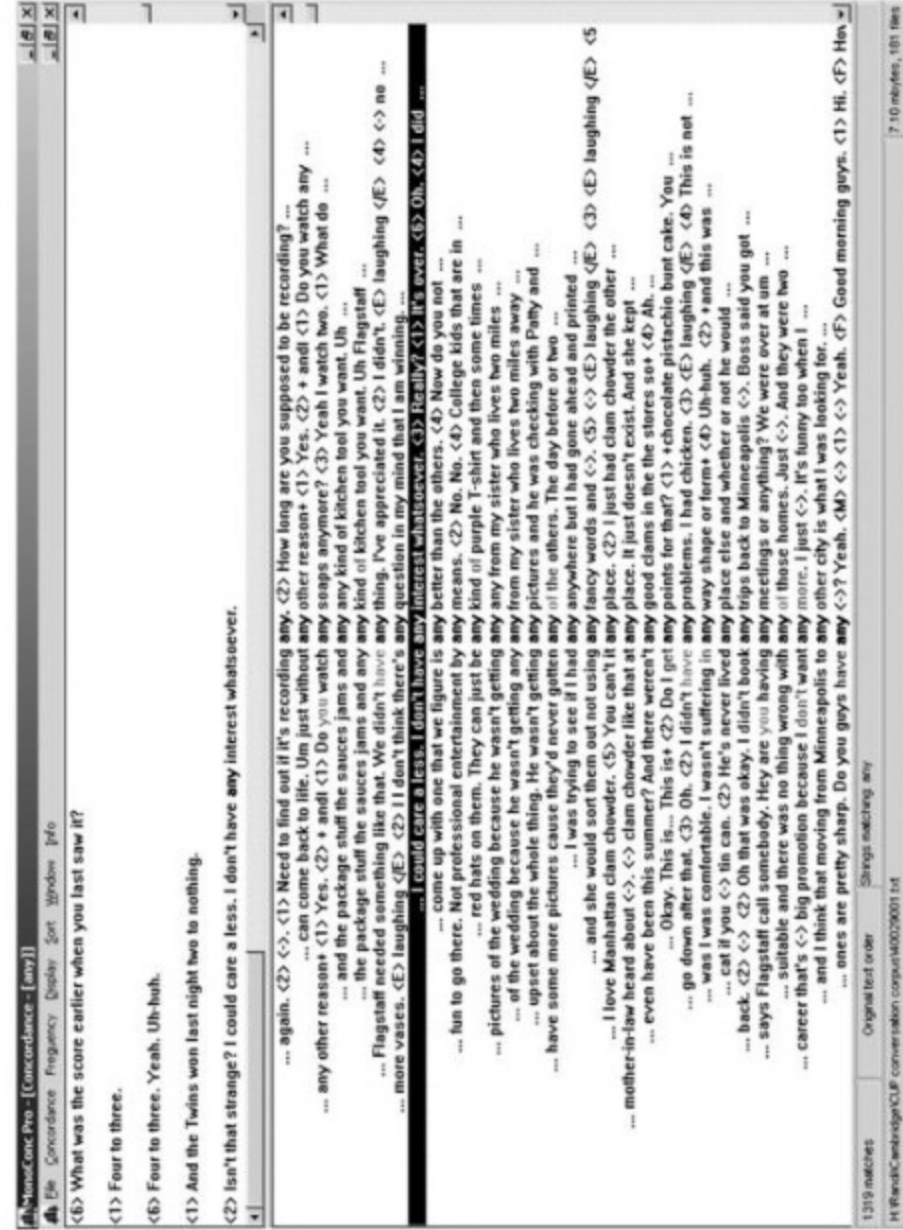


Figure 1.2: KWICs of the target word any in MonoConc Pro 2.2, from a corpus of spoken conversation.



The Sample Activity below is an example of how a teacher used information from corpus research and provided students with KWICs to help guide the students in learning the more frequent uses of the word *any*.

A corpus study by Mindt (1998) concluded that 50 percent of *any* use is in affirmative statements, 40 percent in negative statements, and only 10 percent in interrogatives. The exercise below uses 10 representative corpus examples. The purpose of this exercise is to get the students to discover use patterns and their relative frequency.

Sample Activity¹

The word *any* is often taught in the following way:

Interrogatives: Are there *any* Turkish students in your class?

Negatives: No, there aren't *any* Turkish students in my class.

Affirmatives: Yes, there are *any** Turkish students in my class.

* Not grammatical

Part 1

Read through the following lines taken from a concordance of the word *any*.

concordance lines above can be exploited for other purposes as well, such as defining some of the functions of *any* and common language chunks that use *any*.

Your turn

Make a list of other words that you think might lend themselves to this type of guided discovery activity. Save your list so that you can use it to create some activities in Chapter 4.

OTHER WAYS TO USE CONCORDANCE LINES

Using concordance lines, teachers and researchers can also explore co-occurrence patterns. Certain lexical patterns and some basic grammatical patterns can be discovered through concordancing. Concordance output can also provide learners with information about word use and how the same word can sometimes have more than one meaning, as will be explored in several of the activities presented in later chapters.

Using tagged texts

One aspect that adds complexity to teaching a language is the fact that words can have different grammatical roles. For example the word *can* might be a noun referring to a can of soda, or it might be a modal expressing ability/ possibility: *I can go with you.* Even in this case, a corpus can be a valuable resource for teachers. Some corpora are *tagged*, that is each word in the corpus is labeled for its part of speech or grammatical category (e.g., first person pronoun, verb, relative clause). The process of creating a tagged corpus is quite sophisticated and is beyond the scope of this book as it requires computer programming skills. Although developing and processing a corpus for tagging requires sophisticated skills, using tagged corpora does not require sophisticated programming skills and can provide useful information to the user.


Text Sample 1 on p. 12 is an excerpt from a tagged text. The sentence, “I’m gonna ask you to pay sixty or seventy dollars for a textbook.” has been tagged using a tagger developed by Doug Biber (for more information see Biber 1988; Conrad and Biber 2001). The labels in the right-hand column do not appear in the program but were added here to make the tag codes clearer. Although this particular example uses Biber’s tagger, the codes are similar to those found in other tagging programs.

Appendix B lists some resources for tagging texts.

Text sample 1: Example of tagged text

Word	Tag	Label
I	^pp1a+pp1+++	first person pronoun
'm	^vb+bem+aux++0	BE verb, auxiliary, contracted form
gonna	^md+prd+++0	modal, contracted form
ask	^vb++++	base verb
you	^pp2+pp2+++	second person pronoun
to	^to++++	to clause
pay	^vbi++++	infinitive form of verb
sixty	^cd++++	cardinal number
or	^cc++++	coordinating conjunction
seventy	^cd++++	cardinal number
dollars	^nns++++	noun – plural
for	^in++++	preposition
a	^at++++	article
textbook	^nn++++	noun – singular
.	^.+clp+++	end punctuation


By using texts that have been tagged, teachers can answer a range of different types of questions. For example, we can see what grammatical



structures tend to co-occur, such as which verbs most frequently take complement clauses (e.g., I think that . . . , I know that . . .). Tagged texts are also very useful when dealing with words that have multiple functions, such as *well*, which can be a noun, verb, adjective, adverb, interjection, or discourse marker. Teachers and students can use tagged texts to search and create lists of nouns or prepositions that occur in a text and then look at these in context to get a clear sense of the meaning. In Chapter 3 we will explore some of the ways that tagged texts can be used in the classroom.

The role of register

Language teachers are often concerned with the different contexts of language use and helping learners understand how language can vary depending on these different contexts. Corpora can be a valuable teacher resource for creating activities that help learners understand variation due to situational factors, or register. *Register* is a term used to describe varieties of texts that are defined by situational characteristics (e.g., spoken vs. written, edited vs. real-time). The term *register* can be used at various levels of specificity. For example, spoken language vs. written language would be a broad register distinction. The register of spoken language could be further subdivided into the registers of face-to-face



conversations and phone conversations; the subregister of phone conversations could then be further divided into business vs. personal phone conversations, as illustrated in Figure 1.3.

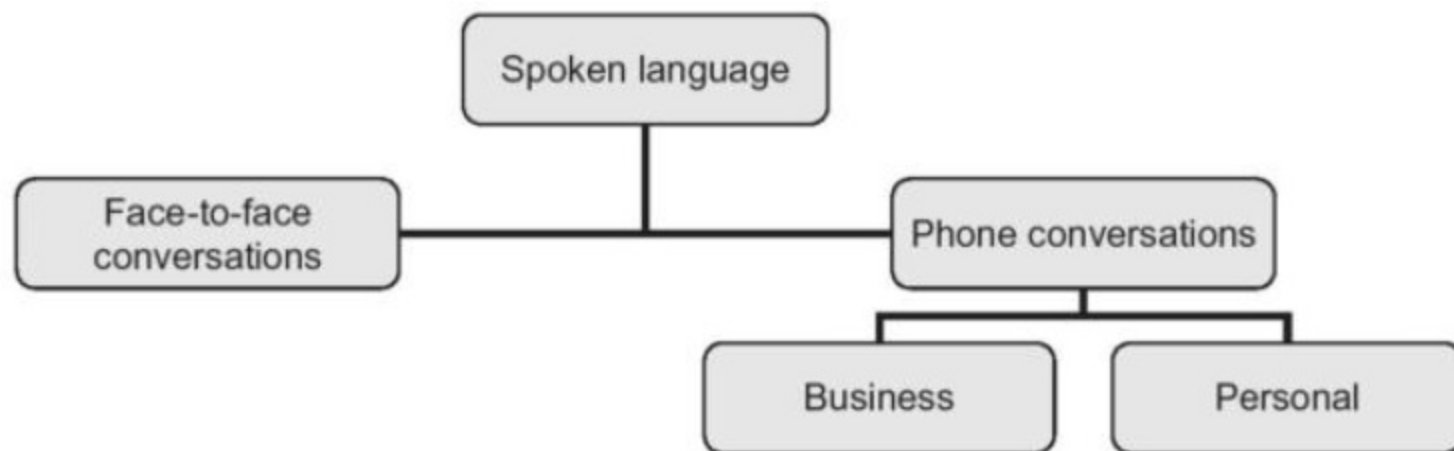


Figure 1.3: Some of the subregisters of spoken language.

The situational characteristics of language use, or register, has a strong impact on the linguistic features associated with that register. Each register has its own unique linguistic patterns. For example, one of the features associated with face-to-face conversations is repetition (e.g., “I bet I . . . I bet I, I put an s on it”). This type of repetition is likely due to the speaker producing spoken language under real-time constraints. This type of repetition is seldom found in written language since writers

usually have time to produce language and do not need to “buy time” to think of a word while maintaining their position as a speaker in the conversation.

When using a corpus for instructional purposes it is important that the corpus represent the register that is being taught. For example, we would not want to use a corpus of informal conversation as a basis for academic writing instruction, or vice versa.

Your turn

Think about the classes that you plan to teach. Make a list of each register that you are planning to teach your students (e.g., business English, conversational English, medical English). Now make another list of the different situations that can occur within those registers (e.g., phone conversations, memo writing, note taking). What kinds of authentic texts, or corpora, would you want available to teach your classes? Keep this list, so that in Chapter 4 you can create materials based on your needs.

What will corpus-based materials look like?

Some activities that are corpus-informed, such as those using concordance lines, might look different than traditional textbook exercises; however many of the activities and formats look much like non-corpus-informed activities. It is the content that will differ. The order of information presentation may also differ. For example, in a beginning conversation class the verbs that are introduced in corpus-informed material may contain irregular verbs because we know from corpus research that the verbs, *think*, *have*, *go*, and *come*, are extremely frequent in conversation. Therefore, in corpus-informed materials, these verbs would be presented early on, unlike the more traditional approach where irregular verbs are kept for later units. Although the activities shown in Figure 1.4 may look similar to those found in many ESL/EFL textbooks, the content being presented is not typical since it introduces an irregular verb at a very early stage, in the third unit in this case. These activities also reflect some of the many different meanings of the verb *have*.

The example in Figure 1.4 is a nice example of how corpus linguistics can be used to inform instruction and materials development, and yet retain the familiar appearance that learners and teachers are accustomed to, thus not scaring anyone away from the material being presented.

What will corpus-based materials look like?

Some activities that are corpus-informed, such as those using concordance lines, might look different than traditional textbook exercises; however many of the activities and formats look much like non-corpus-informed activities. It is the content that will differ. The order of information presentation may also differ. For example, in a beginning conversation class the verbs that are introduced in corpus-informed material may contain irregular verbs because we know from corpus research that the verbs, *think*, *have*, *go*, and *come*, are extremely frequent in conversation. Therefore, in corpus-informed materials, these verbs would be presented early on, unlike the more traditional approach where irregular verbs are kept for later units. Although the activities shown in Figure 1.4 may look similar to those found in many ESL/EFL textbooks, the content being presented is not typical since it introduces an irregular verb at a very early stage, in the third unit in this case. These activities also reflect some of the many different meanings of the verb *have*.

The example in Figure 1.4 is a nice example of how corpus linguistics can be used to inform instruction and materials development, and yet retain the familiar appearance that learners and teachers are accustomed to, thus not scaring anyone away from the material being presented.

often incorrect.

In addition to these types of materials, there are now several ESL/EFL textbooks that use the Academic Word List (Coxhead 2000) to provide teachers and students with ready-made, corpus-informed material for vocabulary instruction.

Exercises

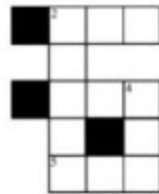
3.1 Fill in the blanks. Use words from B, C, and D on page 6.

1. I never have a big *breakfast* in the morning.
2. I have an with the doctor at 1 o'clock.
3. I had a yesterday, so I studied all night.
4. Mike is having a on Saturday night. Are you going?
5. I'm too busy, so I don't have to take a vacation.
6. I have a terrible I keep sneezing. Atchoo!
7. I had a with Maria last night. We went out to dinner and a movie.
8. Keiko is going to have a She thinks it'll be a girl.

3.2 Answer these questions about *yourself*.

1. Do you have any brothers or sisters? If yes, how many?
2. Which days do you have classes?
3. What do you usually have for lunch?
4. On weekends do you have to get up early in the morning?
5. Do you have coffee or tea with breakfast?
6. Is there anything you don't have at home that you want to have? What is it?
7. Do you ever have trouble understanding English? When?
8. Do you have to study hard to learn English? Why?

3.3 Do the crossword puzzle.



Across

1. You can have one in a restaurant.
3. Some people like to have one on New Year's Eve.
5. You have this between meals.

Down

2. You have these at school.
4. If you don't want coffee, you can have

3.4 Complete the sentences using *have*.

1. A: I'm thirsty!
B: Why don't you ?
2. A: I feel sick today.
B: Do you ?
3. A: Bye, everyone! I'm leaving for my trip to Hawaii!
B: Bye !
4. A: Can you come to my party on Friday?
B: I can't, I have a big test on Monday and I this weekend.

Figure 1.4: An example of a corpus-informed vocabulary book. The corpus was used to determine the order of the information presented in the vocabulary units. (*Basic Vocabulary in Use*, McCarthy & O'Dell 2010, p. 7)

Exercises

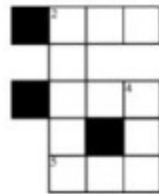
3.1 Fill in the blanks. Use words from B, C, and D on page 6.

1. I never have a big *breakfast* in the morning.
2. I have an with the doctor at 1 o'clock.
3. I had a yesterday, so I studied all night.
4. Mike is having a on Saturday night. Are you going?
5. I'm too busy, so I don't have to take a vacation.
6. I have a terrible I keep sneezing. Atchoo!
7. I had a with Maria last night. We went out to dinner and a movie.
8. Keiko is going to have a She thinks it'll be a girl.

3.2 Answer these questions about *yourself*.

1. Do you have any brothers or sisters? If yes, how many?
2. Which days do you have classes?
3. What do you usually have for lunch?
4. On weekends do you have to get up early in the morning?
5. Do you have coffee or tea with breakfast?
6. Is there anything you don't have at home that you want to have? What is it?
7. Do you ever have trouble understanding English? When?
8. Do you have to study hard to learn English? Why?

3.3 Do the crossword puzzle.



Across

1. You can have one in a restaurant.
3. Some people like to have one on New Year's Eve.
5. You have this between meals.

Down

2. You have these at school.
4. If you don't want coffee, you can have

3.4 Complete the sentences using *have*.

1. A: I'm thirsty!
B: Why don't you ?
2. A: I feel sick today.
B: Do you ?
3. A: Bye, everyone! I'm leaving for my trip to Hawaii!
B: Bye !
4. A: Can you come to my party on Friday?
B: I can't, I have a big test on Monday and I this weekend.

Figure 1.4: An example of a corpus-informed vocabulary book. The corpus was used to determine the order of the information presented in the vocabulary units. (*Basic Vocabulary in Use*, McCarthy & O'Dell 2010, p. 7)

3 Building language

Unit 2 Interests

A Listen. What does Carla think of the band? Practice the conversation.

Alex Listen. What do you think of this song?

Carla It's good - I like it. Who is it?

Alex A new band . . . some local guys. Do you like them?

Carla They're local? Really? They're pretty good. Who's the lead singer? I like her. She sounds like Mariah Carey.

Alex Yeah, everybody says that. It's my friend Lori.

Carla Who's the guy singing with her? I'm not sure about him.

Alex Uh . . . actually, that's me. I'm in the band, too.

Figure it out

B Complete the conversations. Then ask a partner your questions. Give your own answers.

1 A What do you think of _____ (male singer)?

B I like _____.

2 A Do you know _____ (female singer)?

B No, I don't know _____.

4 Grammar Object pronouns; everybody, nobody

I'm a singer. That's **me** on the CD.

You're a musician? I'd like to hear **you**.

She's pretty good. I like **her**.

He's not a good singer. I don't like **him**.

It's a nice song. I like **it**.

We play in a band. Come listen to **us**.

They're local guys. Do you like **them**?

Everybody

Everyone

Nobody

No one

likes pop.

A Complete the questions with object pronouns. Complete the answers with *everybody* or *nobody*. Then practice with a partner.

1. A I listen to hip-hop a lot. Do you listen to it, too?

B Yes. Everybody in my school listens to hip-hop.

2. A Ricky Martin - he was on TV last night. Do you know _____?

B Of course I do. _____ knows Ricky Martin.

3. A I don't really like classical music. Do you ever listen to _____?

B Yes, but with earphones - _____ in my family likes it.

4. A I like Alicia Keys. She's a good singer. What do you think of _____?

B Oh, almost _____ is a fan of Alicia Keys.

5. A My favorite band is Coldplay. I think they're great. Do you like _____?

B Yeah. They're the best. _____ plays rock like they do.

In conversation . . .

Everybody and nobody are more common than **everyone and no one**.

_____ **everybody**

_____ **everyone**

_____ **nobody**

_____ **no one**

Figure 1.5: An example of corpus-based information presented in a format similar to traditional course books (Touchstone Level 2, McCarthy, McCarten, & Sandiford 2004, p. 15).

Putting it all together

As we see in this chapter, information from corpus linguistics can be used to inform language teaching in several ways ranging from helping teachers decide course content or informing the order of presenting information to having learners actually interact with corpora. Corpora that represent the types of language, or registers, that are being taught can be a useful tool for teachers and students. And equally important, although the area of corpus linguistics is complex, the task of using corpora or information from corpus research in the language classroom does not have to be overwhelming. The next three chapters will provide information and detailed examples of how to use information from corpus research to inform teaching decisions and step-by-step instructions for how to use corpus resources for language instruction.

Additional reading

- Aston, G., Bernardini, S., & Stewart, D. (2005). *Corpora and language learners*. Amsterdam: John Benjamins.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Exploring language structure and use*. Cambridge: Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* 34:213–38.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (ed.), *Learning with corpora* (pp. 108–37). Houston, TX: Athelstan.
- Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal* 55:238–46.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Meyer, C. (2002). *English corpus linguistics*. Cambridge: Cambridge University

Press.

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.

Reppen, R., & Simpson, R. (2002). Corpus linguistics. In N. Schmitt (ed.), *An introduction to applied linguistics* (pp. 92–111). London: Arnold.

Sinclair, J. (Ed.). (2004). *How to use corpora in language teaching*. Amsterdam: John Benjamins.