

Language teachers with corpora in mind: from starting steps to walking tall

Angela Chambers*, Fiona Farr and Stephanie O’Riordan

School of Languages, Literature, Culture and Communication, University of Limerick,

Limerick, Ireland

Abstract

Although the use of corpus data in language learning is a steadily growing research area, direct access to corpora by teachers and learners and the use of the data in the classroom are developing slowly. This paper explores how teachers can integrate corpus approaches in their practice. After situating the topic in relation to current research and practice in ICT and language learning, we examine some easily available resources, suggesting how they can provide examples of naturally occurring discourse for use in the language classroom. Beginning with easily available online corpus resources with built-in concordancers which require no prior technical training, we continue by discussing the challenges which more advanced use of corpora presents. This is illustrated by examples from corpora of business communication and casual conversation. We conclude by emphasising that there is considerable scope for both research and dissemination in relation to corpus applications in language learning.

Introduction

The use of corpus data in language learning materials is not a new phenomenon. While early reference corpora of the 1960s and 1970s such as the Brown Corpus of American English and the LOB (Lancaster-Oslo-Bergen) Corpus of British English, usually referred to by its acronym, were seen primarily as resources for linguistic research, later corpora of hundreds of millions of words, such as the BNC (British National Corpus) and the Bank of English have also been used as the basis for language learning resources. While English language researchers initially led the field in corpus linguistics, corpora have now been created in many languages, such as the Mannheim corpora in German and the Corpus del Espanol in Spanish. (A list of web links for several online corpora referred to in this study is provided at the end of the article.) Developments such as these provide resources based not on a standard ‘correct’ language determined by written norms as perceived by the authors of the resources, but rather on attested examples of what is typically written or said. The Cambridge International Corpus, for example, contains approximately one billion words of British and American English, including written texts such as newspaper articles, novels, business and legal texts, web sites, language learners’ examination scripts, academic books and journals, and spoken texts such as everyday conversation and business discourse. This is used as a resource for the authors of language learning text books, dictionaries and, more recently, grammars (Carter and McCarthy 2006) published by Cambridge University Press. For

commercial reasons, this corpus is not publicly available. Examples of corpus-based resources in other languages include the Oxford Hachette French–English Dictionary and the Frequency Dictionary of Czech (Frekvenci´ Slovník Cestiny).

From the mid 1980s, publications on the use of corpus data directly by individual teachers/researchers increased, with Johns (1986) and Tribble and Jones (1990) among the first to emphasise the potential of concordance data as the basis of language learning materials. Despite these positive developments, however, we are a long way from a situation where corpus data are commonly consulted by language teachers and learners. Romer (2006: 121) points out that English language teaching practice and EFL initial teacher training are still largely unaffected by developments in corpus linguistics. Indeed, researchers involved in integrating corpora into their language teaching and into language teacher education programmes (O’Sullivan and Chambers 2006: 62–3; Farr 2008) discover not only positive reactions from learners and student teachers, such as access to real and up-to-date language use, and uncovering functions and structures not found in grammar books, but also negative reactions or obstacles, mainly relating to the effort needed to master the technology and to analyse the results, and the difficulty in gaining access to appropriate corpora.

It is these obstacles which have motivated us to choose this topic, to explore the possibilities open to teachers interested in integrating corpus data in their practice. Our aim is to answer the following questions:

- (1) What can we learn from current research and practice in the use of corpus data in language learning and teaching?
- (2) How can a teacher introduce corpus data into the classroom without prior training in concordancing software?
- (3) What challenges does a more advanced use of corpus data present for teachers and learners?

These discussions will be illustrated by a number of practical examples from corpora of business communication and casual conversation. Both the research publications quoted and the practical examples involve the teaching of modern languages, mostly but not exclusively at advanced level (Kennedy and Miceli’s [2001] learners are at intermediate level). The article is intended to be of interest to language teachers, in particular to teachers who see themselves also as action researchers, using their teaching as the basis of their research.

Research and practice in corpora and language learning

Teachers using technology

The use of corpus data by language learners and teachers can be related to second language acquisition research in a number of ways. Farr, Chambers and O’Riordan (2010), for example, develop principles for integrating corpora based on research in language learning materials development. Research into ICT and teacher education, which is also relevant in this context, reveals that ensuring the transfer of innovative methods from LTE to the classroom is a challenging task. Surveying the literature on this topic, Egbert, Paulus and Nakamichi (2002: 111) conclude that in general ‘teachers are using technology in ways that fit their current practice, rather than transforming their practice through the use of technology’. They do, however, accept, citing Cuban (1986: 70), that ‘teachers will alter classroom behaviour selectively to the degree that certain technologies help them solve problems they define as important’. The applications which Egbert et al. list as the most commonly used are nonetheless limited, namely word processing, drill and practice exercises, and the Internet. The introduction of corpus consultation, requiring considerable investment of time for both training and corpus analysis, is thus a challenging task. Teachers need to become aware of its benefits before they invest time and effort. For this reason we propose a graded approach, beginning with activities which require only basic Internet skills, and moving on to more complex applications only when the simpler activities have shown the contribution which corpus data can make to the language learning environment.

Empirical studies of corpora in language learning: why use corpus data?

Since the early 1990s researchers have published accounts of their use of corpus data with learners, using either large reference corpora which are publicly available, such as the BNC (Bernardini 2000) or, more commonly, smaller corpora of either familiar texts such as newspaper articles or of genres which the learners wish to master, such as academic writing (Johns 1997), thus ensuring that the texts would be appropriate for the learners’ needs, would have an impact on them through relevance to their other work, and would be perceived as authentic. Like Johns, the researchers often created these small corpora specifically to use with their learners. Indeed, Braun (2005), underlining the difference between the needs of linguistic researchers for vast corpora and the specific needs of learners, emphasises the need for pedagogically relevant corpora to be created.

Researchers justify their use of corpora with learners in a number of ways; several emphasise the potential for developing strategies for language learning. Corpus consultation encourages an inductive, problem-solving approach (Kennedy and Miceli 2001: 88) and discovery learning: ‘Every learner a Sherlock Holmes’ (Johns 1997: 101). It also encourages reflection on the part of the learner, increasing language awareness (Chambers 2005). It changes the role of the teacher from a transmitter of knowledge to a facilitator of student learning

(Kennedy and Miceli 2001). It enables learners to learn more about the target language by providing a large number of relevant attested examples of the use of a word or phrase in an instant, compensating to a certain extent for the fact that the learner would otherwise not have time to read or interact verbally enough to encounter all these examples (Gaskell and Cobb 2004: 304). McEnery and Wilson (1997: 6) summarise the pedagogic benefits of using corpora in four ways. It is a means of encouraging discovery learning and also divergent learning, as different learners will 'take different paths through the data and find slightly different things'. It also involves mediated learning, as the learner must interact with the data, rather than learning from their contents. Finally, it involves learning directed or facilitated, but not led by the teacher.

Underpinning the use of multiple examples of attested language use is the conviction that learners do not only acquire language skills by applying the open choice principle (Sinclair 1991: 109–10), assembling nouns, verbs, prepositions, etc, in grammatically acceptable ways, but also by the idiom principle learning to use 'a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments'. Finally, corpus data provide multiple examples in context, what Cobb (1997: 303) terms 'multicontextual learning'. It is important to note that researchers in applied linguistics are not unanimous on this issue. While many of the researchers who report on their use of concordance data with learners appreciate that a concordance provides examples in context, Widdowson (2004: 71) points out that the words accompanying the search word are co-text, and that context is not provided by massed concordance lines. The researchers' choice of specially created corpora of a genre familiar to the learners, and their experimentation with learners at advanced level, are no doubt intended to create a situation where the learners are able to appreciate the context of the concordance lines. An example taken from a familiar genre often included in language courses, the business letter, will serve to illustrate this in a later section. Moreover, as Braun (2005: 59) illustrates, a corpus can be used to select a complete text on which to base a class, with the concordance lines introduced not as an end in themselves, but to provide supplementary examples. It is this use of concordances which is implied in the present article. In other words, corpus-based activities can exist alongside non corpus-based activities, introduced when the provision of multiple examples, perhaps as the basis for an inductive exercise, seems appropriate to the teacher.

It is difficult to come to firm conclusions on the effectiveness of the use of corpora in language learning, as the empirical studies are mostly qualitative in nature, investigating what the students feel they have learned and their reactions to the process. Reactions here are generally very positive, with some reservations concerning the time-consuming activity of analysing the data and the amount of training required (Gaskell and Cobb 2004: 315; Yoon and Hirvela 2004: 2741). Very few studies attempt to quantify the learning resulting from consulting concordances as opposed to more traditional methods. In a rare study comparing the study of concordances with more traditional gap-filling exercises, Stevens (1991)

concluded that students performed better in the learning of vocabulary when studying concordances. The qualitative studies are, however, important in that they support both the linguistic and pedagogic benefits outlined below. The learners appreciate the abundance and relevance of examples (Bernardini 2002: 179; Chambers 2005: 120; Yoon and Hirvela 2004: 275), the authenticity of the data (Chambers 2005:120), and the self-directed nature of the activity (Bernardini 2002: 179).

In summary, the benefits of using corpus data rather than the small number of what are commonly invented examples in a course book are both linguistic and pedagogic. In the linguistic context, teachers can be confident that the language which they are teaching conforms to actual language use. In other words, corpus informed pedagogy is essentially evidence-based. Teenage learners, for example, could have access to the SACODEYL corpora, pedagogic corpora consisting of online interviews with teenagers in seven European languages, watching the videos and using the built-in concordancer on the site to discover if an expression they observe is also used by the other teenagers. In the pedagogic context, McEnery and Wilson (1997), citing Fligelstone (1993), divide the use of corpora in language learning into three areas: teaching about corpora; teaching to exploit; and exploiting to teach. In the context of the first of these areas, they underline the importance of raising awareness of the methodology of corpus use (1997: 7). Teaching to exploit corpus data involves training learners in the use of concordancing software, which in turn makes them learners who 'know what it means to use a corpus, [to] know how to extract material and follow their own hunches in a corpus, and [to], consequently, learn a great deal about language through learning via a corpus' (1997: 7). They also point out, however, that a teacher can simply use printed concordances as the basis for classroom exercises, exploiting to teach while skipping the first two stages. While printed concordances will no doubt still be a useful first stage for many teachers, developments since 1997 make another developmental path possible, namely beginning with simple online resources with built-in concordancers, before moving on to Fligelstone's three stages.

Some easily accessible corpora and their potential uses: what can be done with corpora without extensive training?

The basic definition of a corpus is easy for a teacher to understand, notwithstanding various controversies among researchers concerning authenticity, representativity and sampling. A corpus is a collection of machine-readable, authentic texts, sampled to be representative of a particular language or language variety (McEnery, Xiao and Tono 2006: 5). However, corpora can be classified in an increasingly large number of ways. They may be large reference corpora containing a wide variety of written and spoken text types, or small corpora containing texts of one genre or a limited number of genres. They may be monolingual, or bi- or multilingual (either parallel corpora of texts and their translations, or comparable corpora of, for example, academic articles in two or more languages). They may include only texts by

native speakers, or by expert non-native speakers, or texts produced by language learners. They may exist only in written form, including transcriptions of spoken discourse, or they may be multimodal, with audio- and video-recordings. They may be annotated or tagged in various ways to assist the user, or they may be untagged. Many corpora are tagged to identify different parts of speech. More recently, corpora created for language learning, pedagogic corpora, may be annotated in relation to the themes they deal with, or the ways in which the teacher could use them. (See, for example, the SACODEYL corpora.) A frustrating aspect of wanting to use corpora in language teaching is thus the confusion a teacher might feel while searching for a readily available corpus to use. More detailed information is provided later in this article on the main types of corpora likely to be of interest to teachers at an advanced stage, namely written and spoken, general and specific, small and large, and learner corpora. As this section focuses on the first steps a teacher can take, the focus is on those corpora, whatever their size and category, which can be easily accessed online and used to extract examples of naturally occurring discourse which can be used in language learning materials.

Fortunately, there is a simple solution to the problem of the confusion a teacher might experience when first searching for and using corpus data, in that a number of corpora are freely accessible via the Web and have simple built-in concordancers, which make it possible for a teacher with basic Web skills to produce teaching materials without any technical training. Indeed, WebCorp, which allows the Web itself to be used as a corpus, is becoming increasingly popular (Renouf, Kehoe and Banerjee 2007). While huge language corpora on which reference publications such as grammars and dictionaries are based are usually not available free of charge for obvious commercial reasons (The British National Corpus is a notable exception), in many cases sample data are made freely available to the public. The Bank of English, on which the Cobuild dictionaries are based, allows access to a sampler version which can be queried without any prior training. The simplest introduction for a teacher experimenting with a corpus for the first time would thus be to access one of the online concordancers linked to large corpora, such as the Cobuild Concordance and Collocations Sampler, which provides 40 concordance lines as the result of a simple search, or the BNC, which provides 50. To illustrate how a teacher can make use of this, examples of uses of *nice* in the BNC are listed below:

Extract 1: Concordances of the word nice from the BNC

- 1: So you could make a really nice outfit
- 2: She did not think the masks were nice toys for young children.
- 3: You were allowed so few clothes that it was nice to be able to wear something different.
- 4: Kate had envied her mother her nice clean life.
- 5: 'I must say,' said Constance, 'it's nice to see a lot of animals sitting peacefully in their kennels like a lot of monks.

6: ‘Well,’ he said, ‘we’ve got a nice, juicy chicken leg for you!’

7: Can you imagine Roy’s face when he comes home to everything nice and clean and homely?

8: I do a lot of speaking to various organisations, in a variety of halls, some nice and warm, some freezing cold

The examples of the use of nice and, for example, as in ‘nice and clean’ (Line 7) and ‘nice and warm’ (Line 8), could provide additional attested examples for a text using this expression, as well as giving the learners the opportunity to see other examples of the uses of the adjective. Indeed, the various instances of nice in the concordance lines bring to the fore the polysemic aspect of this word, which can mean pretty or lovely (Line 1), adequate or appropriate (Line 2), tasty (Line 6) and so on and so forth. One can easily imagine an ESL teacher asking learners to infer such synonyms of nice from the immediate context provided in the concordance lines. Resources such as this make it easy for the teacher to produce materials which can be used in a data-driven learning approach, in which the learners inductively work out the different ways in which an item can be used. As previously stated, the intention is not to suggest whole classes based only on concordance lines, but rather a situation where a concordance could enrich the encounter of one example of an expression in an individual text. For an additional example, see Appendix 1.

Other freely available and easy-to-use resources include small monolingual native-speaker corpora of texts of one genre available online, such as ELISA, a pedagogic corpus of recordings of individuals from various English-speaking countries talking about their lives and careers, or the one-million-word Business Letter Corpus of British and American English, both with built-in concordancers requiring no prior training. Corpora in languages other than English are now increasing in number, and useful resources for language teachers wishing to experiment include one-million-word corpora of journalistic (Chambers–Rostand) and academic French (Chambers–Le Baron), both freely available but which would have to be loaded onto a concordancer to produce language learning materials. Of particular interest for secondary school teachers is the SACODEYL corpus of teen talk, which includes video-recorded interviews and transcripts of approximately 10 minutes with 20–25 teenagers in each of the following languages: English, French, German, Italian, Lithuanian, Romanian, and Spanish. This is a pedagogic corpus, including on the site both a simple concordancer and other pedagogic material for teachers. All of these corpora are freely available via the web.

One of the more substantial resources in this category is Cobb’s ‘Compleat Lexical Tutor’, also known as Lextutor, which provides easy access to a number of corpora in English and French, and a simple built-in concordancer. The site has been consciously designed to aid

teachers in finding out information which can only be retrieved with the help of corpus tools such as concordancing or frequency list software, to name the two most common types of language queries performed on corpora. The most attractive feature of such a web site is that it presents information which can easily be understood by 'lay' corpus users. This freely accessible web site with a non-threatening interface has proved very popular with student teachers in the University of Limerick, Ireland. Clearly inspired by the data-driven learning approach to language teaching and learning first introduced by Tim Johns (1986), Lextutor can serve many purposes. The one which teachers may find most attractive is that of providing countless illustrations of the actual use of particular language features which may pose problems for learners. A common example is that of the use and meaning of prepositions in English. *Get back*, for example, has a number of distinct meanings, which are illustrated in the examples below, taken from Lextutor using the BNC:

Extract 2: Concordances of get back from Lextutor using the BNC

- 1: Sharpen up in a big way tomorrow and get back to winning ways after a draw
- 2: Are you tryin to get back at your dad because of this stupid
- 3: By that time we were really longing to get back to England.
- 4: But we must get back to the Vicar, who is awaiting his final

As in the case of Johns (1994), a teacher could use concordances such as this to provide additional examples if the expression had proved problematic in a text which was being studied, or simply to point out to the learners that the expression does not always have the meaning in the text being studied. Consistent use of web pages such as Lextutor allows language educators to complement their intuitive knowledge of the target language whether it be their first, second or other language. A recent example illustrating this point is that of a student teacher trying to say: 'students will be confused' and aware that 'Les e' le`ves seront *confuse' s' (sentence uttered by a student teacher during a feedback session in 2007) did not sound right in French. Using Lextutor as a source of examples the teacher educator could provide students with examples of how they could express themselves more effectively using the expressions below:

Extract 3: Concordances of the French word confusion from the 1 million + corpus of articles from Le Monde (1998) available on Lextutor

- 1: Mais des rapprochements mal-maîtrise' s entraînent une confusion nocive pour les re'alisations . . . [But the badly-controlled tight schedules lead to a confusion which

is harmful to film-making.]

2: accroître les divisions au sein de la majorité plurielle et de **semer la confusion** dans l'opinion publique. [. . . to increase divisions within the multi-party majority and to confuse public opinion.]

3: Ce terme d' " euthanasie passive " **seme la confusion** : décider collectivement de . . . [This term 'passive euthanasia' is confusing: to decide collectively to . . .]

Language Learning Journal 7

4: Ces faits contradictoires **entretiennent une confusion** qui, à son tour, nourrit des jugements . . . [These contradictory facts maintain a state of confusion which, in turn, feeds into judgements . . .]

5: Les parties civiles sont intervenues pour **entretenir la confusion** en faisant le procès de la fonction publique. [The plaintiffs intervened in order to maintain a state of confusion by putting the civil service in the dock.]

Introducing student teachers to resources such as this, which are easy to use, may well help to overcome the problem of the slow pace of corpus integration.

Using a genre-specific corpus: business correspondence

Concordance data are particularly useful in providing a variety of examples in genres where formulaic phrases occur frequently. The Business Letter Corpus, created by Yasumasa Someya and used as the basis for his master's thesis (Someya 1999), includes over one million words of British and American English. Someya's analysis of the corpus reveals a number of features of this genre, including, for example, that modals play an important role. A teacher can thus obtain information from the thesis on the most frequently used phrases and obtain genuine examples of their use by carrying out a simple search on the corpus web site. For example, a search of the words I and you reveals that appreciate (1005 occurrences), appreciated (257 occurrences), and grateful (247 occurrences), are used to express both gratitude and a request. In almost half of the examples grateful is qualified, as shown in Table 1.

It is easy to produce examples of various aspects of the usage of native speakers, for example, the variation in the use of modals, as seen in Extract 4 below.

Extract 4: Use of modals in concordances of the word grateful from the Business Letter Corpus (Someya 1999)

- 1: I would be very grateful if you could set aside some time soon t
- 2: and I would be very grateful if you personally supervised it.
- 3: We would be very grateful if you would allow us to use this metho
- 4: I should be very grateful if you would give me an interview.
- 5: I'd be very grateful if you'd give me a call to let me know

While the authenticity of this task and its relevance outside the classroom are clear for learners wishing to master the writing of business letters, the meaning focus is limited. Unfortunately, because of copyright conditions, it is not possible to view whole texts, so a teacher would have to find letters from another source, and use the concordance lines to provide extra examples of a particular item. The teacher's role

Table 1. Grateful in the business letter corpus.

Expression	Occurrences	Expression	Occurrences
Very grateful	54	Really grateful	1
Most grateful	43	Sincerely grateful	1
Extremely grateful	5	Truly grateful	1
Deeply grateful	5	Mighty grateful	1
How grateful	4		

as mediator is particularly important here, for instance, in encouraging the learners to reflect on the reasons why one writer was able to use contracted forms in the above examples. In the extract below, even the brief concordance lines provide some clues as to the contexts in which 'deeply grateful' could be appropriate:

Extract 5: Concordances of the words deeply grateful from the Business Letter Corpus (Someya 1999)

- 1: We are deeply grateful for the cooperation you have extended u

- 2: difficult and demanding period, and I am deeply grateful for your dedication
- 3: with my being offered the post, and I am deeply grateful to you for the reference you provided
- 4: in rearranging my manuscript, and I am deeply grateful to you.
- 5: You ran a good show, Chris, and I am deeply grateful.

It is clear that these lines deal with important matters, although it is still tantalising not to be able to check the full context, as would be the case if the teacher had access to the full text of a corpus and was using a concordancer such as Wordsmith Tools (Scott 2004) or MonoConc (Barlow 2000), where one click would give instant access to the full text. In other words, these concordances would have to be used as complements to an authentic business letter provided by the teacher. Despite this limitation, this easy access to concordance examples, without any of the training needed to master the use of a concordancer such as the two mentioned above, provides teachers with an extremely useful resource for developing a wealth of materials, or advising learners on how best to use the resource in their future professional practice.

Using corpora in more involved ways

This section focuses on corpora and data-driven learning (DDL), developing the discussions in the previous sections and considering what is involved in the selection of appropriate corpora which the teacher can use with concordancing software, and the questions surrounding language variation and possible editorial decisions. In these cases, the corpora are not all immediately available on the web as in the previous section, but can be purchased, and the teacher will have to learn how to use concordancing software, such as Wordsmith Tools, MonoConc or the freely available AntConc.² While the easily available resources discussed in the previous section have many advantages, it is important to note that the potential of choosing a corpus and using a concordancer is much greater, in particular in terms of the choice of corpora and the possibility of checking the context of a concordance line at any time, plus many added functions such as the generation of statistical information, keywords, collocations and clusters, and the ability to reorganise and present the data in a range of ways.

The type of corpus to use

Much choice exists in the range of corpora available, and there is also the possibility of building a corpus. However, a teacher must always consider that all corpora are built for a specific purpose (Hunston 2002: 14) and therefore their usefulness to the materials developer

must be assessed with this in mind. As we have seen, corpora can be divided into many groups depending on the criteria that are to be applied.

Written and spoken corpora

Written language is more readily available, and written corpora thus tend to be more accessible and larger. For this reason, early corpora, such as the Brown Corpus (American English) and the LOB corpus (British English), which were based on similar templates, contain written texts from a range of categories from press reports to fiction, to biographies and essays. More recent examples, which contain millions (soon to be a billion in some cases) of words of written language, include those funded by the major publishing houses, often in conjunction with universities. Some examples are: the Cobuild/Bank of English (approximately 440m words of written language) and the Cambridge International Corpus (approximately 780m words of written language), the Corpus of Contemporary American English (over 300m words of written language) and also a range of what have become known as the National Corpora, such as the British National Corpus (90m words of written language), and the American National Corpus (approx 18m words of written language). Spoken language corpora, on the other hand, are a much scarcer commodity, owing to issues of cost, access, quality of recording, time, and difficulties associated with transcriptions (McCarthy 1998). Nonetheless, there has been an increase in the number and size of spoken corpora. CANCODE (the Cambridge and Nottingham Corpus of Discourse in English) is one of the more significant and currently consists of five million words, but is not publicly available. Other examples include: the Santa Barbara Corpus of Spoken American English, the Corpus of London Teenage Language (COLT), and the Hong Kong Corpus of Spoken English (HKCSE). The choice of using a written or spoken corpus in designing materials for teaching purposes will depend on curriculum aims and student needs, and often teachers will find it useful to consult with both at different times. In fact, it can be very illuminating to examine differences and similarities between spoken and written corpora as these will highlight salient aspects of each. And we now have an emerging blend of these modes in online communication and the type of language used on the WWW (see Hundt, Nesselhauf, and Biewer 2007). To illustrate the spoken written divide, O’Keeffe, McCarthy, and Carter (2007: 140) provide an example of the frequency of the word *right* across two spoken and two written corpora, the results of which can be seen in Figure 1 below.

Based on these results, O’Keeffe et al. conduct a more detailed examination and find that *right* has important discursual functions as a boundary marker and is also used extensively as a response token. Extract 6 below, which represents a random sample of concordance lines from the Limerick Corpus of Irish English (L-CIE, one million words of mainly casual conversation) (Farr, Murphy and O’Keeffe 2004) illustrates the boundary marker and response token uses quite well:

Extract 6: Concordances of the word right from L-CIE

1: nothing of I can't make it out what it is. Right. I only looked at it twice really. I know because I

2: Susan which is a bit bit awkward. That's right yes yes no no there's nothing nothing like that am.

3: yes no no there's nothing nothing like that am. Right and wha= you you really did do you do not like this I

4: was a notice in the ah office you have to be eighteen. Right. But the one thing worries me is am a risk of

5: out.' ' Yeah. umhuh I didn't freak out but I was upset. Right so I mean you may find out whether lasers do or

6: in any gangs or you know criminals or anything like that. Right. You know a lot you can see on the television a lot o

7: really know if she knew how strongly we felt about it. Right right she'll probably be terrified to meet you now. P

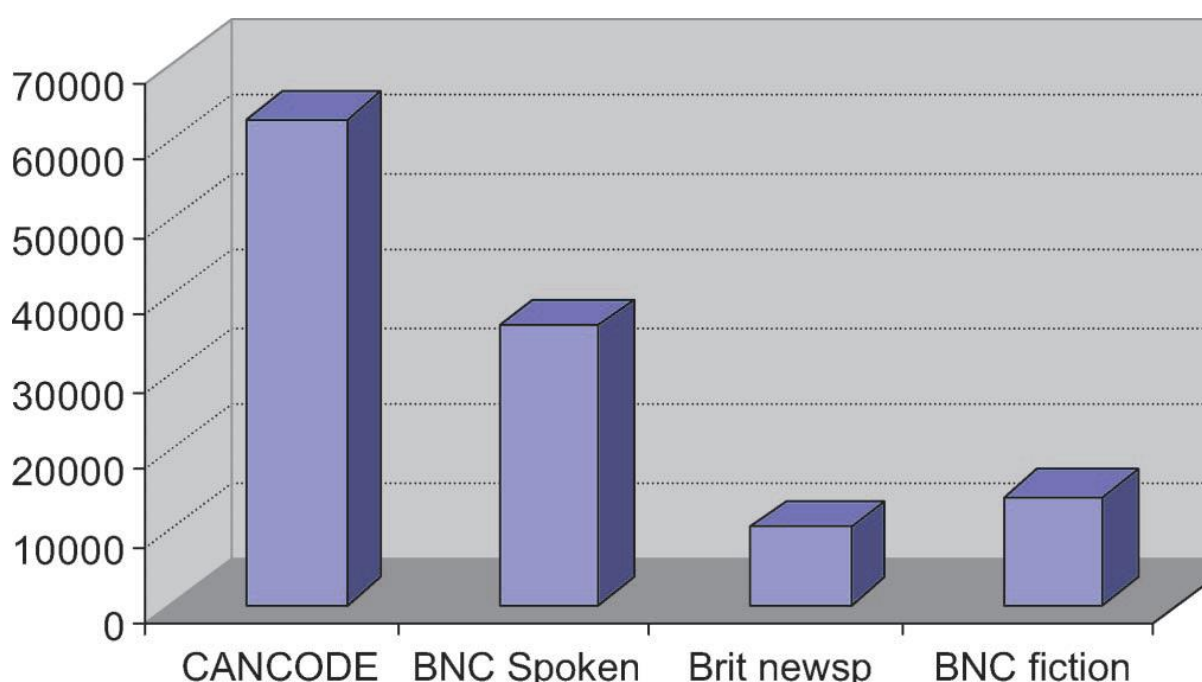
8: know if she knew how strongly we felt about it. Right right she'll probably be terrified to meet you now. Probably

9: errified to meet you now. Probably but I hope not . Right okay so basically you want to know any kind of

The examples indicate the very frequent use of right to attend to what the other person is saying, either as a single response token (for example, Line 1), or as a response token which leads into a turn takeover (for example, Line 3). The boundary marker use is also prominent in examples such as that in Line 5. Both of these functions come through strongly, not only in this nine line random sample, but in the 3096 occurrences of right in L-CIE. This type of awareness raising activity can be usefully employed with, for example, EFL students living in an English speaking environment, who wish to improve their casual conversation interactional strategies. And while the teacher will need to select relevant lines from the 3096, or have the software choose a random sample of 20–30 occurrences, to make such activities feasible within a classroom setting, our experience has been that getting students to consult with corpus evidence in this way is highly motivational as it reflects their real world, out of class encounters. Such explorations provide data that can have much potential for the development of materials to compare written, spoken and on-line language, or indeed, when looking at one mode in isolation, for example, spoken corpora, to improve students' awareness and production of spoken language. In addition, the importance of locally contextualised data cannot be underestimated in a learning environment for which the variety (in this case, Irish English) is not, on the whole, represented in textbooks.

In this section, we have touched on two important issues, and while the present paper does not allow for an extended discussion of either, they have both been given more detailed coverage in previous publications (see O’Keeffe and Farr 2003; McEnery, Xiao and Tono 2006). Firstly, there is the question of how many examples are required to allow a corpus user to generalise from the results. There is no definitive answer to this question, but repeated examples are required in order to reach general conclusions. The number of examples will depend on the size of the

Figure 1. Frequency of right across spoken and written corpora.



corpus, and a more prudent way to approach this issue is to foster critical interpretation skills among students and to encourage them to interpret corpus data in conjunction with information they might have from reference and text books, their own experiences and intuitions. Part of these critical skills includes interpreting the number of occurrences in conjunction with the type of corpus from which they emanate and the purpose for which the examples are being assessed. The second issue is the choice between allowing students to have direct access to corpora and the teacher mediating the material. It has been suggested that a progressive approach is advisable, with students initially being exposed to mediated data in the form of edited printouts, which are more controlled and help to develop the type of skills just mentioned. Once students are more confident and astute, they can then be slowly introduced to the software and direct manipulation of the data. In a teaching context where ICT and associated skills are highly valued, we have taken this approach, but it is possible to imagine many situations in which the teacher does not move beyond producing mediated

hand-out materials. As is the case with all methodologies and materials, the approach deemed most suitable for the particular group of students will be determined by the teacher in the local context.

General and specific corpora

Another choice involves whether, or when, it is appropriate to consult a general or a specialised corpus. A general corpus aims to be relatively balanced and includes data from ‘genres and domains that typically represent the language under consideration’ (McEnery, Xiao and Tono 2006: 59). The national corpora and those used in the production of reference books normally aim to be as general as possible so that broad conclusions can be reached. However, there are ongoing debates surrounding the issues of balance and representation (see Biber 1993; Sinclair 1995), and many involved in corpus exploration and building, particularly on a smaller scale, often opt for a more specialised corpus to meet their explicit needs. Indeed, research suggests that there is a very strong case to be made for the examination of specific domains as they may contain not only their own lexical patterning but also grammatical structures that differentiate them from other contexts of use (see Biber et al. 1999; Carter and McCarthy 2006). Many specialised corpora exist, particularly in the contexts of academic and business English, from example, the Michigan Corpus of Academic Spoken English (MICASE), the British Academic Spoken English (BASE) Corpus, and the Corpus of Spoken Professional American English (CSPAEE). Those involved in the delivery of LSP programmes can benefit greatly from the use of such corpora. Using comparative techniques, through the generation of frequency and keyword lists, students and teachers can appreciate the generic fingerprints of different types of language (for example, Farr 2007).

Small and large corpora

The notions of large and small are relative to whether one is referring to spoken or written corpora, and also to the technological capabilities of the period during which the data is collected. The one-million-word Brown Corpus of written data seems small compared with today’s one-billion-word Cambridge International Corpus, but for the late 1960s it represented a significant collection. Written corpora can reach millions of words and are therefore useful for the extraction of low-frequency words, but do not lend themselves to any large-scale qualitative analysis, other than using random sampling, as handling becomes problematic. On the other hand, smaller corpora will produce more manageable results for students and teachers to work with, as well as providing opportunities for accessing the original files to provide more context. Arguments have been made for students to initially experience corpus linguistics through smaller corpora as a sort of training forum during which their conceptual schema can adapt in preparation for using larger corpora (Gavioli and Aston 2001; O’Keeffe and Farr 2003).

Learner corpora

Corpus collections often contain data from language learners or proficient users, sometimes in a lingua franca context. They can consist of spoken or written productions from a range of domains. Learner corpora can have lots of applications ranging from pedagogic to socio-linguistic. Some examples include the Longman Learners' Corpus (10 million words of written language), which contains data from a range of first language backgrounds, the Cambridge International Corpus, mentioned above, which contains 20 million words from Cambridge exam scripts written by learners of English, and the International Corpus of Learner English, which contains three million words in the form of written essays from learners coming from a range of L1 backgrounds. Although the first two of these are not in the public domain, researchers have been examining such data to determine the type of difficulties learners have, and this in turn can inform the most appropriate content for materials and curricula (see for example Granger 2002). Researchers like Seidlhofer (2002, 2004) have been creating and building collections of English as a Lingua Franca (ELF) as part of their VOICE corpus, which is moving towards one million words of naturally occurring face-to-face conversations from reasonably fluent speakers. A similar corpus is ELFA (Mauranen 2003). In this type of research, emanating from socio-linguistic perspectives, an examination of the core features of successful communication between speakers is crucial to the discussions of the status of English and its use as an international language.

Editing the corpus

Due to the technical requirements of the corpus analysis software, annotation and tagging, the text in a corpus can look very far removed from the type of word processed material that the typical learner, even 'digital native' (Prensky 2001) may be familiar with. The following extract comes from the tagged version of the LOB corpus:

A01 2 ^ *' _*' stop_VB electing_VBG life_NN peers_NNS **' _**' _.

A01 3 ^ by_IN Trevor_NP Williams_NP._.

A01 4 ^ a_AT move_NN to_TO stop_VB \0Mr_NPT Gaitskell_NP from_IN

While a tagged version such as this may be of great interest and use to grammarians or researchers in particular fields, it probably needs to be modified if it is to be of benefit to a teacher or learner. It is always useful to have access to unmarked versions of corpora, and even then other information necessary for processing may still need to be edited out. Take the following example from the L-CIE, which has minimal coding, yet we often convert such

texts into something like what appears below it so that students are not unnecessarily distracted:

Extract 7: Converting corpus extracts to more reader friendly formats

<\$3> that routine maintenance doesn't be <\$G1> too much detail <\$E4> laughs
</\$E4> <\$O3> nevertheless it's it's a it's another job done </\$O3>
<\$4><\$O3> that's that's what <\$G2></\$O3> there's other sp = spuds peeled
there a while now
<\$3> I don't think we I think we'll do them in their jackets and I I'll actually <\$X>
d'ya j do you </\$X> know what you could do you could peel these ones and I might
just pop them in and roast them wouldn't that be an idea
<\$4> be nice all right.
<\$3> I'll give you those three <\$E> clattering sound </\$E> give you <\$O3>
those three and then we'll </\$O3>+

Michelle: That routine maintenance doesn't be [inaudible speech]. Too much detail
[laughs]. Nevertheless it's, it's a. another job done. [overlapping speech]

Kate: That's that's what [overlapping speech] . . . There's other sp. spuds peeled
there a while now.

Michelle: I don't think we I think we'll do them in their jackets and I I'll actually d'ya
know what you could do? You could peel these ones and I might just pop
them in and roast them. Wouldn't that be an idea?

Kate: Be nice all right.

Michelle: I'll give you those three [clattering sound] give you those three and then
we'll . . . [overlapping speech]

In this case, seven steps have been taken to make this more like the kind of text students and teachers are used to. We removed annotations, added conventional punctuation, extended and

contextualised extralinguistic information, edited the format, removed some of the repetitions and hesitations, gave the speakers fictitious names, and inserted line numbers. The degree to which such editing is desirable will depend on the student audience and on the aim of the activity. In many corpora, contextual information such as personal and place names will have been removed by the corpus creators, but if teachers decide to build their own databases, they should take great care to ensure that this is done, for ethical and legal reasons. Extract 7 also raises another important issue which becomes apparent when looking at informal spoken data, that of non-standard language. There are three types exemplified in the extract. First of all, there is an example of a lexical item, the word *spuds*, which is understood by all speakers of Irish English and also considered ‘acceptable’ by most in informal speech, and possibly even in written language. Teachers may consider such language a valuable addition to a student’s passive or active lexical repertoire as it can give valuable cultural insights. Contrasted with this, we can see in the first turn, the truncated verb phrase *doesn’t be*, and soon after, the subject verb object disagreement *there’s other spuds*. Many might consider the first example to be very much non-standard and not the type of language that learners need to be unnecessarily exposed to, as many educated native speakers would not use such structures. On the other hand, the use of *there is* + plural object is more ubiquitous and found in other varieties also. Therefore, though non-standard in the traditional sense, it is part of the relatively stable variety of Irish–English commonly used by many speakers and it may therefore be preferable not to edit it out for the purposes of materials development.

Also found in corpora is another type of ‘bad’ language, but this time we refer to taboo language, including the use of religious references and swear words. Speech is generally transcribed as it is spoken, in all of its colour and without censorship in order to capture as much authenticity as possible.

Table 2. Taboo language items in the LCIE.

	Frequency (words per million)
Religious	
God	715
Jesus	459
Christ	65
Hell	59
Lord	33
Saint	28
Bless*	16

Sexual

Fuck*	773
Dick*	15
Whore*	15
Cunt*	9
Fag*	9
Wank*	2

Body

Shit*	211
Crap*	80
Piss*	65
Arse*	43
Bloody	26
Bull(shit)	22
Boll*	21
Fart*	15
Bleed*	7

Mind

Eegit/idiot	38
bastard	36
Spastic/spa	12
Nut*	9
morons	1

Animals

Dog*	126
bitch	55
Pig*	17
cow	13

Misc.

Feck*	104
lousy	4

Notes: *is known as the wildcard and indicates that all words beginning with these letters are included as well as the word itself.

This means that transcriptions can be laced with what might be considered to be obscene language. L-CIE contains very high frequencies for two reasons. Firstly, it is approximately 80% casual conversation, and secondly, it represents a variety of English in which speakers are known to use taboo language more frequently and in a wider range of contexts than in other varieties (Farr and Murphy 2009). Table 2 illustrates the most frequent language items with the potential to function in an offensive way.

Some of these items are highly frequent and some are significant in terms of cultural insights and would indeed be of interest to many sociologists or corpus linguists, but the decision on whether these items should be excluded in pedagogic contexts is one for the materials developer to make based on the teaching context and culture. Of course, exclusion is only logistically feasible if using the data in a qualitative way. It would be extremely time-consuming to comb a corpus of any notable size before allowing students to run concordances. This, along with the other factors mentioned in this section, is a serious consideration when deciding if and how to use corpora.

Conclusion

Corpora clearly have an important role to play in the language learning and teaching environment, making available multiple examples of attested language use and providing access to specialised registers. For many practising teachers, using easily accessible corpus data with built-in concordancers for materials development will be the first step, alongside the use of corpus-based resources such as dictionaries, grammars and course books. Even achieving this level of corpus consultation, which conforms to the findings of Egbert et al. on teachers' practices in the use of technology, could greatly enhance the language learning environment for teachers and learners. For those who wish to travel further along the path of using corpus data with their learners, resources are increasingly becoming available, and some language teacher educators are now also including corpus consultation and analysis in their programmes. This is still an area, however, where research and practice have yet to meet. While there have been notable successes in the use of corpus data for dictionaries, course books and grammars, corpus consultation literacy is not yet seen as an essential skill for language teachers, let alone learners. With few exceptions – Braun's (2007) analysis of pupils in secondary education in Germany is one – empirical studies involve researchers in higher education focussing on their own students, with the result that very few language learners in secondary education are aware of the existence of corpora, let alone the possibility

of using them either in their work at school or in their higher education or future careers. There is obviously much scope for future research and dissemination in this area.

Notes

1. Yoon and Hirvela's study is both quantitative and qualitative.
2. AntConc. http://www.antlab.sci.waseda.ac.jp/antconc_index.html (accessed October 14, 2010).

References

- Barlow, M. 2000. *MonoConc Pro*. Houston: Athelstan.
- Bernardini, S. 2000. Systematising serendipity: proposals for concordancing large corpora with language learners. In *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*, ed. L. Burnard and T. McEnery, 225-234. Frankfurt am Main: Peter Lang.
- Bernardini, S. 2002. Exploring new directions for discovery learning. In *Teaching and learning by doing corpus analysis*, ed. B. Kettemann & G. Marko, 165-182. Amsterdam and New York: Rodopi.
- Biber, D. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8, no. 4: 243-257.
- Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman grammar of spoken and written English*. London/New York: Longman.
- Braun, S. 2005. From pedagogically relevant corpora to authentic language learning contents. *ReCALL* 17, no. 1: 47-64.
- Braun, S. 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL* 19, no. 3: 307-328.
- Carter, R., and M. J. McCarthy. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Chambers, A. 2005. Integrating corpus consultation in language studies. *Language Learning and Technology* 9, no. 2: 111-125.
- Chapelle, C. A. 2001. *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.
- Cobb, T. 1997. Is there any measurable learning from hands on concordancing? *System* 25, no. 3: 301-315.
- Cuban, L. 1986. *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.
- Egbert, J., T. M. Paulus and Y. Nakamichi. 2002. The impact of CALL instruction on classroom computer use: A foundation for rethinking technology in teacher education. *Language Learning and Technology* 6, no. 3: 108-126.
- Farr, F. 2007. Spoken language as an aid to reflective practice in language teacher education: using a specialised corpus to establish a generic fingerprint. In *Spoken corpora in Applied Linguistics*, ed. M.-C. Campoy, and M.-J. Luzón, 235-258. Bern: Peter Lang.
- Farr, F. 2008. Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users. *Language Awareness* 17, no. 1: 25-43.

- Farr, F., A. Chambers and S. O’Riordan. In press. Corpora for materials development in language teacher education: Principles for development and useful data. In *Perspectives on language learning materials development*, ed. F.Mishan and A. Chambers. Bern: Peter Lang.
- Farr, F., and B. Murphy 2009. Religious references in contemporary Irish-English: ‘for the love of God almighty...I’m a holy terror for turf’. *Journal of Intercultural Pragmatics* 6, no. 4: 535-560.
- Farr, F., B. Murphy and A. O’Keeffe. 2004. The Limerick Corpus of Irish English: design, description and application. In *Corpora, varieties and the language classroom. Special Edition of Teanga 21*, ed. F. Farr and A. O’Keeffe, 5-29. Dublin: IRAAL.
- Fligelstone, S. 1993. Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME 17*, 97-109.
- Gaskell, D. and T. Cobb. 2004. Can learners use concordance feedback for writing errors? *System* 32, no. 3: 301-319.
- Gavioli, L. and G. Aston. 2001. Enriching reality: language corpora in language pedagogy. *English Language Teaching Journal* 55, no. 3: 238-246.
- Granger, S. 2002. A bird's-eye view of learner corpus research. In *Computer learner corpora, second language acquisition and foreign language teaching*, ed. S. Granger, J. Hung and S. Petch-Tyson, 3-33. Amsterdam: John Benjamins.
- Hundt, M., N. Nesselhauf and C. Biewer. 2007. Corpus linguistics and the web. In *Corpus Linguistics and the Web*, ed. M. Hundt, N. Nesselhauf and C. Biewer, 1-6. Amsterdam: Rodopi.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johns, T. 1986. Micro-concord: A language learner's research tool. *System* 14, no. 2: 151-162.
- Johns, T. 1994. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In *Perspectives on pedagogical grammar*, ed. T. Odlin, 293-313. Cambridge: Cambridge University Press.
- Johns, T. 1997. Contexts: The background, development, and trialling of a concordance-based CALL program. In *Teaching and language corpora*, ed. A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles, 100-115. London: Longman.
- Kennedy, C. and T. Miceli. 2001. An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology* 5, no. 3: 77-90.
- Levy, M. 1997. *Computer-assisted language learning: Context and conceptualization*. Oxford: Clarendon.
- Levy, M. and G. Stockwell. 2006. *CALL dimensions: Options and issues in computer-assisted language learning*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Mauranen, A. 2003. The corpus of English as a lingua franca in academic setting. *TESOL Quarterly*, 37, no. 3: 513-527.
- McCarthy, M. J. 1998. *Spoken language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McEnery, T., & Wilson, A. 1997. Teaching and language corpora. *ReCALL*, 9 no. 1: 5-14.
- McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based language studies. An advanced resource book*. London and New York: Routledge.
- O’Keeffe, A. and F. Farr. 2003. Using language corpora in language teacher education: pedagogic, linguistic and cultural insights. *TESOL Quarterly* 37, no. 3: 389-418.
- O’Keeffe, A., M. McCarthy and R. Carter. 2007. *From Corpus to Classroom*. Cambridge: Cambridge University Press.

- O'Sullivan, Í. and A. Chambers. 2006. Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing* 15: 49-68.
- Prensky, M. 2001. Digital natives, digital immigrants Part 2: Do they really think differently? *On the Horizon* 9, no. 6: 1-6.
- Renouf, A., A. Kehoe and J. Banerjee. 2007. WebCorp: an integrated system for web text search. In *Corpus Linguistics and the Web*, ed. M. Hundt, N. Nesselhauf and C. Biewer, 47-68. Amsterdam: Rodopi.
- Römer, U. 2006. Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54, no. 2: 121-134.
- Scott, M. 2004. *Wordsmith Tools Version 4.0*. Oxford: Oxford University Press.
- Seidlhofer, B. 2002. Pedagogy and local learner corpora. Working with learning-driven data. In *Computer learner corpora, second language acquisition and foreign language teaching*, ed. S. Granger, J. Hung and S. Petch-Tyson, 231-234. Amsterdam: John Benjamins.
- Seidlhofer, B. 2004. In search of 'European English': Or why the corpus can't tell us what to teach. Paper Presented at EuroCALL, Vienna, 3-5 September.
- Sinclair, J. 1991. *Corpus, concordance, collocation*. London: Longman.
- Sinclair, J. M. 1995. Corpus typology - a framework for classification. In *Studies in Anglistics*, ed. G. Melchers and B. Warren, 17-46. Stockholm: Almqvist and Wiksell International.
- Someya, Y. 1999. A corpus-based study of lexical and grammatical features of written business English. MA dissertation, University of Tokyo <http://www.someya-net.com/09-MA/> (accessed 25 October, 2009).
- Stevens, V. 1991. Concordance-based vocabulary exercises: A viable alternative to gap-fillers. In *Classroom concordancing: English Language Research Journal 4*, ed. T. Johns and P. King, 47-63. Birmingham, England: Centre for English Language Studies, University of Birmingham.
- Tribble, C. and G. Jones. 1990. *Concordances in the classroom: A resource book for teachers*. Harlow: Longman.
- Widdowson, H. G. 2004. *Text, context, pretext. Critical issues in discourse analysis*. Oxford: Blackwell.

Corpora available online (all accessed 25 October, 2009)

- British National Corpus. <http://www.natcorp.ox.ac.uk/>
- Business Letter Corpus. <http://www.someya-net.com/concordancer/>
- The Chambers-Le Baron Corpus of Research Articles in French/Le Corpus Chambers-Le Baron d'articles de recherche en français. Oxford: Oxford Text Archive.
<http://ota.ahds.ac.uk/headers/2527.xml>
- The Chambers-Rostand Corpus of Journalistic French/ Le Corpus Chambers-Rostand de français journalistique. Oxford : Oxford Text Archive.
<http://ota.ahds.ac.uk/headers/2491.xml>
- Cobuild Concordance and Collocations Sampler.
<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>
- COLT: Corpus of London Teenagers. Department of English, University of Bergen.
<http://torvald.aksis.uib.no/colt/>
- Corpus del Español. <http://www.corpusdelespanol.org/>
- Elisa. http://www.uni-tuebingen.de/elisa/html/elisa_index.html
- Lextutor. <http://www.lexutor.ca/>
- SACODEYL. <http://www.um.es/sacodeyl/>

Appendix 1: Playing games in French corpora

Below are examples of concordance lines which a teacher might use to enhance a class involving teaching how to refer to playing games and musical instruments in French. Examples 1 and 2 are taken from the Chambers-Rostand Corpus, and give examples of how the verb *jouer* is used with the preposition *à* to denote playing a game, while the preposition *de* denotes playing a musical instrument.

Example 3 is taken from SACODEYL, a corpus of recordings of French-speaking teenagers talking about their lives, including their interest in sport. Interestingly, they mostly do not use the verb *jouer*, but rather *faire*.

These simple concordance lines present the teacher with a challenge. Should she simply teach the use of *jouer*, traditionally presented in course books as the way to convey playing a sport, and indeed attested in quotations in the journalistic corpus, or should she also present the learners with the opportunity of accessing the SACODEYL corpora, video-recordings of French-speaking teenagers available online, where they more commonly use the verb *faire* to refer to playing games. The concordance lines also allow the possibility of presenting both examples to the learners.

Example 1

« Moi, mon père ne voulait pas que je joue au foot, raconte Miriam, 16 ans.
« Des filles qui jouent super bien au foot et qui ne peuvent pas venir au club, il y en a plein
« Il m'a appris à jouer. Je ne savais pas jouer au tennis avant lui.

Example 2

AU « GUINNESS BOOK RECORDS »? Jouer du violoncelle en même temps que du didjeridoo et laisser sa main pianoter
Alors qu'il veut apprendre le saxophone, ses proches lui répondent que personne n'en joue dans la famille et qu'il ne fera pas exception.
Si Alice est censée être au piano, Agathe Alexis fait simplement mine d'en jouer en agitant ses doigts.
Hasna joue de la guitare électrique, du oud, de la derbouka, du bendir et même du banjo

Example 3

Oui, alors je fais du sport. Au lycée, je fais du badminton.
Et j'aime bien faire de l'ordinateur.
Oui, je fais du football dans un club.
Ben, eh ben, je fais les quatre appareils : le sol, le saut, les barres asymétriques et la poutre
Ben tout ce qui est du sport, généralement, j'aime bien. Je fais de la piscine en dehors, enfin avec le lycée,
Le mercredi, je fais du rock et de la salsa.
je fais de la piscine. Sinon, si vous voulez que j'énumère tous les sports que j'ai faits, on n'a pas fini.
Alors, ça fait six ans que je joue au football.