INTEGRATING CORPUS CONSULTATION IN LANGUAGE STUDIES

Angela Chambers
University of Limerick

ABSTRACT

Alongside developments in language research, the potential of corpora as a resource in language learning and teaching has been evident to researchers and teachers since the late 1960s. Despite publications which emphasise the benefits of corpus consultation for language learners (Bernardini, 2002; Kennedy & Miceli, 2001), there is little evidence to suggest that direct corpus consultation is coming to be seen as a complement or alternative to consultation of a dictionary, course book, or grammar by the majority of learners. There is thus a need for research to underpin the integration of corpora and concordancing in the language-learning environment.

This study begins with an account of published research relating to course design and structure in the area of corpus consultation by language learners. The focus then narrows to the initial training of learners in corpus consultation, using as an example a course involving undergraduate students on several language degree programmes. The results of the students' consultation of the corpora are examined, including choice of search word(s), analytical skills, the problems encountered, and their evaluation of the activity. The results reveal how corpus consultation can complement traditional language-learning resources, while also raising questions concerning its integration in the language-learning environment.

INTRODUCTION

Since large computerised corpora of English were created in the 1960s, there has been a steady increase in the number of publications devoted to their use in the context of language teaching and learning. The pioneering work of Johns (1986) and Tribble and Jones (1990) was followed by an explosion of studies devoted to various aspects of the use of corpora in language learning in various contexts, for example the publications resulting from the TALC (Teaching and Language Corpora) conferences on teaching and language corpora (see, e.g., Burnard & McEnery, 2000; Kettemann & Marko, 2002). From the early 1990s onward, corpora were clearly being consulted by language teachers, and also by learners, at least in courses run by researchers and enthusiasts, and this activity was gaining in popularity by a process which McEnery and Wilson (1997, p. 5) describe as percolation. This has created a need for research to underpin this new development, focusing on aspects such as the type of corpora to be consulted, large or small, general or domain-specific, tagged or untagged.

Other pedagogic issues also require investigation, such as the advantages of direct access to corpora as opposed to mediation by the teacher through the preparation of corpus-based worksheets, the strategies which learners need to acquire to benefit from direct consultation, and, last but not least, the means by which this new activity can best be integrated into the language-learning environment. Some of these issues are already receiving considerable attention from researchers, with a number of studies recommending the use of small corpora tailored to the learners' needs (Aston, 1997; Roe, 2000), while others champion large corpus concordancing (Bernardini, 2000; Cheng, Warren, & Xun-feng, 2003). Direct access to corpora by learners is the subject of a number of studies (see, e.g., Bernardini, 2002; Chambers & O'Sullivan, 2004; Kennedy & Miceli, 2001, 2002), with a cautionary note from Johns (1997, p. 113) recommending the use of corpus results mediated by the teacher as a first stage. While there is already a substantial and increasing body of research in several aspects of direct corpus consultation by learners, there is still considerable scope for developments, particularly in the area of course design and

structure, concerning how one can successfully integrate corpus consultation into a programme of language study in higher education.

The publications on corpus consultation quoted in this study give varying amounts of information on the types of course structure within which they are operating, including the aims of their courses and the time allotted to them. But all this is presented as a given, understandably so, as the studies do not aim to investigate issues arising from course design and structure. The aim of this study is to examine a number of aspects of course design in corpora and language learning involving direct access by learners, focusing not on the training of corpus linguists but rather on the popularisation of corpus consultation by a wide spectrum of learners. After a brief overview of the types of courses which are described in the studies referred to above and other similar publications, one example will be examined in more detail, namely a section of a second-year undergraduate course on language and technology which aims to encourage the learners to use corpora as a resource in their language learning alongside other resources such as the dictionary, course book, and grammar. The course aims, structure, content, and assessment will be briefly described, paying particular attention to the training provided in concordancing and corpus analysis, the corpus resources used, the students' choice of an aspect of the language to be studied, the strategies which they require to benefit from the corpus consultation, their success or otherwise in analysing the results, and their evaluation of the activity. This will enable us to draw some conclusions concerning the factors which favour the integration of corpora and concordancing into the language-learning environment and the obstacles which remain to be surmounted.

COURSE DESIGN IN CORPORA AND LANGUAGE LEARNING

Within the disciplinary area of language studies, corpora and corpus-based methods are increasingly used outside language learning per se, in areas such as the teaching of literature (see, e.g., Kettemann, 1995; Louw, 1997) and of translation (see, e.g., Bowker, 1998; Zanettin, 2001). This section, however, will include only research concerning those wishing to learn about language either as linguistic researchers or language learners. Fligelstone (1993, p. 98) proposes what he terms a simple framework for assessing "the factors relevant to good teaching practice," grouping corpus-related activities into three categories:

> TEACHING ABOUT (i.e., teaching about corpora/corpus linguistics)
>
> TEACHING TO EXPLOIT (i.e., teaching students to exploit corpus data)
>
> EXPLOITING TO TEACH (i.e., exploiting corpus resources in order to teach)

Even from reading only the small selection of studies of direct corpus consultation by learners referred to above, it is clear that there is considerable variation in the nature of the courses on which they are based, ranging from courses clearly designed as part of a programme of study in linguistics, to a limited amount of training included in a language course so that the learners can benefit from consulting a corpus. Davies (2000), for example, uses corpora of historical and dialectal texts when teaching an advanced course in Spanish linguistics. Similarly, the description of Paul Thompson's (2004) postgraduate module in corpora in applied linguistics in the University of Reading clearly situates it within the discipline of corpus linguistics. At the other end of the scale, in the sense not of being inferior but of having very different aims and therefore content, certain courses, mostly at undergraduate level, include a very limited amount of training in corpus consultation with the practical aim of enabling the learners to consult corpora to improve their language skills. A comparison of one such course, part of a second-year undergraduate module at the University of Limerick, and the Reading postgraduate course, reveals both the similarities and differences between them (see Table 1).

Table 1. General and Specialised Courses

	Undergraduate Year 2	Postgraduate
Course	Part of course/ Core	Complete course/ Specialist option
Hours	9	20
Corpus resources	Small	Small + large
Corpus creation	No	Yes
Corpus analysis	Yes	Yes
Tagging	No	Yes
Assessment	3,000-word essay	3/4,000-word essay

While both courses include lectures on corpus linguistics and on the analysis of corpora, alongside practical laboratory sessions, the postgraduate course is a specialist option embedded within the already specialised context of masters programmes in Applied Linguistics and ELT. It is allotted time to allow for greater depth of study and familiarisation with the tagging of corpora, while the core undergraduate teaching is, as we shall see, part of a second-year module and is obliged to make room for other aspects of technology and language study, also considered as core elements of the degree programme. It is this situation which creates the challenge of popularising corpus consultation, informing students of its potential benefits and giving them the skills to benefit from it in a very limited amount of time, as well as providing access to resources for future use and guidelines on how best to benefit from them.

Before examining the undergraduate course in more detail, it is important to note that the publications relating to corpus consultation by learners do not all fit neatly into the two types of course described in Table 1, or into one of Fligelstone's three categories. Several other studies contain elements of both the postgraduate and undergraduate courses, supporting Fligelstone's (1993, p. 98) comment that there is a certain amount of interaction between his three categories. Aston (1997, p. 61), for example, notes that the analysis of small corpora for language-learning purposes can serve as a useful starting point for students who may later wish to move on to the analysis of larger corpora in a research context. Dodd (1997), referring to the use of unedited corpus data with advanced students at undergraduate and postgraduate level, comments,

> At this level, several teaching aims are likely to coincide. These include improving the practical proficiency of the learner, improving the learner's formal knowledge about the language (and about language in general), and giving the student an insight into the work of the descriptive grammarian. (p. 132)

In another context, Cheng et al. (2003) are able to devote a much more substantial amount of classroom and laboratory contact hours over two semesters to corpus design and analysis than their Limerick counterparts. Corpora and concordancing are taught by them as a substantial part of second-year undergraduate courses on Information Technology and Discourse Analysis, within an English language major undergraduate programme. Their aims include both research in corpus linguistics and the practical benefits of language learning, firstly, placing the students "in the role of language researchers finding out for themselves about the English language" (p. 178), and secondly, at the same time encouraging them "to reflect on their experiences as language learners and English language majors from this form of data-driven learning" (p. 178). The much greater amount of time available to them enables them to introduce the students to work with larger corpora and to move further into the study of corpus linguistics as a discipline than the shorter undergraduate course. In the context of popularising corpus consultation, however, the Limerick course is interesting by its very limitation, in that it can be seen as a component of a course which one could reasonably envisage being included in all undergraduate language degree programmes. Kennedy and Miceli (2001, 2002) are very possibly examples of other researchers working within similar parameters, in that there is no evidence in their publications that the degree programmes

involved have a noticeable bias towards Information Technology or Discourse Analysis, as in the case of Cheng et al. Looking at the variety of course design and structure within the publications which study corpus consultation by learners, it seems clear, without in any way undermining the validity of Fligelstone's framework, that the range of courses or parts of courses devoted to direct access to corpora can be situated on a continuum rather than within a clearly defined category.

CORPUS CONSULTATION AT POPULARISATION LEVEL

While a limited amount of training in corpus consultation at undergraduate level may be the first step in a career as a corpus linguist, it seems reasonable to assume that the majority of those undertaking such a course have at this stage no ambition to become experts in corpus linguistics, any more than they desire to become lexicographers when using the dictionary or materials designers when learning from a course book. Their interest is thus most likely to be aroused if they perceive the activity as being of benefit to their language learning. Thus, while a small number of graduates of the Limerick course are active in research involving corpus-based methods, the main aim of the course is to encourage all students to consult corpora as part of their language learning.

Corpora and concordancing are included in a second-year module on language and technology, which aims to introduce students to the major pedagogical, professional and research applications of technology in modern languages and to enable them to integrate these into their studies. Corpora and concordancing is one of four components, each of which is taught for 3 weeks, with one lecture and a 2-hour session in a computer laboratory. For the module assessment students submit two 3000-word essays, selecting any two of the areas covered. While there is some variation in the students' choice of topics from year to year, the essays are more or less equally divided between the four areas, namely introduction to technology and language learning, WELL (Web-enhanced language learning) evaluation and personalisation of language technologies, corpora and concordancing, and machine translation. The students taking the corpora and concordancing module come from a variety of different courses. Most are following either the BA in Applied Languages or the BA in Applied Languages with Computing in which it is a core module; some take it as an optional module on the BA in Languages and Cultural Studies; others are exchange students, mostly from France, Germany, Romania, and Spain. All of these students are studying at least one language to degree level, some are studying two, and a number are studying three, including English, French, German, Irish, and Spanish. This brings together a relatively diverse community of university language learners in the module, making it a suitable terrain for an evaluation of the potential for popularising corpus consultation.

The lectures on corpora and concordancing include an introduction to corpus linguistics and an account of the types of research based on corpora, focusing in particular on the use of corpora in language learning. In the laboratory sessions, using Wordsmith Tools (Scott, 1999), students receive training in the use of the software and guidance on corpus consultation and analysis. Within the general aim of the module to encourage students to integrate technology into their studies, the classes on corpora and concordancing are intended to show how corpora can be used a resource for language learning alongside their more traditional counterparts, the course book, grammar, and dictionary. It is emphasised that corpora can complement these resources, and, as we shall see, the students particularly appreciate the access to a large number of examples, and to language use which one describes as "authentic, up to date, and relevant."

The laboratory sessions also provide guidance on the use of the corpora for problem-solving activities, and students are encouraged to bring examples of their written work to the classes and to try to improve them through corpus consultation. They are given advice on the selection of appropriate search words, for example the choice of a noun as a search word in order to find what verbs accompany it. (For a more detailed analysis of the strategies used by students in corpus consultation and the type of guidance provided by the teacher, see Chambers & O'Sullivan, 2004, and Kennedy & Miceli, 2001.)

The fact that five languages are involved makes the choice of the corpora used for training purposes a difficult one. After experimenting with a number of options, it was decided to create what we call "training corpora" in the five languages involved. Version 1 of these corpora includes for each language a journalistic corpus of 100,000 words and a corpus of academic writing of 50,000 words (published research articles and parts of masters and doctoral theses). In 2003 only the journalistic corpora were available. Each contained articles on a wide variety of topics, including current affairs, editorials, reviews, and sport, collected in 2002-2003 from two newspapers, in the case of French, for example, Le Monde and L'Humanité, and in the case of English, The Irish Times and The Independent (the Irish newspaper of that name). While the limited size was determined largely by time and resources, and while it is intended to expand the size for future cohorts, it is interesting to note that certain other researchers with experience of using corpora with undergraduate learners also choose to use very small training corpora. Kennedy and Miceli (2001, p. 79), for example, created sub-corpora of 50,000 words from their corpus of approximately 570,000 words for initial training in corpus analysis, and Gavioli (2001, p. 108) comments that 50,000 words is a lot for a learner. Dodd's (1997, p. 131) comment that "a modest corpus of a million or so words is certainly enough to make a valuable teaching aid" may well provide some common ground between those aware of the difficulties experienced by undergraduate students analysing a corpus for the first time, and the champions of large corpus concordancing such as Bernardini (2000, 2002).

Large corpora clearly provide superior resources for the study of language, providing many examples of a much larger proportion of the words in a language than their smaller counterparts. They do, however, present a number of disadvantages for beginners with limited time available for training, in particular long waiting times for searches for common words. In addition, the need to cope with examples from several sub-corpora in different genres would further complicate what is already a challenging task for the learners. In the particular context of this course, providing training in five languages using large corpora, all different in size, content, and means of access, would clearly be impossible. As we have seen, however, researchers working in one language only also show a strong preference for using small corpora for initial training, suggesting that their experience of the reality of using corpora in the classroom or computer laboratory has led them to the conclusion that using small corpora is most likely to succeed. It is also interesting to note that the exceptions to this, namely Bernardini (2000) and Cheng et al. (2003), are working in more specialised contexts at undergraduate level (translation, and information technology and discourse analysis, respectively) and have more time available for training over a much longer period. This is not to suggest that the learners in this and similar studies should be limited to using small corpora throughout their degree programmes and beyond. As we shall see, this study raises questions concerning the resources, support, and guidance provided as a follow-up to this type of introductory training.

The theoretical and pedagogical basis of this course is firmly situated within what Benson (2001) terms technology-based and teacher-based approaches to learner autonomy, favouring "independent interaction with educational technologies" and emphasising "the role of the teacher […] in the practice of fostering autonomy among learners" (p. 111). Thus, although considerable guidance is given in the choice of the aspect of the language which they may wish to analyse for the assessed coursework, that choice is theirs. As the aim is to encourage them to see corpora as a resource in language learning alongside the course book, dictionary, and grammar, they were asked to choose a problem which they had encountered in their language work and to compare the usefulness of the corpus with what they could learn from a course book or grammar. In the lectures they were introduced to the concept of lexical grammar, and it was made clear that they could investigate not just traditional grammatical concepts, but common words where the corpus, despite its limited size, might reveal relevant lexico-grammatical patterns which students at this level might not master and where a grammar or course book are of little or no use.

A number of examples were given, including "la question" in French and "time" in English. None of the students chose this option, perhaps because it seemed easier to choose an aspect of the language which was clearly identified in the course book or grammar, or perhaps because, influenced by traditional

language-teaching methods, it seemed more beneficial to them to identify traditional aspects of grammar which were problematical for them in their language work and to see what they could learn from the corpus. As we shall see, however, from their traditional grammatical starting points, they often made discoveries which were lexico-grammatical rather than solely grammatical in nature, thus benefiting from the corpus-based approach. The choices of the 14 students who submitted essays for this part of the course in 2003 were as follows:

Table 2. Essay Topics

Language	Essays	Topics
English	3	Verb + to + infinitive/Verb + gerund
		Since
		Come/phrasal verb
French	8	*En* (3)
		Pour
		Connaître/*savoir*
		Negation
		Celui/ceux/celle/celles
		Pronominal verbs
Irish	1	Is
Spanish	2	*Ser/estar*
		Hasta

Given the variety of languages and topics, it is not possible to give a detailed individual analysis of all 14 students' work within the limited scope of this study. The analysis will therefore be limited to comments on the success or otherwise of the students' analyses, with particular reference to the strategies they used and their evaluation of the activity. As it is interesting to observe the relationship between the students' analyses, strategies, and evaluation, a number of individual students' work will be used as starting points, with examples from others added subsequently to illustrate points common to many of the essays. It is important to note that there was very considerable variation in the amount of learning resulting from the corpus consultation. In one case, for example, a student did not reduce the 837 occurrences of the search word through random selection, and the analysis consisted solely of isolated comments on individual expressions which she noticed among the 837 results. In this case, poor corpus consultation and analysis skills were clearly an obstacle to the learning experience. This, however, was the exception rather than the rule, and the great majority of the students benefited from the activity. The cases examined in more detail below represent varying degrees of success. (Essays involving Irish and Spanish are not included here, as the author has no knowledge of these languages and had to rely on assistance from colleagues in assessing them.)

To examine how successful the students were in this activity, it is important to ask the question, What did they learn from the corpus? Clearly, given the limited size of the corpora, the concept of the language learner as researcher cannot be applied in any literal sense although, as we shall see, discovery learning is possible. In the lectures, following Dodd (1997, pp. 135-136), two possible methods of analysis were suggested: deductive and inductive. In the deductive method, which I presented as the easier of the two and the more appropriate for initial work on a small corpus, the grammar or the course book would be studied first, and the student would then apply the rules to the concordance results and compare the two resources. A student adopting the more demanding inductive approach would try to infer the rules from the concordance results, only attempting this if a large number of results was obtained. Of the 14 students, 9 chose the latter method, perhaps suggesting that they found the inductive approach more natural or more interesting when using corpus data. It is difficult to draw firm conclusions on this subject, however, as it is possible that they simply decided to present their results in this way. A clear example of how the

inductive method was presented is found in the study of verb + to + infinitive and verb + gerund. Only 49 examples of the first were found, along with 41 of the second. The student inferred rules from these, which she[1] subsequently confirmed in the grammar of her choice, commenting that the grammar included far fewer examples. She noted that only one area covered in the grammar did not appear at all in her concordance, namely the use of go + gerund to describe a recreational activity, for example to go fishing, boating, birdwatching, and so forth. She did, however, discover "to go missing" in the corpus, which was not covered in the grammar, illustrating the role of the corpus as a source of serendipitous discoveries, as noted by Johns (1988, p. 21) and Bernardini (2000, p. 225). This led her to reflect on why the corpus did not include examples of go + gerund to refer to recreational activities, which she attributed not to its limited size but to its nature, concluding that it was a better resource for "formal language use" than for what she termed "everyday language," recommending the addition of magazine articles to remedy this. Her conclusion was similar to that of the majority of the students, namely that the grammar is useful for explaining rules and -- especially -- exceptions, but that the corpus complements it, giving a much larger number of up-to-date examples that are easier to remember, and sometimes giving information which is not found in the grammar.

A clear example of the deductive approach was found in the study of *connaître* and *savoir.* More interestingly, this student's essay also reveals how a learner can derive benefit from corpus consultation without realising the full potential of the activity. With careful use of the wildcard, the student found most of the 75 examples of all forms of the lemma *savoir* and of the 50 of *connaître.* Following a brief presentation of how this area, which was problematic for her, was covered in two grammars, she proceeded to analyse the concordance results. The analysis of the occurrences of *savoir* focuses on several of the colligational contexts in which it is used, including *savoir* as noun, *savoir que, savoir ce que, savoir* + infinitive, *savoir* + noun, *faire savoir, savoir si, savoir à quel point,* in all a much richer presentation than is found in the grammar. Expressions such as *à savoir, reste à savoir,* and *en savoir gré,* which are found in the corpus, remain unexplored by the student, as does the use of the conditional tense, where the meaning changes from *savoir* in the sense of "to possess knowledge" and is used as a modal, as in the examples below.

de dollars sur dix ans. Ce plan	*ne saurait*	*cependant s'apparenter à un quelconque*
la brutalité qui éviscère et méprise	*ne sauraient*	*en être absentes. Les petits marins*
au creux du crâne ! Bobby Lapointe	*ne saurait*	*être absent d'un tel répertoire,*
avec une drôle de langueur. On	*ne saurait,*	*pour vous mettre en bouche, citer tous*
les joies de la libre entreprise	*ne sauraient*	*dissimuler l'absence de compassion*

The analysis of *connaître* provides another example of how a student can benefit from corpus consultation while not realising the full potential of the activity for discovery learning. While most of the 50 results confirm the explanation in the grammars that *connaître* is used in the sense of being acquainted with something or some one, in 9 the subject is not a person, and the sense is rather to experience, as in the examples below.

Véritable foire aux idées, il a	*connu,*	*depuis sa création, en 2001, un succès*
mondiaux, à Davos, le Forum social	*connaîtra*	*un temps fort particulier: le*
salaire moyen est de 30 cents. Le pays	*connaît*	*à nouveau la sécheresse, après deux ans*
ce n'est pas l'unique cause. Le pays	*connaît*	*une sécheresse très importante, une des*
Champagne-Ardennes et l'Île-de-France ont	*connu*	*un véritable enfer, qui a provoqué*
guérilla, et le Kosovo pourrait	*connaître*	*une dangereuse radicalisation. Il est*
montre la voie. La capitale toscane a	*connu*	*durant cinq jours un événement qui*
des " sages " n'a visiblement pas	*connu*	*un grand succès auprès des technocrates*
SPORTS Affaires : le foot français ne	*connaît*	*pas de trêve Football. Fernandez*

This use of *connaître,* which would be very appropriate in student essays in French, received no comment. The expression "*on connaît la chanson,*" which has nothing to do with singing but rather has the sense of "the same old story" also remained unexplored. Despite this, the student fulfilled her objective of using this exercise to solve her difficulty in distinguishing between the uses of these two verbs, particularly appreciating the example of both in one concordance line: "*les gens qui me connaissent savent que ma joie est tout intérieure.*" She concluded that the best problem-solving strategy for her consisted of an initial consultation of a grammar followed by a study of the much greater variety of examples of actual language use in the corpus, well illustrated in this case by her analysis of the occurrences of *savoir.* Although this essay reveals limited corpus consultation skills, it is interesting in the context of popularising corpus consultation in that the student still benefits from the activity, just as learners with limited dictionary skills still find solutions in the dictionary to some of the problems they encounter.

To complete the individual case-studies, it is interesting to compare the three essays on the preposition *en* in French, good examples of Kennedy and Miceli's comment on "the fatal lure of prepositions" (2001, p. 83) and also good illustrations of the students' independence of mind and autonomy in choosing to ignore the lecturer's advice to avoid selecting very common words such as this as search words. These students will be referred to as Students 1, 2, and 3. Student 1, deciding to analyse about 130 of the 1,530 occurrences, randomly selected 240, as she wished to remove the "obvious patterns" which she had observed during an initial trawl of all the results, namely *en* + year/season/month, *en* + country/continent/province, *en* + gerund, and *en* + number. Having eliminated these, she then concentrated her analysis on the remaining 136 occurrences. She observed that, in addition to the grammatical functions listed in the grammar book and dictionary which she consulted, the concordance revealed examples of expressions such as *la mise en* + noun, *en tout* + noun, and *en plein* + noun, information which, she noted, was not referred to in any structured way in either the dictionary or the grammar. The concordance results then aroused her interest in phrasal verbs, so she returned to the 1,530 occurrences, sorted them alphabetically, and studied the 91 phrasal verbs which this revealed. *S'en sortir, s'en prendre, s'en tenir,* and *s'en aller* proved to be the most common, none of which was given in the dictionary or grammar, although both gave examples of seven phrasal verbs with *en,* all of which were present in the corpus with the exception of *s'en ficher.* Like her colleague who had studied English verbs, she reflected that *s'en ficher* was understandably absent from this corpus, as the verb is "reserved for the spoken language." Despite such a successful journey of discovery in the corpus, she noted in what was generally a very positive conclusion that she found a number of aspects of the whole activity tedious, tiring, and laborious, in particular counting frequencies, deleting what she considered irrelevant concordances such as Aix-en-Provence, and reading from a screen.

Student 2 analysed a random sample of 300 occurrences of *en.* She began by analysing the use of *en* + gerund, noting that it would have been interesting if it had been possible to compare the use of the gerund with and without *en.* (This is possible albeit a little laborious.) Concluding the analysis of *en* + gerund, she observed, "While this aspect of grammar can be covered quite successfully without concordancing, the use of the software can certainly help cement in the learner's mind that which has already been covered." This student also noted the presence of phrasal verbs with *en,* but chose not to explore that discovery. She emphasised rather that the concordance results not only provided more examples than the grammar book, they acted as a sort of extension of it, in that, for example, while the grammar covered simple expressions of time such as *en quatre semaines,* the concordance also included related expressions such as *en l'espace de quatre semaines.* The discovery of phrases such as *mettre en* + noun and *la mise en* + noun was highlighted by the student as "one of the most interesting discoveries made from this concordance." Indeed, this may suggest that the students were right to ignore the lecturer's advice concerning prepositions, as it would be hard to discover these phrases unless one knew in advance that they existed or was given the search words by the teacher. This student was particularly proud to discover

several examples of a category not covered in the grammar book, namely *en* meaning "as" in the sense of "in the role of": "*Cécile Brune, en amante anglaise*;" "*Anne Kessler, en beauté chlorotique.*" Finally, the student noted that *en* is used as a preposition in a large number of adverbial and prepositional expressions, while the grammar gives only seven examples. As the student noted, "By quite a great deal this is the area in which the concordancer has proven itself to be the most useful, easily surpassing how the topic is dealt with in the selected grammar." This student's conclusion is extremely positive, with the corpus and concordancer clearly appreciated as a much richer learning environment than the grammar.

The third student who analysed *en,* also selecting 300 occurrences for analysis, found the experience much less worthwhile and stimulating, and it is important to note that differences in level of linguistic or analytical skills cannot provide the explanation for this. While Student 2 found the examples of *en +* gerund a useful complement to the examples in the grammar, Student 3 preferred the "very basic and easy to understand" examples in the grammar (all three students used the same grammar). Interestingly, this was the only student in the group who found the truncated concordance lines frustrating: "I did not find the concordance helpful in establishing these rules as the full context is not given." The generally very competent level of her corpus consultation skills suggests that she was aware of the possibility of accessing the full text in each case, but understandably did not consider this a realistic activity to undertake for each and every example. Only in the case of the adverbial and prepositional expressions and phrases did Student 3 accept the superiority of the concordance, as it provided many more than the seven examples in the grammar. Her conclusion, like that of most students, is that the combination of grammar and concordance is the ideal method, but, exceptionally in this case, that view is expressed in an almost grudging manner:

> I have concluded that the latter [the grammar] proved far more suitable in helping me to understand this point of grammar. I did although obtain a great deal of useful expressions that I know are up to date and widely used by native speakers from the concordance list.

The considerable variations in three learners' reactions to the very same activity cannot be analysed in any greater detail here, as no detailed information is available on their motivation or learning styles. The negative reaction of Student 3, however, is interesting in that it may suggest that not all learners will experience the interest and exhilaration which is evident in the reactions of many learners to the discoveries they make from corpus consultation.

The five brief case studies exemplify the main features of the 11 essays analysed here. In relation to the success or otherwise of the attempts at corpus analysis and the strategies used, there was a considerable amount of variation in the students' ability to explore the corpus to see what, if anything, it could add to the presentation in the course book or grammar, ranging from complete mastery and enjoyment of the exploratory nature of the activity, to a mechanical analysis of the results which added little to the student's knowledge of the language, at times missing points which seem worth commenting on. Basic command of the software did not seem to be at issue in the great majority of cases, although some obvious strategies were missed. S'*, for example, was not used in the study of pronominal verbs, although using *se* still provided a considerable amount of data for analysis. In general, however, all these students showed the ability to use Wordsmith Tools (Scott, 1999) sufficiently well to derive some benefit from their analysis of the results. It would seem rather that differences in motivation or learning styles may explain the considerable variation in the success of the activity. In addition to the variation in analytical ability, there was also considerable variation in the students' ability to reflect on the nature and limitations of the corpus, an ability which came easily to some students, but was totally lacking in others.

STUDENT EVALUATION

The students' evaluation of the activity in the final section of their essays reveals several clearly recurring patterns. Unlike Widdowson (2000, p. 7), these students do not call into question the authenticity of the

concordance results, despite a brief reference by the lecturer to this view. As one student wrote, "the French used in these articles is authentic, up to date, and relevant." The word "real" is also used to describe the corpus, in contrast to the invented examples in course books and grammars, which are described by one student as "unreal and sometimes stupid." The up-to-date nature of the contents, relating to news from just a few months previously, was appreciated by many students. For a number of them, this authenticity and familiarity made for what one termed "easy memorizing." However, while most were happy to accept the limited size of the corpus for their first attempt at corpus consultation, a few queried the choice of texts, revealing a preference for "simple colloquial language out of magazines or novels." The rich learning environment created by a large number of examples was also appreciated, in contrast to the limited number of examples given in course books, dictionaries, and grammars. As one student wrote, "The sheer amount of entries given by the software was impressive, and it made learning about the choice made [demonstrative pronouns in French] much quicker and easier when there were numerous examples to look at."

The positive features of discovery learning are mentioned by a number of the students. As one Erasmus student wrote: "Working out lexical or grammatical patterns on his or her own may help the learner to memorise problematic aspects better than it would be the case when 'spoonfed' with rules." Furthermore, although the term is not used, a sense of empowerment is evident in some of the positive evaluations. As Student 2 concluded, "I discovered that achieving results from my concordance was a highly motivating and enriching experience. I've never encountered such an experience from a textbook." These positive reactions suggest that corpora and concordancing certainly have their place in a language-learning environment focusing on learner autonomy and discovery learning.

Even where students' overall reactions were positive, however, they did not hesitate to express strongly worded views on the disadvantages. Firstly, none saw the corpus and concordancer as replacing the grammar book or course book. Indeed there was a general consensus in their conclusions that the grammar still had its place, with two students concluding strongly in its favour. As one expressed it, "old friends are still best." A third, who studied "since" in English, while greatly appreciating the corpus consultation, remained faithful to the grammar book as the ultimate authority, innocently destroying the whole edifice of descriptive corpus linguistics: "Besides, a grammar book can confirmed [sic] to us if the grammar structures employed in a text are really right or if there is a mistake." Secondly, the limitations of the small corpus were noted by a number of students, in the case of phrasal verbs with "come" for example, although as a result of the choice of very common words, few students made this criticism. It will possibly be solved in this course in the near future, if the size of the corpora and the variety of the texts increases. It will be interesting to see, however, if this influences in any way the third criticism, namely the students' perception of the analysis of the results as tedious, time-consuming, and laborious. It is possible that a larger resource will encourage them to choose less commonly occurring terms, in particular to avoid prepositions, but the task of classifying and counting will still have to be undertaken. It is difficult to reach a firm conclusion here, however, as prepositions are a common source of problems for learners in a number of languages, English and French for example. Fourthly, several students included in their section on disadvantages, comments on the need for training and appropriate analytical skills, showing uncertainty as to whether or not they had reached a sufficient level. As one student (one of the negative conclusions) noted, "In order to really derive any benefit from the use of corpora and concordancing, one would need quite intensive training and much practice." Fifthly, a number of students mentioned in their section on the disadvantages that the benefits of direct corpus consultation were limited to advanced students, and would be of no use to beginners because of their lack of comprehension of the text surrounding the search word. (This limitation could, of course, be solved by creating an appropriate corpus.) Finally, the lack of availability of corpora was cited by several students as a disadvantage. As they had been provided with an extensive list of available corpora in several languages, with easily accessible links through the module Web page, this perceived disadvantage raises important

questions which go beyond the issue of course design and lead us to ask what learning environment is necessary for learners if corpus consultation is to be integrated into their language learning. This issue will be dealt with in the following section.

SUGGESTIONS FOR FURTHER RESEARCH

Although the earliest attested use of corpora in the language classroom was as early as 1969, according to McEnery and Wilson (1997, p. 12) who refer to Peter Roe's use of it in Aston University, Birmingham, research focusing on the extent to which learners actually benefit from corpus consultation and analysis is relatively recent. Chambers (2004) has identified 12 studies since the early 1990s, including four quantitative studies (Cobb, 1997; Gaskell & Cobb, 2004; Stevens, 1991; and Yoon & Hirvela, 2004) and nine[2] qualitative studies including the present article (Bernardini, 2000, 2002; Chambers & O'Sullivan, 2004; Cheng et al., 2003; Johns, 1997; Kennedy & Miceli, 2001, 2002;[3] Sun, 2003; Yoon & Hirvela, 2004). While French (Chambers & O'Sullivan, 2004) and Italian (Kennedy & Miceli, 2001, 2002) are included in these publications alongside the languages involved in the present study, the majority of the studies involve learners of English. Although the results of these experiments are largely positive, they raise a number of issues which point the way for future research. Firstly, they focus on the context of university education, understandably as they can all be classed as action research, with the students of the researchers forming the subjects of the experiment, as in this article. There is thus a need for studies involving other sectors of education, possibly in the form of concordances prepared by the teacher rather than direct access.

Secondly, they mainly use written corpora as a resource, leaving scope for investigations of the use of spoken corpora. Braun (in press) describes an initiative in which an extract from a spoken corpus is used as the basis for a language class, with concordances of specific items which are encountered providing additional examples. The learners' activities and reactions are not, however, included in her study.

Thirdly, the existing publications raise several issues which merit further study: for example, the benefit of direct consultation of corpora by learners as opposed to consultation of concordances provided by teachers; the learner strategies used in corpus consultation and analysis, and the teacher's role in providing guidance; the role of corpus consultation in specific areas of second language acquisition such as the acquisition of vocabulary, grammar, and writing skills; and the role of the corpus alongside other resources. Although the existing studies involve a variety of research methods, including accounts of the activities undertaken by the learners and presentations of the results of their work with the corpora, learner evaluations of the activity, sometimes but not always questionnaire-based, observations by the researcher, trialling, and think-aloud protocol, there are few examples of each of these, and there is thus a need for more quantitative and qualitative studies. Finally, there is clearly scope for studies involving languages other than English, particularly now that corpora are available in several languages.

The increasing availability of vast corpora and their potential in language learning raises the fundamental issue of the ease with which learners can access these resources and benefit from them. There is as yet little evidence to suggest that the learners in the various studies cited here move on to regular corpus consultation in the course of their degree programme.

This situation suggests that classroom-based activity, which has so far provided the setting for all of the publications related to direct corpus consultation by language learners cited in this study, cannot on its own lead to any large-scale integration of corpus consultation into language studies. In addition to this, the student evaluations quoted above suggest that they do not easily make the transition from an initial training course, however much they appreciate it and benefit from it, to regular corpus consultation using the large Web-based corpora which are available. University language resource centres or the writing centres common in universities in the United States of America may provide a solution here, as they can serve as repositories of resources suitable for learners, and also as providers of learner training and

guidance. This is not to say that each individual centre should have to create its own resources. Learners' needs are sufficiently similar in many universities for shared resources to be appropriate. And it is possible that the most appropriate resources for all but the most proficient students in corpus consultation and analysis will not be the huge corpora of hundreds of millions of words which have been created for research purposes, but of corpora created for learners, such as BNC Baby to give one example (Burnard, 2004). This four million word corpus, extracted from the British National Corpus and consisting of one million words each of written academic and journalistic English, fiction and spoken English, is part of a project involving the Oxford Text Archive and the Open University in the UK. Similarly, one-million word corpora of written academic and journalistic French are currently being created at the University of Limerick, and will be deposited with the Oxford Text Archive, the journalistic corpus in May 2005 and the academic corpus in 2006. If resources such as these and other relevant corpora were replicated in other languages, learners could benefit from accessing them independently and using them alongside the traditional resources of dictionary, course book, and grammar. Centres providing resources and guidance, which are at least as important as the classroom in a language-learning environment which aims to promote learner autonomy, could then become the focus for future research projects involving corpus consultation by learners.

CONCLUSION

The example of this course, together with the evidence from several other publications on related topics cited in this study, suggests that corpus consultation as a language-learning activity has many positive features, particularly in a language-learning environment which favours learner autonomy and discovery learning. It is, however, the disadvantages noted by these learners which are of particular interest here, as they provide a list of problems to be solved. Issues relating to the size and nature of the corpus are no doubt the easiest to remedy, although a delicate balancing act is required to ensure that the size of the corpus used for initial training does not unduly increase what they already perceive as the laborious and tedious analytical work, even when only a randomly selected concordance of about 150 occurrences is involved. An increased allocation of time for training could provide the answer, although probably not within the context of most language degree programmes, where the curriculum is already under pressure from the competing disciplines of literature, cultural studies, area studies, linguistics, and language learning per se. Finally, as we have seen in the previous section, corpora and concordancing is an area where the whole language-learning environment has a role to play, not just the classroom but also facilities for independent and collaborative learning, and where there is thus ample scope for further research and development.

NOTES

1. As only one of the students was male, all students are referred to as "she" to preserve their anonymity.

2. Yoon and Hirvela's analysis, being both quantitative and qualitative, is included in both categories.

3. Two reports based on different aspects of the same study.

ACKNOWLEDGMENTS

ABOUT THE AUTHOR

Angela Chambers is professor of applied languages and director of the Centre for Applied Language Studies in the University of Limerick, Ireland. She has co-edited two books on computer assisted language learning. Her current research focuses on the role of corpora in language learning.

E-mail: Angela.Chambers@ul.ie

REFERENCES

Aston, G. (1997). Small and large corpora in language learning. In B. Lewandowska-Tomaszczyk & J. P. Melia (Eds.), *Practical applications in language corpora* (pp. 51-62). Lodz, Poland: Lodz University Press.

Benson, P. (2001). *Teaching and researching autonomy in language learning.* London: Longman.

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 225-234). Frankfurt: Peter Lang.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165-182). New York: The Edwin Mellen Press.

Bowker, L. (1998). Using specialised monolingual native-language corpora as a translation resource: A pilot study. *Meta, 43*(4), 631-651. Retrieved December 9, 2004, from http://www.erudit.org/revue/meta/1998/v43/n4/002134ar.pdf

Braun, S. (in press). From pedagogically relevant corpora to authentic language learning contents. *ReCALL, 17*(1).

Burnard, L. (Ed.). (2004). *BNC Baby* [CD-Rom]. Oxford, England: Research and Technology Service, Oxford University. Available at http://www.natcorp.ox.ac.uk/babyinfo.html

Burnard, L., & McEnery, T. (Eds.). (2000). *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora.* Frankfurt: Peter Lang.

Chambers, A. (2004). Popularising corpus consultation among language learners and teachers. Paper presented at TALC (Teaching and Language Corpora) conference, University of Granada.

Chambers, A. & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL, 16*(1), 158-172.

Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System, 31*(2), 173-186.

Cobb, T. (1997). Is there any measurable learning from hands on concordancing? *System 25*(3), 301-315.

Davies, M. (2000). Using multi-million word corpora of historical and dialectal Spanish texts to teach advanced courses in Spanish linguistics. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 173-185). Frankfurt: Peter Lang.

Dodd, B. (1997). Exploiting a corpus of written German for advanced language learning. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 131-145). London: Longman.

Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME, 17*, 97-109.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors. *System 32*(3), 301-319.

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108-137). Houston, TX: Athelstan.

Johns, T. (1986). Micro-concord: A language learner's research tool. *System, 14*(2), 151-162.

Johns, T. (1988). Whence and whither classroom concordancing. In T. Bongaerts, P. De Haan, S. Lobbe, & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9-27). Dordrecht, The Netherlands: Foris.

Johns, T. (1997). Contexts: The background, development, and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology, 5*(3), 77-90. Retrieved December 9, 2004, from http://llt.msu.edu/vol5num3/kennedy/

Kennedy, C., & Miceli, T. (2002). The CWIC project: Developing and using a corpus for intermediate Italian students. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 183-192). New York: The Edwin Mellen Press.

Kettemann, B. (1995). Concordancing in stylistics teaching. In W. Grosser, J. Hogg, & K. Hubmeyer, (Eds.), *Style: Literary and non-literary. Contemporary trends in cultural stylistics* (pp. 307-318). New York: The Edwin Mellen Press.

Kettemann, B., & Marko, G. (Eds.). (2002). *Teaching and learning by doing corpus analysis.* Amsterdam; New York: Rodopi.

Louw, B. (1997). The role of corpora in critical literary appreciation. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles, (Eds.), *Teaching and language corpora* (pp. 240-251). London: Longman.

McEnery, T., & Wilson, A. (1997). Teaching and language corpora. *ReCALL, 9*(1), 5-14.

Roe, P. (2000). The ASTCOVEA German Grammar in conText Project. In B. Dodd (Ed.), *Working with German corpora* (pp. 199-216). Birmingham, England: University of Birmingham Press.

Scott, M. (1999). *Wordsmith Tools version 3*. Oxford, England: Oxford University Press.

Stevens, V. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. In T. Johns & P. King (Eds.), *Classroom concordancing: English Language Research Journal 4* (pp. 47-63). Birmingham, England: Centre for English Language Studies, University of Birmingham.

Sun, Y-C. (2003). Learning process, strategies and Web-based concordancers: A case-study. *British Journal of Educational Technology, 34*(5), 601-613.

Thompson, P. (2004). The University of Reading master's program in applied linguistics [Web page]. Retrieved December 9, 2004, from http://www.rdg.ac.uk/AcaDepts/cl/slals/maal.htm

Tribble, C., & Jones, G. (1990). *Concordances in the classroom: A resource book for teachers.* Harlow, England: Longman.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 3-25.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes towards corpus use in L2 writing. *Journal of Second Language Writing, 13,* 257-283.

Zanettin, F. (2001). Swimming in words: Corpora, translation, and language learning. In G. Aston (Ed.), *Learning with corpora* (pp. 177-197). Houston, TX: Athelstan.