ReCALL **16** (1):158–172. © 2004 Cambridge University Press
DOI: 10.1017/S0958344004001211 Printed in the United Kingdom

Corpus consultation and advanced learners' writing skills in French

ANGELA CHAMBERS and ÍDE O'SULLIVAN
Department of Language & Cultural Studies, University of Limerick, Limerick, Ireland
(*e-mail: {Angela.Chambers,Ide.Osullivan}@ul.ie*)

Abstract

In the rapidly changing environment of language learning and teaching, electronic literacies have an increasingly important role to play. While much research on new literacies focuses on the World Wide Web, the aim in this study is to investigate the importance of corpus consultation as a new type of literacy which is of particular relevance in the context of language learning and teaching. After briefly situating the theoretical and pedagogical context of the study in relation to authenticity and learner autonomy, the paper describes an empirical study involving eight postgraduate students of French. As part of a Masters course they write a short text and subsequently attempt to improve it by using concordancing software to consult a small corpus containing texts on a similar subject. The analysis of the results reveals a significant number of changes made by the learners which may be classified as follows in order of frequency: grammatical errors (gender and agreement, prepositions, verb forms/mood, negation and syntax); misspellings, accents and hyphens; lexico-grammatical patterning (native language interference, choice of verb and inappropriate vocabulary); and capitalisation. The conclusion notes that the situation in which these students found themselves (i.e. faced with a text on which the teacher had indicated phrases which could be improved) is replicated in many cases every day, and suggests that corpus consultation may have a useful role to play in the context of interactive feedback, particularly in cases where traditional language learning resources are of little use.

1 Introduction

Developments in Information and Communications Technologies have not just simply provided new ways of teaching the traditional literacies of reading and writing; but have themselves become a new type of literacy, producing a rapidly changing learning environment and presenting a major challenge to learners, teachers and researchers. The concept of literacy now includes not only the knowledge and skills which are traditionally associated with that concept, but also the ability to select, evaluate and use the electronic tools and resources appropriate for the activity which is being undertaken. It is not surprising that much of the literature relating to these new literacies focuses on the use of the World Wide Web (Warschauer, 1999; Cope & Kalantzis, 2000; Kasper, 2000;

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

Murray, 2000; Richards, 2000), and yet it is clear to those researching in the area of learner access to corpora that for language learners at an advanced level, corpus consultation skills may form a particularly important new literacy. Despite a number of recent studies (Cobb, 1997; Bowker, 1998; Kennedy & Miceli, 2001), there is a dearth of empirical studies which could form the basis of developments in language pedagogy.

This article presents the results of a pilot study as part of a research project involving students of French at Masters level, who are given the opportunity to improve a short text which they have written in French by using concordancing software to consult a small semi-specialised corpus of texts by native speakers on a similar subject. The aim is to discover if corpus consultation skills can improve aspects of their writing skills, and what particular aspects of writing skills are involved. The experiment includes changes relating to morphological and syntactical accuracy, but focuses primarily on the learners' ability to use – and more particularly to learn to use – the type of lexico-grammatical patterns found in the writing of native speakers on the same subject. It is this aspect of language which receives most attention from researchers. As Tognini-Bonelli (2001:24) observes: "What is consistently shown by corpus work is the strict correlation between lexical and grammatical choices which extend the boundaries of the initial unit". After a brief examination of the pedagogical context, relating in particular to authenticity and autonomy, the corpus and the empirical study will be described and the results examined.

2 Pedagogical context: authenticity and autonomy

The theoretical and pedagogical context of this study draws on a number of areas within second language acquisition research in addition to corpus linguistics itself, most notably authenticity and autonomy. Despite Chomsky's criticism of the representativity of the corpus (1957:159), it is now generally accepted as a valuable source of information on actual language use. However, once one begins to consider the use of the corpus by learners as a resource, the situation becomes more complex. In the well-known debate between Widdowson (2000; 2001), Stubbs (2001) and others, the corpus has been criticised for lacking authenticity: "What is not taken into account is the pedagogic perspective, the contextual conditions that have to be met in the classroom for language to be a reality for the learners" (Widdowson, 2000:7). However, counterarguments have been put forward, drawing on the concept of authenticity as involving an activity or a process rather than an inherent quality, a concept developed by Widdowson himself (1983:30) and later elaborated by van Lier (1996). According to these counterarguments, the learner recontextualises or authenticates the corpus in the process of consulting it. Mishan (forthcoming), pointing out the necessity of giving learners access to what Widdowson (1983:30) would call 'genuine' texts, argues that means of authenticating corpora for teaching and learning purposes must be found, and puts forward a number of recommendations. Furthermore, Gavioli and Aston (2001:241) note that:

> […] there is an alternative way of authenticating discourse, by adopting the role of an observer (Aston, 1988). While the participant interacts with the text as an intended recipient, the observer views this interaction from the outside, adopting a critical, analytic perspective. Observer as well as participant roles can allow learning: observation allows strategies of interaction to be noticed, while participation

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

allows such strategies to be tested.

This implies that the learner's interaction with the corpus, involving the observation and then the selection and use of items which have been observed, can be a valid language-learning activity. A detailed consideration of the debate on the authenticity of corpora and corpus consultation is beyond the scope of the present study, but it is important to note that in this case Widdowson's main criticism of corpora as "decontextualized language" (2000:7) does not apply. The texts are not totally decontextualised here, as we shall see. The students are already familiar with this type of writing and with the context in which the texts were published, even with many of the actual texts, and they are consulting the corpus to assist them in the production of texts which deal with topics similar to those in the corpus.

The second area of research within second language acquisition which is of particular relevance to this study is learner autonomy. In consulting the corpus to assist them in writing French, the students are part of a "paradigm-shift from a process-control model of language instruction to an information-resource model in which learners explore the language for themselves and the role of instruction is to provide tools and resources for doing so" (Cobb, 1997:301). McEnery and Wilson (1997) emphasise that even when consulting a concordance printout, let alone a corpus, the student becomes a researcher. For them, corpus-based learning is synonymous with discovery learning, divergent learning, mediated learning and a move away from teacher-led learning, to a situation in which the teacher "may point the general direction for the students, that is, tell them what points to examine and what questions to ask. But it is the students who lead themselves and one another through the learning process" (1997:6). At a later stage in the process, the students could of course decide for themselves what questions to ask and become even more self-directed. In other words this scenario corresponds to many of the central features of learner autonomy. As Johns pointed out as early as 1986, corpora and concordancing transform language learning into "a species of research activity" (1986:160) and transform the role of the teacher into that of adviser or guide. Indeed, the inspiration for this particular study came partly from a comment by Johns in this same article:

> The third, and most important, potential use of an interactive concordancer is as a learning resource to be used freely by students on their own initiative with the role of the teacher restricted to suggesting points at which it may help to solve learning difficulties. One possibility with which we have experimented is its use in helping students to correct their written work, some mistakes being underlined and a 'C' placed in the margin signifying 'You have used this word in a way which is different from how an English person would use it: if you get a concordance of the word you should be able to work out a suitable correction for yourself (1986:161).

Johns is clearly referring to his own students here, who would have been familiar with the consultation of concordance printouts and no doubt influenced by his infectious enthusiasm for concordancing. However, we have found that students require considerable guidance if they are to benefit from corpus consultation, and a number of recent research publications focus on this area (Kennedy & Miceli, 2001; 2002).

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

3 The small semi-specialised corpus

The two most important decisions concerning the corpus to be provided concern the nature of the texts and the number of words. It was decided to provide a corpus of texts on topics similar to those studied in this part of the course, which could be updated from year to year. As the students' task is to write a commentary on a text, the corpus would ideally contain similar commentaries written by native speakers, but this was not possible. Therefore, it was decided to include texts in French relating to the history and development of the French language and to current issues relating to the language, written by educated and informed native speakers, thus providing the students with models of well-written contemporary standard French. The resulting untagged corpus, 165,000 words at the time the study was carried out, contains two sub-corpora. The first (125,000 words from *Le Monde*) deals with current issues relating to the French language, and is divided into two equally sized sub-corpora, the first dealing with the defence of the French language in general and the second more specifically with the controversy surrounding the non-ratification by France of La Charte européenne des langues régionales ou minoritaires. The second sub-corpus (40,000 words) includes a text on the history and development of the French language (Leclerc, 2000). This is part of a substantial Web-based text which is recommended to the students as an introduction to the subject. It is academic writing in a didactic rather than a research context. A number of issues relating to the appropriateness, as well as the availability of these texts, were taken into consideration. The Leclerc text was deemed appropriate as it dealt with the subject of the commentary in an academic context, and the journalistic articles were included as they briefly presented and argued opinions on issues related to the same current issues which were the subject of the students' written work. In that sense they were closer to the student writing (a short commentary analysing a French text) than the longer academic text. Availability was not an issue, as permission had been acquired from those holding copyright to include these texts in the corpus. It is intended to add a number of academic articles on relevant topics to the corpus in the future, when permission is obtained from those holding the copyright. This corpus corresponds to Aston's definition of a small semi-specialised corpus as typically between 20,000 and 200,000 words (1997:54). To paraphrase Aston (1997:55–58), small corpora have advantages over their larger counterparts for such uses: they are easier to manage; easier to become familiar with; easier to interpret; easier to construct; easier to reconstruct; more clearly patterned; and their limits are clearer. The corpus analysis tool or concordancer being used for the purpose of this study is WordSmith Tools.

4 Empirical study – methodology

4.1 Methodology and context

As studies involving corpus consultation by learners are few in number, there is as yet no clearly defined methodology on which one can rely. Published studies include quantitative research (Cobb, 1997; Bowker, 1998), qualitative research focusing on learner strategies and teacher guidance (Kennedy & Miceli, 2001; 2002) and a longitudinal case study of learners (Turnbull & Burston, 1998). The aim of this study is to obtain information on

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

the type of changes made by the learners in order to complement existing research and to assist us in presenting future guidance to students. In this sense the study can be classified as action research, with the aim of informing future professional practice.

The students involved in the study are eight advanced learners of French following the MA in French at the University of Limerick in the 2002–2003 Academic Year. Two features of the programme make it suitable for this empirical study, in that it includes both a module entitled La langue française aujourd'hui, involving the study of the history of the French language and current issues relating to it, and a module entitled *Technologies de l'Information et de la Communication et études de langues*, focusing on ICT and language studies, and including the use of corpora in language learning. As part of the module requirements for *La langue française aujourd'hui*, students complete a written assessment. Indeed, the programme involves many writing tasks such as commentaries and essays; therefore, students are highly motivated to participate in a project with the aim of improving their writing skills.

4.2 The stages of the study

1. Written assessment
2. Training in corpus consultation skills
3. Empirical study – two-hour session

Stage 1: Written assessment. This consists of a short commentary in French (600 words) on a text in French relating to an aspect of the French language. Students are asked to situate the passage in its context, examine the issues involved and comment on the author's stance. The text chosen in 2002 was an editorial from *Ouest-France* on 18 June 1999 concerning France's refusal to ratify the European Charter for Regional or Minority Languages. The commentary is written in a two-hour supervised session, during which students may consult dictionaries. At this time none of them has had any training in corpus consultation skills (this is confirmed later in the questionnaires). The marking process took the form of underlining errors/mistakes[1] and placing x in the margin to indicate a basic inaccuracy, for example in gender, agreement, verb forms etc. In addition, words and phrases which did not correspond to standard written French were underlined. No explanation or feedback was provided at this stage, as would normally have been the case. Instead the students were informed that this feedback would take the form of an empirical study, and their agreement was confirmed. The commentaries, which were then typed and e-mailed to the researcher, were reformatted to leave multiple spacing of approximately three centimetres between each line of text so that students could easily write any changes they made above the original version. The underlining made by the lecturer as part of the original grading was added to the printed version, and each underlining was numbered.

Stage 2: Training in corpus consultation skills. As part of the module *Technologies de*

1 It is not always possible for the researcher/lecturer to distinguish between errors, where the student is unaware of the correct form, and mistakes, which may be the result of hastiness etc. Henceforth, the terms 'errors' will be used to refer to both errors and mistakes unless otherwise stated.

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

l'Information et de la Communication et études de langues, the students are introduced to corpora and concordancing. Over a three-week period, this part of the module consists of a one-hour lecture and a one-hour session in a computer laboratory each week. The first lecture includes an introduction to corpus linguistics and to the history of corpora, a brief account of the types of research based on corpora, and an evaluation of the current situation concerning corpora and corpus-based research involving the French language. The other two lectures provide the students with information on the types of project which they may choose for the assessment of this module, namely a corpus-based discourse analysis or a study of how a corpus could be analysed using a concordancer to assist a learner in mastering an aspect of French language use. In the laboratory sessions students are shown the different uses of the concordancer in certain areas of language study. In the first week they become familiar with the functions of Wordsmith Tools, in particular how to devise strategies to help them find what they are looking for, and ways to fine-tune their searches, which would be useful in helping them discover words and how they function with others. Various applications of the corpus and concordancing software for language learning are also included, not just their application for improving writing skills. In the second session students are shown how to create a corpus, and in the third session they work on their projects, which they choose as part of the assessment for this module, creating a corpus and analysing it. Thus, while the main focus of these classes is not to provide the basis for this study, it nonetheless makes it possible by providing a short course in corpus consultation skills.

Stage 3: Empirical Study. This took the form of a two-hour session in the computer laboratory in the week following the lectures and seminars described above. The session was organised in three stages. After an introduction of ten minutes outlining what they had to do, the students were allowed up to 100 minutes to consult the corpus and make any changes they wished, and then given ten minutes to fill in the evaluation forms. The only assistance provided during this process was technical (relating to the use of the software); no linguistic assistance was provided. During this time the students were asked to try to improve those parts of the texts which had been underlined. They had to study the words and phrases underlined in their text, search the corpus for ways to help them improve these phrases, look at the results of the concordance and decide if they wanted to change their text. If so, they had to

a. Write the correction or the new word or phrase above the text of the original essay (only indicating those changes which resulted as a consequence of using the corpus).

b. Fill in the form which was attached with details of the search word they used, the results of the concordance and what they discovered.

Students were also asked not to limit themselves to the phrases which had been underlined. If they had time they were asked to choose a phrase/sentence from their text that they were unhappy with and see if there was any way in which they could use the corpus to improve this. None took advantage of this suggestion, and, as we shall see, one student noted in the evaluation questionnaire that it would be difficult to improve an unmarked piece.

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

5 Analysis of results

5.1 Classifying the errors and changes

The commentaries produced using traditional resources were compared with those improved with the aid of a corpus, in order to evaluate the influence which corpus consultation had on the students' writing skills in French. The evaluation forms completed by the students were also analysed in order to gain information on their reactions to this process. It is important to note that the number of changes made was not confined to a small number of students – see Table 1.

The texts were then examined and changes were classified according to the categories indicated in Table 2. For example, *noun + verb* indicates that the noun identified was either used with an incorrect verb or a verb that was overused, or used in the wrong context, and that the student proceeded to do a concordance search on a noun to help them find an alternative verb to be used in that context. This table classifies the changes made in order of frequency, and indicates if the resulting change was positive, negative, or neither of these. The most frequent changes were in the areas of grammar (31) followed by misspellings (23), lexico-grammatical patterning (18), and capitalisation (7). The smallest number of changes was made in the areas of register (1) and punctuation (1). It was not always obvious how certain changes should be categorised. For example, *au même temps* (at the same time) and *dans une façon positive* (in a positive way) have been placed in the category of prepositions, although there is very probably an element of native language interference involved as well. However, they were classified under grammatical errors involving prepositions. This enabled us to include under native language interference only examples involving lexico-grammatical patterning.

The analysis of the changes will focus firstly on those errors which could have been corrected using traditional resources and then on changes in areas where traditional resources would have been of little or no use.

5.2 Correcting basic errors

Gender and agreement were involved in fifteen cases where the corpus was used to check the correct form, with thirteen successful changes, where nouns had been assigned the incorrect gender (10), or where the agreement between the noun and adjective was incorrect (3). Given that the errors involved here are of a very basic nature, it is impossible to prove that the students could not have produced the correct version simply by looking carefully at the word or phrase. However, the students do appear to have followed the guidelines to include information only on changes made using the concor-

Table 1. *Overview of results*

Summary of results	
Total number of changes	85
Positive changes	64
No changes	17
Negative changes	4

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

Table 2 *Total number of changes made according to type of error*

Classification of error/mistake	Positive changes	No changes	Negative changes	Total no. changes
Grammatical errors				
Gender and agreement	13	2	0	15
Prepositions	5	0	0	5
Verb forms/mood	5	0	0	5
Use of the negative	4	0	0	4
Syntax	1	1	0	2
Total	28	3	0	31
Misspelling/accents/hyphens				
Total	16	6	1	23
Lexico-grammatical patterning				
Native language interference	8	1	0	9
Noun + Verb	3	3	0	6
Inappropriate vocabulary	1	1	1	3
Total	12	5	1	18
Capitalisation				
Total	6	1	0	7
Other				
Total	2	2	2	6
Total no. of errors/mistakes	64	17	4	85

dancer. Many of them spotted errors or mistakes, in particular accents, and changed them without consulting the corpus and without commenting on these on the attached forms. These are not included in the analysis. Where basic errors were corrected using the concordancer, this could of course have been done with the help of a dictionary, but it is interesting to note that, as we shall see, students commented in their evaluation forms that corpus consultation was a quick way of checking such issues.

Errors involving the subjunctive, verb forms, negation and syntax were also corrected with the help of the concordancer. For example, one student discovered that *le fait que* triggers the subjunctive. Another used the concordancer to distinguish between past participles and the present tense (*remet* and *remis*), and a third used it to check if a verb should be used in its reflexive form or not. Similarly the four attempts made to correct the use of the negative were successful, as shown in Table 3.

Two attempts were made to correct syntax, one of which was successful. Firstly, the phrase *ce que c'est la démocratie* was changed to *ce qu'est la démocratie*. The student also discovered an alternative way of expressing this: *en quoi consiste la démocratie*. In contrast to this, the phrase *En s'y opposant la position logique* was not changed to *En s'opposant à la position logique*. Having carried out a search of opposer, the student

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

Table 3 *Positive changes in the category of use of the negative*

Incorrect expression	→	Corrected expression
pas sans succès	→	non sans succès
ne pas ratifié	→	ne pas ratifier
il a vu aucun danger	→	il n'a vu aucun danger
n'avait pas vu aucun danger	→	n'avait vu aucun danger

failed to identify the problem and thus resolve it. If he had carried out a search of the lemma opposer and its forms using the wildcard facility, he would have discovered 13 examples of *s'oppos* à*, and the pattern of using the preposition *à* would have been very obvious.

In the case of misspellings, accents and hyphens, traditional resources could have been used to correct the errors, but corpus consultation provided the answer in some cases. In total, students attempted to change 23 misspellings, accents, and hyphens, 16 of which were successfully corrected. It is clear from the results, however, that the dictionary is a simpler alternative here. While the single consonant in the misspelt constitutionelle is easily corrected using constitution* as the search word, the concordancer is of little use in many misspellings such as the missing accents on words such as *refuter* and *lacheté*.

5.3 Capitalisation

Checking the corpus for capitalisation proved helpful as students were able to retrieve information which could not be found in the dictionary, namely when exactly to use capital letters in expressions such as, *président de la République* and *la Charte européenne*. Six of them had spelt *président* and *européenne* with capital letters and all six instances were rectified using the concordancer. One student noted that this has now "reminded me of the difference in French *vis-à-vis* English" with regard to capitalisation. Only one example of the misuse of a capital letter remained uncorrected, *la situation Corse*, as the student was unable to identify the problem. Having carried out a search of *la situation Corse*, and finding no occurrences, the student abandoned the search. A concordance search of the word *corse* would have revealed the pattern of using a capital letter to refer to the country (94 occurrences), and a small letter in the case of the language (55 occurrences) and adjectives (60 occurrences).

5.4 Lexico-grammatical patterning

The initial stimulus for this study arose from an awareness that corpora and concordancing software provide an excellent means of identifying lexico-grammatical patterning. Although as we have seen, changes in other aspects of language use were more common, the identification of appropriate lexico-grammatical patterns nevertheless provides some of the most interesting results, as they could not have been obtained easily from a dictionary, grammar or course book. Native language interference was clearly

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

involved in nine cases, eight of which were corrected by consulting the corpus. Examples are included in Table 4, with the search word underlined in each case. They highlight the importance of using the noun as a search word in order to find the correct lexico-grammatical patterning.

In the corpus it is interesting to note the lexico-grammatical patterning of the word question, which can be used as the equivalent of the English word 'question', or in a broader sense meaning issue or problem. A search on this word produced 85 results, the clearest patterns of which are summarised in Table 5.

The limited number of results compared with what one would find in a larger corpus does not pose a major problem here. What is interesting to note is that for advanced learners a limited concordance result such as this can provide sufficient information to assist them in improving their writing skills.

Another category of error within lexico-grammatical patterning involves the use of an incorrect, inappropriate or overused verb to accompany a particular noun. This is referred to as Noun + verb. One such verb in French which tends to be overused by native English speakers is *dire* (to say), with students tending to overuse expressions such as *l'auteur dit*. To encourage them to discover a greater variety of reporting verbs the researcher recommended, as part of the tips on the instruction sheet, doing a concordance search on nouns such as *Chirac, Jospin, Poignant, président, premier ministre, l'auteur, le texte, le rapport*. Thus in one case where *l'auteur dit* had been underlined, a search of the noun *auteur* enabled the student to propose alternatives such as *rejeter, raconter, affirmer, suggérer,* and *souligner*. In another case the sentence *la phrase est comme si le Président n'avait pas considéré l'importance de sa décision* was replaced by *les mots véhiculent la notion que le Président n'avait pas considéré l'importance de sa décision*, as a result of the use of *texte* as a search word.

5.5 Prepositions

One category within grammatical errors which is not easily corrected using traditional resources is the use of prepositions in general, and the prepositions which accompany

Table 4 *Positive changes in the category of native language interference.*

Incorrect expression	→	Corrected expression
D'abord il y avait la question de la signer	→	D'abord il a été question de la signer…
La question qu'il faut se demander	→	La question qui se pose maintenant
hors de la question	→	hors de question…
son ton d'incrédulité	→	le ton d'incrédulité
en personne du Premier Ministre	→	en la personne du Premier Ministre

Table 5 *Lexico-grammatical patterning of question.*

La/une question de (33)
La question se pose/sera posée – il faut se poser la question….. (12)
Il est/a été question de (5)
Re/mettre en question (5)

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

verbs in particular. There are no simple rules which can be learned to determine the choice of preposition in a given phrase, and checking each verb in the dictionary can be tedious. All five attempts to correct prepositions were successful. Thus *Chirac a une tendance de changer son avis* became *Chirac a tendance à changer d'avis*, as a result of a search of tendance. Similarly *l'auteur fait mention à* was corrected to *fait mention de* by carrying out a concordance of *fait mention*. One student was convinced that the expression *au même temps* was correct, a very common error at undergraduate level which in this case had persisted; a concordance of *même temps* revealed otherwise (twelve occurrences of *en même temps*). This is an example of where the use of the incorrect preposition is very possibly the result of native language interference. One other such example is *dans une façon positive*, which was easily corrected by carrying out a search of *façon*. This student commented that consulting the corpus was "good for unlearning errors".

6 Student feedback

On completion of the above activity, students were asked to provide written feedback on the changes which they made, and to complete a questionnaire commenting on the corpus consultation process and on their evaluation of this activity as a means of improving writing skills. Firstly, they were asked to answer a number of questions providing information on their computer literacy in general, and more specifically any experience they may already have had with corpora or corpus consultation (none in the case of this study). They were then asked to comment on how helpful they felt that working with the corpus was in improving their writing skills in French (Q7), to evaluate the training provided, if this was sufficient to help them complete the task they had been given (Q8), if not, why not (Q9), if they would consult corpora in the future to help their writing skills (Q10), and if not, why not (Q11). Finally they were asked to comment on what developments in the University would help them to use corpora to improve their writing skills (Q12) and to give details of what they believe to be the main advantages and disadvantages of consulting a corpus to improve their writing skills in French (Q13A and B). The principal results of this evaluation are summarised here.

Responses to Question 7 (To what extent do you feel that working with this corpus has helped improve your writing skills in French?) were generally positive. All found this activity of some help in improving their writing skills: one student replied that it was extremely helpful, four replied that it was very helpful and three replied that it was helpful. None of them found it of little or no help at all. Comments added in response to this question included:

- "Good for 'unlearning' errors – i.e. making me search to find out what is wrong with something that I have convinced myself (somehow!) is correct."
- "Useful to check common usage or verbs/expressions that I might be unsure of."
- "It has opened my eyes."

Seven of the students reported that the training was sufficient to help them complete the task they had been given (Q8), while one reported that he was unsure but suggested that: "group work with a proficient tutor per group" would improve the training (Q9).

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

When asked if they would consult corpora in the future to help their writing skills (Q10), five said that they would while three of the students were unsure. It is evident from the questionnaires that this hesitation to use corpora again in the future relates primarily to external factors such as the availability of corpora and the software (all three comment on the need for greater availability of corpora in their recommendations), and not to their reaction to the corpus consultation process (two of these students found it helpful, while one found it very helpful). It is interesting to note that one student who found the corpus very helpful and would use corpora again in the future to improve writing skills, specified that it was a qualified yes: "The errors need to be highlighted first. I don't believe I could significantly improve an unmarked piece." In response to Question 12 (What developments in the University would help you to use corpora to improve your writing skills?) the following recommendations were made:

- Increase the availability of corpora
- Allow greater access to the software, possibly on a rental basis
- "Need more hands-on time in order to become more proficient and incisive and observant of the software"
- More demonstration on the uses of corpus consultation skills

In response to Question 13(A), the following is a list of the advantages which the students felt were to be gained from using the corpus to improve their writing skills in French. Corpora allow you to:

- Write accurate and good French, to highlight errors and to give options in fixing errors
- Avoid the use of Anglicisms
- Avoid the use of franglais
- Autonomously discover your errors
- Access large bodies of texts
- See words and expressions in context
- Confirm an idea/hunch/phrase
- Find alternatives for any word you tend to overuse
- Check a grammatical query
- Check vocabulary/gender/syntax

However, on the negative side the following comments were made in response to Question 13(B):

- It is time consuming
- You have to have a clear idea of what alternatives to look for so your own subject knowledge has to be pretty advanced
- Not readily available
- It doesn't always give you suggestions on how to change a phrase
- There isn't always an occurrence of the word required

In summary, the students' evaluations reveal both positive and negative reactions,

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

suggesting that corpus consultation can be of help in a number of areas, but that, for various reasons involving training and the availability of resources, it does not automatically provide a miracle cure for all the problems of language learners. As Gavioli and Aston point out (2001:245):

> […] more research is needed into the design of corpus-based activities, and into their selection and grading. What is it worthwhile to look at in concordances, for whom, and from what corpora? How can corpus-based activities best be integrated with 'normal' language teaching, at different levels of proficiency? How can learners (and teachers) best be trained to profit from these resources?

7 Conclusion

This study reveals a number of positive results arising from direct learner access to corpora and also raises a number of important issues in relation to the optimal conditions for such access. On the positive side, the researchers' assumption that the identification of appropriate lexico-grammatical patterns would figure in the results was confirmed, with native language interference being reduced as a result of corpus consultation, and idiomatic phrases adopted which the students would have had difficulty producing as a result of consulting a dictionary, grammar or course book. This is not, however, the most common result of the corpus consultation, and an unexpected result was the enthusiasm of students at this level for the use of the concordancer as a means of checking basic aspects of language use such as gender and agreement, which could also have been done by using the dictionary. Furthermore, a minor but interesting advantage was the usefulness of the corpus as a resource for checking capitalisation. The semi-specialised nature of the corpus was particularly important here, as a much more substantial general corpus would have been required to ensure the presence of references to a document such as the European Charter for Regional or Minority Languages. Apart from this obvious example of the benefit of the semi-specialised corpus, however, it is difficult to draw firm conclusions on its role in the project. It is possible that the high number of positive changes was only possible because the corpus, focusing as it did on the very topic on which the students were writing, was able to provide relevant search results despite its very limited size. In addition, the journalistic articles, which mostly expressed the opinion of the journalists on the controversial topic of France's hesitation in signing and subsequent refusal to ratify the Charter, involved exactly the context, namely expressing an opinion about a controversial issue, which was involved in the student writing. The prevalence of the word question in both the searches and the results is an interesting illustration of this, with a search based on question producing 85 results in the corpus. On the other hand, researchers such as Bernardini (2000) have also produced positive results using large corpora such as the British National Corpus. For those working on languages where huge corpora are not available, it is interesting to note Dodd's view that "huge corpora are not necessary for language-teaching purposes. A modest corpus of a million or so words is certainly enough to make a valuable teaching aid and is realistically within the reach of most teaching institutions" (1997:131). For the purposes of more specialised study, 500,000–1,000,000 words may well be an appropriate size, and it is not unrealistic to envisage a scenario where the language resource areas of language

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

departments would make semi-specialised corpora available to students in the areas being studied.

Finally this study raises questions about the nature of the guidance required for learner access to be successful. It is interesting that one student noted that this particular study would not have been successful if the areas requiring improvement had not been indicated on the text. Underlining or otherwise marking inappropriate use of language in student writing is, however, a very common activity in the language learning context, with the result that the situation in which these students found themselves (i.e. faced with a text on which the teacher had indicated phrases which could be improved) is replicated in many cases every day. There is thus very considerable scope for corpora, of whatever size and nature, to be consulted in the context of interactive feedback, whereby learners see the corrected text not as punitive proof of their weaknesses but rather as a positive stage in their language learning, an interim document which they can take to the language resource area to consult the appropriate corpus, or simply access corpora which could be made available via the World Wide Web. Without the availability of corpora, interactive feedback is limited to the correction of errors in areas covered in traditional resources, whereas the corpus can cover these areas as well as helping to reduce native language interference and other aspects of language use where traditional resources are of little or no benefit.

Acknowledgements

Ída O'Sullivan wishes to acknowledge the assistance of the Irish Research Council for the Humanities and Social Sciences which provided a postgraduate scholarship to support her research.

References

Aston, G. (1988) *Learning Comity: an approach to the description and pedagogy of interactional speech*. Bologna: Cooperativa Libraria Universitaria Editrice.

Aston, G. (1997) Small and large corpora in language learning. In: Lewandowska-Tomaszczyk, B. and Melia, J.P. (eds.), *Practical applications in language corpor*a. Lodz: Lodz University Press, 51–62.

Bernardini, S. (2000) Systematising serendipity: proposals for concordancing large corpora with language learners. In: Burnard, L. and McEnery T. (eds.), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the third international conference on teaching and language corpora*. (Lodz studies in language). Frankfurt am Main: Peter Lang, 225–234.

Bowker, L. (1998) Using specialised monolingual native-language corpora as a translation resource: a pilot study. *Meta* **43**(4): XLIII. (Also at: http://www.erudit.org/revue/meta/1998/v43/n4/002134ar.pdf

Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton and Co.

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

Cope, B. and Kalantzis, M. (eds.) (2000) *Multiliteracies: Literacy learning and the design of social future*. London: Routledge.

Dodd, B. (1997) Exploiting a corpus of written German for advanced language learning. In: Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds.), *Teaching and Language*

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press

Corpora. London and New York: Longman, 131–145.
Gavioli, L. and Aston, G. (2001) Enriching reality: language corpora in language pedagogy. *ELT Journal* **55**(3): 238–246.
Johns, T. (1986) Micro-concord: a language learner's research tool. *System* **14**(2): 151–162.
Kasper, L. (2000) New technologies, new literacies: focus discipline research and ESL learning communities. *Language Learning and Technology* **4**(2): 105–128. Available http://llt.msu.edu/vol4num2/kasper/default.html
Kennedy, C. and Miceli, T. (2001) An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology* **5**(3): 77–90. Available http://llt.msu.edu/vol5num3/kennedy/default.html
Kennedy, C. and Miceli, T. (2002) The CWIC project: developing and using a corpus for intermediate Italian students. In: Kettemann, B. and Marko G. (eds.), *Teaching and Learning by Doing Corpus Analysis*. *Proceedings of the fourth international conference on teaching and language corpora*, Graz 19–24 July. Amsterdam: Rodopi.
Leclerc, J. (2000) Histoire de la langue française. Available http://www.tlfq.ulaval.ca/axl/francophonie/histlngfrn.htm
McEnery, T. and Wilson, A. (1997) Teaching and language corpora. *ReCALL* **9**(1): 5–14.
Mishan, F. (forthcoming) Authenticating corpora for language learning: a problem and its resolution. *ELT Journal* **58**(3).
Murray, D. E. (2000) Changing technologies, changing literacy communities? *Language Learning and Technology* **4**(2): 43–58. Available http://llt.msu.edu/vol4num2/murray/default.html
Richards, C. (2000) Hypermedia, Internet communication and the challenge of redefining literacy in the electronic age. *Language Learning and Technology* **4**(2): 59–77. Available http://llt.msu.edu/vol4num2/richards/default.html
Stubbs, M. (2001) Text corpora and problems of interpretation: a response to Widdowson. *Applied Linguistics* **22**(2): 149–172.
Tognini-Bonelli, E. (2001) Corpus linguistics at work. Studies in *Corpus Linguistics, Vol. 6*. Amsterdam: Benjamins.
Turnbull, J. and Burston, J. (1998) Towards independent concordance work for students: Lessons from a case study. *ON-CALL* **12**(2): 10–21. Available http://www.cltr.uq.edu.au/oncall/turnbull122.html
van Lier, L. (1996) *Interaction in the Language Curriculum: Awareness, autonomy and authenticity.* London: Longman.
Warschauer, M. (1999) *Electronic Literacies: Language, culture, and power in online education*. Mahwah, New Jersey: Lawrence Erlbaum.
Widdowson, H. G. (1983) *Learning Purpose and Language Use*. Oxford: Oxford University Press.
Widdowson, H. G. (2000) On the limitations of linguistics applied. *Applied Linguistics* **21**(1): 3–25.
Widdowson, H. G. (2001) Interpretations and correlations: a reply to Stubbs. *Applied Linguistics* **22**(4): 531–538.

https://doi.org/10.1017/S0958344004001211 Published online by Cambridge University Press