HAL
open science

Corpus work with ordinary teachers: Data-driven learning activities

Henry Tyne

To cite this version:

Henry Tyne. Corpus work with ordinary teachers: Data-driven learning activities. Input, Process and Product: Developments in Teaching and Language Corpora, Masaryk University Press, pp.114-129, 2012. hal-00679691

HAL Id: hal-00679691

https://hal.science/hal-00679691v1

Submitted on 16 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Corpus work with ordinary teachers: Data-driven learning activities

Henry Tyne
University of Perpignan

This paper presents findings from experimental uses of corpora for language teaching in secondary school education in France. Working alongside two teachers of Spanish, we identified specific learning outcomes in pedagogical sequences and looked at how corpus use could be brought into their schemes of work with a view to facilitating the attainment of these. In each case, the teachers felt that their task had been made easier, insofar as the enquiry methods or techniques allowed them to carry out their work more efficiently. Such approaches to corpus use in language teaching are found to be compatible with the existing practices of ordinary language teachers. However, they do bring into question the idea of corpus work necessarily involving large, end-state electronic corpora.

1. Introduction

In recent years, it has been pointed out that the use of corpora in language teaching has not become as widespread as perhaps might have been expected (e.g. Römer 2006, 2009). One issue concerning the use of corpora in language teaching is communication (or lack thereof) between research and teaching spheres (McCarthy 2008). And whilst researchers and corpus linguists may be quite keen to underline the potential of their work for teaching and learning, there is a lack of user-friendly applications for general, everyday practice. Indeed, if *we* know that learners can do many things with corpora, this does not necessarily provide much comfort for ordinary teachers who may not be linguists at all.

Most existing studies of corpus use in language learning are carried out in university environments with quite advanced, mature(ish) learners. This is the case for second language acquisition research in general as Macaro (2003: 251) observes. Davies (2004: 260) openly defends his study using advanced students since it shows "a more complete picture of the different ways in which students can use large corpora to study and analyze the grammar of the second language." However, whilst this may indeed be a legitimate argument in terms of the *possibilities* for corpus use, it does mean that we are not dealing with younger learners for whom learning foreign languages may just be part of the many obligations of compulsory schooling.

This paper looks at how two ordinary teachers (referred to here as teacher A and teacher B) of Spanish in different *lycées* (high schools, ages 15-18) in the Lorraine region in north-eastern France go about applying what they know about corpora and corpus data to their teaching. They were left to their own devices, i.e. they were not shown specifically how to apply corpora to teaching and then required to test this. It could be argued that the two teachers mentioned here are, in fact, not very 'ordinary', having both completed

postgraduate studies in applied linguistics. However, they are 'ordinary' insofar as they are teachers in 'ordinary' schools, with 'ordinary' teenage learners and are not researchers working in a university environment (cf. Mauranen 2004a, 2004b). Thanks are due to these teachers (and their students) for their availability and their willingness to discuss work.

As part of an open research programme, the teachers were asked to consider the different learning activities they used and to consider how corpora could be brought into their schemes of work as and when they felt able to do so. An emic approach is adopted here, taking an inside view of how practice is developed. In other words, this paper looks at what the teachers actually end up doing that involves *their* application of corpora rather than having conditions imposed on them based on ideas of what could or should be done: from classroom to corpus and not the other way around. Also, rather than looking at what teachers typically say they do (or like to do), this paper presents what they actually did, step by step, trying to see how and why they chose these activities and how, for them, corpora were important for the teaching / learning. In each case, the teachers simply informed the author that they were working on given topics or skills and that work involving corpora seemed to be a useful way forward. They then went about their work (school year 2009-2010) and reported back in detail, either after the particular learning sequence had run its course or during, to talk things over or ask for advice or information. Outcomes were then discussed in informal interviews.

In the following sections, three case studies that exemplify the use of what I will loosely term 'corpus work' are presented. This does not mean that the use of corpora by these teachers is limited to these examples. Rather, it means that these three cases were discussed extensively and can therefore be reported in full here. It should be stressed that, whilst the activities detailed in these sections may challenge mainstream views of what corpora are (or allegedly should be) and how they can be used in language pedagogy, the aim here is first and foremost to present the teachers' uses – how *they* saw corpora and how *they* thought work relating to corpora could help *them*; how *they* integrated 'corpus work' into their teaching. These uses are commented on in the discussion session.

2. Case studies

2.1. From text to corpus with whiteboard concordancing techniques

One of the curriculum requirements for the French *baccalauréat* examination taken at age 18 is to be able to produce a piece of writing on a given topic respecting a genre or type of production such as a short news item or letter to a friend. In this case study, two classes of learners (35 in all) were asked to work on short news items of general interest (*faits divers* or *sucesos* in Spanish).

Teacher A's first step was to look at how well students were acquainted with this genre in their mother tongue (L1) French. She began by simply asking students, without any priming, to write a short news item in French. This initial exercise revealed that students had difficulty in producing this genre in their L1: in both groups, students were quite clearly challenged – 8

out of 35 were incapable of producing anything at all and, on the whole, the productions were reported to be quite poor.

The teacher then asked students to go away and find a short news item in a French newspaper or from an online news source. This they did and, in the following lesson, the teacher asked them what they thought typified a short news item. She reported that students came up with a range of different replies but generally they had grasped the importance of it relating to a given event or incident involving given participants in a given place and usually with given consequences. As students had picked up on the importance of the title in the *faits divers*, the teacher subsequently asked them to study these: students picked out items of information but did not comment on form. In a fashion similar to what Johns (1993: 7) has termed "blackboard concordancing" (see also "hand-concordancing" in Willis 1998), in which groups of learners working on different pages of text produce concordances directly on the board for the whole group to see, the teacher then asked students to call out the titles of their articles: these were written up on the classroom whiteboard one below the other. As the titles appeared, those which contained a verb stood out as exceptions and the recurrent verblessness (which had gone unnoticed until then) of a large proportion of titles was commented on explicitly by the class. As Willis (1998: 50) puts it, "somehow, concordance lines have the effect of enabling us to take a more objective look at the language... you often notice something you hadn't realised was there". This is the case here too, dealing with ad hoc 'concordances' that draw attention to the given feature (in this case absent verbs).

The students next received a collection of seven other texts in French selected independently by the teacher. They were asked to consider the recurrent features in the collection; these were found to relate to the following questions: "who / what / when / where / how?" This stage was deemed to be important by teacher A since it required students to deal with unknown texts as parts of a larger collection and not to attempt to read or 'study' each text in detail. It was also useful since it allowed students to get an idea of the different themes in the short news items and the types of words (and their frequency) they would be likely to encounter when they did the same exercise in Spanish.

Next a collection of seven short newspaper articles were given to students, this time in Spanish. The same wh-questions were asked and students simply had to pick out this information from the target language texts. Again, this is worthy of note since it is an atypical way of approaching written sources in the target language. A pyramid or 'snowballing' technique was then used to get students engaging with the collection of texts: first they were asked to explore each text individually; then in pairs with neighbours; then as a full class exercise. The wh-questions table was completed on the board.

The teacher went on to look at grammar, asking students which tenses they would typically find in the Spanish corpus – they all said past tense with a preference for the *pretérito perfecto* (cf. the *passé composé* in French). The students were then asked to pick out all verbs in the past tense from the articles in Spanish and these were again written up on the board: they were surprised to discover than the *pretérito indefinido* or simple past was common in the Spanish news items. They had originally questioned the pertinence of learning what they thought was a 'rare' tense (cf. the *passé simple* in French) – and now they

saw why. This discovery linked directly into a parallel grammar session on forming the *pretérito indefinido*.

The teacher then returned to the initial preoccupation of producing a written piece. The students had seen the various elements they would need in the set of exercises and accordingly went about producing their own news items on a particular theme, this time in Spanish. All of them went on to produce a *suceso* (see example in Appendix A), using the appropriate tenses and lexical items relating to the given theme, and found it easy and enjoyable. The practical issue of being able to do hands-on work (writing on the texts, highlighting, sorting and comparing them, etc.) without electronic equipment, was seen to be a key feature by teacher A as she sought to link what she saw as useful principles in corpus work to existing practices and more conventional exercises in a language classroom containing one group of learners, one teacher and one whiteboard.

2.2. Knowledge of spoken corpora and exploitation of meaningful examples

As a part of ongoing classwork on writing, the same Spanish teacher as in the previous case study looked at the notion of 'producing a dialogue' which is a requirement for the *baccalauréat* examination. One of the problems identified by the teacher, given her knowledge of spoken corpora acquired in postgraduate studies,[1] was that there are certain features in authentic spoken dialogues (i.e. not staged ones) which, when adapted to the written mode, can serve to categorise speakers (cf. MacCaulay 1991). The teacher's aim then, was to first acquaint learners with spontaneous oral dialogues before going on to work on written pieces in which orality can be more or less present and in which various writing and punctuation devices can be used to great effect to highlight or even exaggerate certain aspects of speech production.

Students were given a short extract of a transcription taken from a corpus of spoken French[2] (see Appendix B) and were subsequently asked to identify features that they thought would not be in a 'typical' written text: predictably, they identified features such as pausing, overlapping speech, unfinished sentences. This extract was in turn compared to three other dialogue extracts from non-spoken sources: in a novel, in a play and in a cartoon strip. The students then went about discussing those elements of 'staged' orality (e.g. a description of the scene in a play, interrupted speech with punctuation marks, font size, etc.) as well as those elements that were recurrent in the transcript but were not to be found in the other dialogues (e.g. overlapping speech, pausing, hesitation and 'fishing' for words). The same exercise was then repeated using similar data in Spanish, and the students went through the

[1] It is worth noting that spoken corpora have occupied a prominent place in corpus linguistics in France since the pioneering work carried out by Claire Blanche-Benveniste and the Aix-based GARS research team (*Groupe Aixois de Recherches en Syntaxe*) in the 1970s and 1980s (see Blanche-Benveniste & Jeanjean 1987). Moreover, issues pertaining to sociolinguistics (including methodological aspects of fieldwork and transcription) have typically received attention in the literature on corpora (Cappeau & Gadet 2007, 2010). Many French students of linguistics will have memories of long hours spent transcribing their own data for coursework assignments.
[2] Extract taken from the Surrey corpus (see Tyne 2009b). Reference to Spain was seen as an additional point of interest here.

extracts identifying the same types of procedures. To complete the sequence, they went on to produce some very convincing dialogues in the target language.

Teacher A commented on the usefulness of corpus data as a resource or source of teaching material to provide authentic extracts for a qualitative approach. This does not exclude working with these data in any other way, but it is interesting to note how this teacher's experience of corpus-building (and in particular transcription and transcription theory) provided a path into the resulting corpus work. This also applies in the following case study.

2.3. Building a corpus of spoken Spanish

In the third case study, teacher B chose to build on existing listening comprehension exercises by introducing transcriptions of material recorded by the students themselves in the form of interviews with their Spanish language assistant. Teacher B's first remark was that transcription could be a useful means for her, as a teacher, to make sure that the recorded extracts had been listened to thoroughly by all students: prior experience had shown that she had little or no control over *how* students would listen to given extracts used in out-of-class exercises, which in turn gave rise to discrepancies in student involvement during follow-up classroom activities. The teacher also had the impression that students enjoyed doing transcription and wanted to find out in more detail which aspects of this exercise they were most attracted to, what they were getting out of it, and why. She was also interested to see what they thought of the possibility of generalising this type of exercise. The seven questions in Table 1 (based on Tyne 2009a) were put to students in the form of a questionnaire, with space to expand on the YES/NO answers in each case.

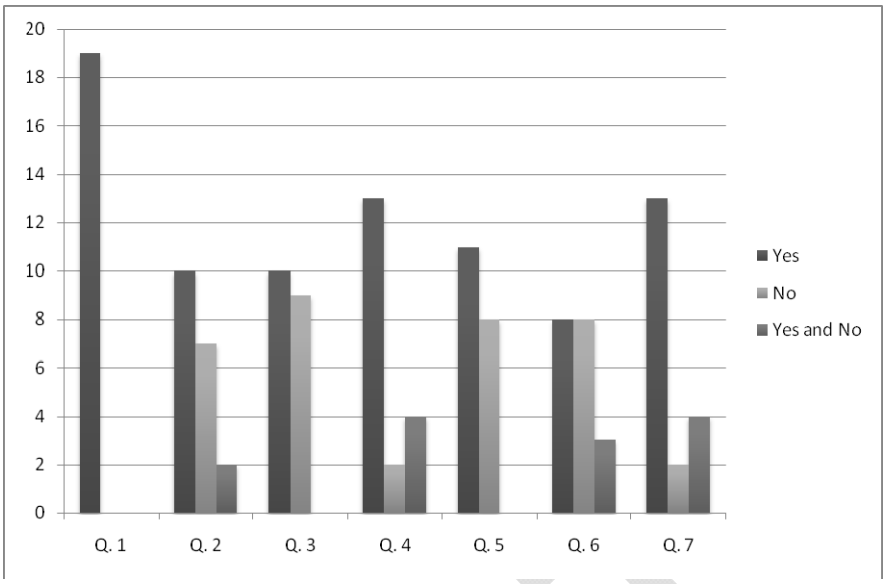
Table 1. Questions put to students (translated from French)

1. Do you think transcription is useful for improving your Spanish?
2. Did you get the impression that you were discovering new things about Spanish through transcription?
3. Did you get the impression that you were gaining confirmation of things you already knew about Spanish through transcription?
4. Did you find transcription hard?
5. Did you enjoy doing transcription?
6. Do you think transcription could be used with less advanced learners?
7. Other than for listening comprehension exercises, do you think transcription (either the **process**, i.e. the act of transcribing; or the **product**, i.e. use of the resulting transcript) could be used for other learning activities (for Spanish or other languages)?

Nineteen usable questionnaires were returned; the responses are presented in Figure 1, and will be discussed in more detail below.

Figure 1. Answers to questions put to students

DRAFT



All 19 students in this study felt that transcription was useful to improve their Spanish in one way or another (see column 1 in Figure 1), commenting in particular on the importance of having to attend to detail, on the new ways of working, and on the contact with authentic speech. Their expanded comments include:

1. *Je pense que la transcription d'un discours espagnol permet d'améliorer grandement la compréhension orale et d'enrichir le vocabulaire car ce genre d'exercice nécessite un travail de recherche qui, indéniablement, apprend de nouveaux mots et expressions aux élèves (s'il est fait sérieusement).*
   [I think transcription of speech in Spanish allows us to greatly improve listening comprehension and to enrich vocabulary because this type of exercise requires us to search and this undeniably allows new words and expressions to be learnt by students (if the work is done seriously).][3]
2. *Quand on transcrit, il se passe dans notre tête ce qui se passe lorsque l'on est en face de quelqu'un qui nous parle espagnol. En plus, cela nous permet d'entendre des gens parler espagnol. Je trouve cela intéressant de corriger les transcriptions des autres, cela nous permet de voir ce que peuvent entendre les autres.*
   [When we transcribe, the same things happen in our heads as when we are with a person who is talking in Spanish to us. What's more, it allows us to hear people speaking Spanish. I find it interesting to correct other people's transcripts as it allows us to see what other people think they hear.]
3. *La transcription nous aide à mieux comprendre une situation et à se forcer en quelque sorte à trouver un mot, le mot manquant.*
   [Transcription helps us to understand the situation better and, in a sense, to force ourselves to find the missing word.]

Concerning questions 2 and 3 about the discovery of new things and the verification of existing knowledge, the answers appear to go both ways (see columns 2 and 3 in Figure 1 above): some students discovered things, others didn't; some felt reassured by seeing things they already knew, others didn't; and so on. We also find the 'discovery' of things already known (but forgotten), the discovery of the ability to identify and transcribe new forms, and the discovery of interesting content:

4. *J'ai eu la surprise de reconnaître certaines expressions espagnoles que j'ai apprises il y a longtemps.*
   [I was surprised to recognise certain Spanish expressions that I had learnt a long time ago.]

[3] English translations are fairly literal and are supplied as an aid for comprehension. All answers were given by students in French.

DRAFT

5. *J'ai découvert que je comprenais assez bien quand la personne parle au calme et que je pouvais orthographier correctement certains mots inconnus.*
   [I discovered that I understood quite well when the person speaks slowly and clearly and that I could spell certain unknown words correctly.]
6. *Oui [j'ai appris des choses] sur les coutumes des Espagnols par rapport aux coutumes françaises (sur Noël).*
   [Yes I discovered things about Spanish traditions in comparison with French traditions (for Christmas).]

The issue of whether or not it is good practice to let learners transcribe extracts that contain unknown elements is not dealt with here. However, it should be noted that these students are all in a typical guided learning context, in which the task of segmenting language and assigning meaning to given forms and structures is generally highly developed (Klein 1986). As one student remarked, "*J'arrive à transcrire presque bien mais pas toujours tout comprendre*" [I can manage to transcribe pretty well but I don't always understand all of it]. It is interesting to note that (question 4) the majority of students found transcription hard (see column 4 in Figure 1 above); rate of speech was typically commented on here.

The answers to question 5 are perhaps a little surprising: whilst all students say that they thought the exercise was beneficial for their Spanish (see column 1 in Figure 1 above), 8 out of 19 students said they did not like the exercise, typically claiming it was too hard, too repetitive[4] or too individualised:

7. *C'est trop long, il faut toujours recommencer.*
   [It's too long, you always have to start again.]
8. *C'était un travail trop personnel, la classe ne travaille pas en commun.*
   [It was too individualised, the class doesn't work together.]

However, those students who did like transcription picked up on aspects such as attention to detail, or the fact that many new things are encountered and that it involves authentic language. Other issues such as the possibility for self-evaluation and also the nature of the task itself were also viewed positively:

9. *Cela permet de voir notre niveau, de nous auto-évaluer.*
   [It allows you to see how good you are, to get an idea of how you're doing.]
10. *J'ai trouvé cet exercice à la fois ludique et enrichissant.*
    [I found this exercise both fun and enriching.]

Question 6 deals with transcription as an exercise with lower-level learners, an important issue for teacher B dealing with different levels of proficiency across different groups. The reactions appear to be quite divided, with some students strongly in favour and some clearly against (in particular due to the difficulty), while others were quite undecided. It is interesting to note that for some students their response related to the nature of the data to be transcribed, as is the case in the following:

11. *Si l'exercice est adapté à leur niveau et est sur un thème qu'ils travaillent, oui.*
    [If the exercise is adapted to their (= the students') level and is on a theme they're working on, yes.]

[4] Teacher B did not make use of transcription software with students. It may be that the repetitive aspect commented on negatively here was reinforced by the inconvenience of having to continually go back and relisten to small portions of the sound file whilst transcribing in a separate text file or on paper.

DRAFT

In the case of question 7, the vast majority of students felt that transcription could be used beyond listening comprehension tasks, suggesting its merits, for example, in improving writing skills, in oral preparation (noticing how others speak in given registers), or simply for exploring the other languages they are learning:

12. *Oui, dans toutes les langues étudiées. Cela aide et la plupart des élèves aiment faire cet exercice.*
    [Yes, in all the languages we study. It helps and most students like doing this activity.]
13. *Cela peut permettre de chercher de nouveaux mots, de savoir les écrire et de chercher leur sens.*
    [It can help you to find new words, to see how to write them and to find their meaning.]

So, not only is transcription found to be a simple means of checking that all students have actually listened in detail to various extracts assigned out of class, it also serves to promote attention to detail (cf. Lynch 2001) and to draw students closer to the target language in general. Furthermore, having each student transcribe several minutes of speech is a step towards the creation of a collaborative local corpus to be used with future generations of learners (cf. teaching to establish resources; Renouf 1997) – essential for languages where few resources are available, and important also for targeting specific objectives (Aston 2002; Koester 2010). However, the questionnaire results provide food for thought and allow the teacher to better evaluate the potential (and limitations) of doing certain 'new' things with students. Again, it is interesting to note that the teacher was attracted by the more practical aspects of corpus work (making the corpus) as a means of connecting with staple language learning exercises (listening comprehension) rather than setting out from more "conventional" corpus consultation methods and activities (Braun 2010: 92).

3. Discussion

In case studies one and two, we see how work on individual texts is used. As Flowerdew (2009) stresses, consciousness-raising actives at the level of text can be a useful starting point in the classroom rather than a bottom-up approach where, typically, "truncated concordance lines are examined in a somewhat atomistic fashion" (Flowerdew 2009: 395). Furthermore, as Charles (2007) and Weber (2001) have found, a top-down, text-to-corpus approach can be extremely useful in getting learners into corpora. And as Braun (2010: 78) points out, for many learners exploring a corpus, "a good grasp of the texts" qua texts can be crucial for understanding certain nuances. This is what teacher A seems to have been doing in case study one, for example: first the students read short texts, next they were asked to skim texts for information, then they picked out forms and made sense of them via a spontaneous whiteboard concordance (i.e. no longer from within the source text context). Each text remains accessible as a text, but at the same time the set of texts is used as a collection – it is a "manageable corpus" (Willis 1998: 46). The teacher clearly recognises that students' typical ways of engaging with the target language will be through familiar activities and tasks around selected topics and documents, and it is perhaps too much to expect these students to engage directly with concordance lines since they will not entirely grasp their significance (i.e. the fact that they represent moments extracted from many texts, and that the repetition of forms is significant for understanding the particular distribution of linguistic features across different registers, etc.). In terms of the students' written productions, as Gavioli (2001: 125) admits, "the idea of asking learners to produce texts by combining

DRAFT

chunks copied from other texts may seem questionable"; nonetheless, such an exercise enabled students to incorporate the various elements they had noticed into their own writing.

In case study two, it is clearly the knowledge of what spoken data is typically like that motivated teacher A's approach. In this case a corpus extract is presented as though it were a text that students are invited to read – noticing arises precisely because it is *not* a written text and so reading is not an easy matter. What is interesting here is that the teacher was using *her* knowledge of spoken corpora (having worked as a student with spoken data in different ways) in order to inform her teaching practice. She used a printout of a transcribed passage as her model. This is akin to the data-driven language awareness highlighted by Hales (1997), in which trainee teachers acquaint themselves with transcriptions.

In the third case study, there is a strong emphasis on the dynamic nature of 'pre-corpus' work. In this case, then, not only is transcription found to be a simple means of checking that all students have actually listened in detail to various extracts, but teacher B also realises that creating a small spoken corpus as a collaborative task can be a learning activity itself (cf. Tyne 2009a). The group of students involved in this activity may never actually access the end-state corpus – but they are still learning through their work with language data, which is the crucial aspect of data-driven learning (Johns 1991). However, they will have carried out a series of activities concerning the practical and methodological aspects of corpus-building. It is clearly possible to learn from corpus data without ever going near a concordancer. Also, teacher B observes that the transcription exercise can be worked on so that it is more compatible with existing student learning activities appropriate for different types of learner and different types of class – 'academic' or vocational. For example, several students mentioned in the questionnaires that they found they worked on their own too much, without conventional classroom interaction with peer and teacher support. Teacher B looked upon this feedback (see example 14) favourably since it showed, amongst other things, engagement: her students were not indifferent to the activity, but were keen to discover things for themselves:

14. *En espagnol on devrait le faire plus souvent mais en classe, comme ça on ne resterait pas sans savoir. Vous pourriez nous aider un peu…*
    [In Spanish we should do this more often, but in the class so as to make sure we don't come away without knowing. A bit of help wouldn't go amiss…]

In the three case studies detailed here there is virtually no mention of corpora in the modern understanding as implied typically in corpus linguistics today, i.e. a large collection of texts stored and available for consultation electronically (Tognini-Bonelli 2010). The two teachers, given their particular expertise, their working conditions and their teaching / learning objectives, clearly appear to be attracted to a small, local corpus approach and, in particular, appear to target the dynamic processes involved rather than end-state corpus consultation. Teacher A did admit to having used Mark Davies' Corpus del Español[5] for lesson preparation, though with limited results (and with a limited degree of patience). This was apparently the case for her attempt to prepare an exercise based on the observation of the imperative following verbs such as *pedir* and *decir* (e.g. *te pido que vengas* [*I'm asking you to come*]).

[5] Available at http://corpus.byu.edu/, accessed 20/06/11.

Similarly, teacher B commented on her attempts to work with the online SACODEYL[6] data, in which youth language samples can be accessed. Here, the authenticity of corpus data (cf. Chambers 2009) was seen to be attractive, although the accents of the young Spanish speakers were found to be problematic for the French learners. However, by getting students actually involved in the processes, these two teachers were bringing them to experience the data directly – surely the heart of 'data-driven learning'. As various scholars have already observed, the 'homemade' approach to corpora can be beneficial for engaging learners with the data: Aston (2002) gives a list of advantages including a sense of ownership, greater focus (exact themes that correspond to the topics studied), increased critical awareness, etc. Elsewhere, Braun (2005) has pointed out that it helps learners to better authenticate the corpus, i.e. to engage with the human dimension involved. This type of approach is rooted in task-based language pedagogy whereby "authenticity of methodology" (Mishan 2004: 225) can be seen as a way of involving learners with the data in meaningful ways.

Both of the teachers felt that their tasks had been made easier and that the corpus methods and techniques allowed them to carry out their work more efficiently. However, it is difficult with this type of research to positively state that students learned more (or even learned more efficiently) than they would have done with more conventional activities. It would be rather demanding to expect ordinary teachers to systematically carry out control and test group observations. In fact, teachers already have at their disposal a range of measures and indicators that inform them of students' achievement and they must use these foremost when they come to take stock of their corpus-based activities, and when they come to decide whether they will continue to use them. For example, at a very basic level, the students' marks are taken as measures of progress; or quite simply, questions such as 'are they interested in the work?', 'did they complete the exercise?' or 'can they do things that they couldn't do at the start?' can be indicators of learner engagement or achievement. But teachers also need to see students working well and enjoying it. As teacher A commented, this too is a simple but effective and perfectly normal means of seeing how well a given activity is working.

4. Conclusion

Mauranen (2004a: 100) stresses just how difficult it can be to try to bring "corpus thinking" into teaching: her experience even with highly skilled teachers shows that it is "surprisingly hard to convey the logic of what corpus data is like, how a corpus works, what kinds of questions can be asked, as well as the more mundane logic of the program" (Mauranen 2004b: 209). The difficulty referred to here is essentially that of accessing language as a paradigmatic display of linguistic structures. This is quite understandable, and for teachers to fully understand and adopt the methods of corpus linguistics is undoubtedly a challenge. Indeed, few ordinary language teachers are also corpus linguists (if they are linguists at all). Thus, for some observers, corpus input into teaching should be all but invisible, with existing materials (i.e. coursebooks) benefitting from corpus enquiry (e.g. McCarthy 2004: 15). For

[6] System Aided Compilation and Open Distribution of European Youth Language (http://www.um.es/sacodeyl/, accessed 16/06/11) supported by the European Commission. See Widmann (2007).

others, corpus enquiry can be introduced directly (e.g. Chambers 2005). But here we come up against the problems of introducing techniques and attitudes which may work with older, more advanced learners, taught by the researchers themselves, but which may, for one reason or another, be quite difficult to reconcile with many ordinary school learning environments (Duda & Tyne 2010: 101).

However, as this paper has attempted to demonstrate, there are cases in ordinary, everyday teaching where activities can be fitted to certain aspects of corpus work – and not necessarily end-state corpora, ready for consultation: recording, listening, transcribing, assembling collections of texts, working with individual texts as well as the whole corpus, etc. Corpus linguists use all kinds of techniques which may be relevant for our learners, but there is no reason why teachers should be restricted to these alone when using corpora. This is the point made by Braun (2010) in her paper, Getting past 'Groundhog Day'.

In the case studies presented here, teachers A and B simply went about applying those aspects of corpus work they felt were reconcilable with their existing day-to-day activities in given sequences. It has been shown that these teachers have quite a pragmatic view of the uses of corpora for language learning: what can *we* get out of it given the many constraints that exist, including the availability of suitable technology (computers or just a classroom whiteboard?), student motivation (e.g. how far can we go getting students to transcribe, and does this work with all students?) and predominant teaching-learning methods within the system? The activities presented here certainly go beyond conventional definitions of corpora applied to language teaching (e.g. Reppen 2010). However, they are unquestionably examples of 'data-driven learning' using corpora, in the broadest sense, and invite us to think more openly about the applied uses of corpora in language teaching in ordinary learning environments (cf. Braun 2009).

References

Aston, G. 2002. The learner as corpus designer. In B. Kettemann & G. Marko (eds), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, p. 9-25.

Blanche-Benveniste, C. & C. Jeanjean. 1987. *Le Français Parlé: Transcription et édition*. Paris: Didier Érudition.

Braun, S. 2005. From pedagogically relevant corpora to authentic language learning contents. *ReCALL* 17(1): 47-64.

Braun, S. 2009. *Corpus-Enhanced Language Learning and Teaching* [Backbone deliverable 3.1 part 1, project report]. http://u-002-segsv001.uni-tuebingen.de/backbone/moodle/file.php/1/reports/02%20Corpus-enhanced%20LLT%20-%20(del%203.1%20part%201%20of%202).pdf, accessed 26/11/10.

Braun 2010. Getting past 'Groundhog Day': Spoken multimedia corpora for student-centred corpus exploration. In T. Harris & M. Moreno Jaén (eds), *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, p. 75-97.

Cappeau, P. & F. Gadet. 2007. L'exploitation sociolinguistique des grands corpus: Maître-mot et pierre philosophale. *Revue Française de Linguistique Appliquée* 121: 99-110.

Cappeau, P. & F. Gadet. 2010. Transcrire, ponctuer, découper l'oral: Bien plus que de simples choix techniques. *Cahiers de Linguistique* 35(1): 187-202.

Chambers, A. 2005. Integrating corpus consultation in language studies. *Language Learning & Technology* 9(2): 111-125.
Chambers, A. 2009. Les corpus oraux en français langue étrangère: Authenticité et pédagogie. *Mélanges CRAPEL* 31: 15-33. http: //revues.univ-nancy2.fr/melangesCrapel/, accessed 16/06/11.
Charles, M. 2007. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes* 6(4): 289-302.
Davies, M. 2004. Student use of large, annotated corpora to analyze syntactic variation. In G. Aston, S. Bernardini & D. Stewart (eds), *Corpora and Language Learners*. Amsterdam: John Benjamins, p. 259-269.
Duda, R. & H. Tyne. 2010. Authenticity and autonomy in language learning. *Bulletin Suisse de Linguistique Appliquée* 92: 87-106.
Flowerdew, L. 2009. Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics* 14(3): 393-417.
Gavioli, L. 2001. The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (ed.), *Learning with Corpora*. Houston: Athelstan, p. 108-137.
Hales, T. 1997. Exploring data-driven language awareness. *ELT Journal* 51(3): 217-223.
Johns, T. 1991. Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (eds), *Classroom Concordancing. English Language Research Journal* 4: 1-16.
Johns, T. 1993. Data-driven learning: An update. *TELL&CALL* 2: 4-10.
Klein, W. 1986. *Second Language Acquisition*. Cambridge: Cambridge University Press.
Koester, A. 2010. Building small specialised corpora. In A. O'Keeffe & M. McCarthy (eds), *Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, p. 66-79.
Lynch, T. 2001. Seeing what they meant: Transcribing as a route to noticing. *ELT Journal* 55(2): 124-132.
Macaro, E. 2003. *Teaching and Learning a Second Language: A guide to recent research and its applications*. London: Continuum.
MacCaulay, R. 1991. 'Coz it izny spelt when they say it': Displaying dialect in writing. *American Speech* 66(3): 280-291.
Mauranen, A. 2004a. Spoken corpus for an ordinary learner. In J. Sinclair (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, p. 89-105.
Mauranen, A. 2004b. Speech corpora in the classroom. In G. Aston, S. Bernardini & D. Stewart (eds), *Corpora and Language Learners*. Amsterdam: John Benjamins, p. 195-211.
McCarthy, M. 2004. *Touchstone: From corpus to coursebook*. Cambridge: Cambridge University Press. http://www.cambridge.org/us/esl/Touchstone/teacher/images/pdf/CorpusBookletfinal.pdf, accessed 16/06/11 via Internet Archive.
McCarthy, M. 2008. Accessing and interpreting corpus information in the teacher education context. *Language Teaching* 41(4): 563-574.
Mishan, F. 2004. Authentic corpora for language learning: A problem and its resolution. *ELT Journal* 58(3): 219-227.
Renouf, A. 1997. Teaching corpus linguistics to teachers of English. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds), *Teaching and Language Corpora*. London: Longman, p. 255-266.
Reppen, R. 2010. *Using Corpora in the Language Classroom*. New York: Cambridge University Press.
DRAFT

Römer, U. 2006. Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik* 54(2): 121-134.

Römer, U. 2009. Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, p. 83-98.

Tognini-Bonelli, E. 2010. Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (eds), *Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, p. 14-27.

Tyne, H. 2009a. Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL* 31 : 91-132. http: //revues.univ-nancy2.fr/melangesCrapel/, accessed 24/11/10.

Tyne, H. 2009b. Style in L2: The icing on the cake? In E. Labeau & F. Myles (eds), *The Advanced Learner Variety: The case of French*. Bern: Peter Lang, p. 243-268.

Weber, J-J. 2001. A concordance- and genre-informed approach to ESP essay writing. *ELT Journal* 55(1): 14-20.

Widmann, J. 2007. The SACODEYL project: Corpus exploitation for language learning purposes. In S. Granger (ed.), *Optimizing the Role of Language in Technology-Enhanced Learning*. Louvain-la-Neuve: IDILL workshop proceedings, p. 69-71. http://telearn.archives-ouvertes.fr/docs/00/19/72/03/PDF/GRANGER-SYLVIANE-2007.pdf, accessed 16/06/11.

Willis, J. 1998. Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In B. Tomlinson (ed.), *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, p. 44-66.

DRAFT

Appendix A. Task guidelines and sample of final product

Apellido y Nombre : Fecha :

TAREA DE EXPRESIÓN PERSONAL - ESCRIBIR UN SUCESO

1- La trama: La caserne des pompiers détruite par le feu = Des recherches sur la cause.

Détruire = destruir – La caserne = el cuartel

détruite = destruida – la cause = la causa

2- Escribe en el cuadro el título de tu artículo.

3- Escribe la lista de los elementos de circunstancias que vas a emplear en tu artículo

¿Quiénes?	¿Cuándo?	¿Dónde?

4- Redacta el artículo (70 palabras), valiéndote (*en te servant*) del vocabulario útil así como de los tiempos del pasado y de las circunstancias que habrás escogido (*que tu auras choisies*).

Cuartel destruido por las llamas

Ayer, durante la noche, un incendio se declaró en el cuartel de los bomberos. Eran las dos de la madrugada, así que, no había nadie en el establecimiento: el equipo de noche trabajaba al exterior. Cuando los bomberos regresaron al cuartel, las llamas tenían destruido una buena parte del establecimiento. A las tres de la madrugada, el cuartel esta completamente destruido, los bomberos no lograron dominar las llamas, era muy tarde. Actualmente, no se sabe exactamente cómo las llamas se declararon y hais que esperar la conclusión del peritaje. Afortunada-mente, no hubo herido grave.

Appendix B. Extract from a corpus used by teacher A

+	pause (+ short , ++ medium, +++ long)
<…>	overlapping speech
x_	lengthened syllable
/…/	different possibilities for transcription (e.g. when the transcriber is not certain)
xxx-	truncated word

A: tu tu connais un peu l'Espagne déjà
B: pardon
A: tu tu connais déjà l'Espagne
B: euh
A: tu es <déjà> <B: non> allée là-bas
B: non
A: non mais tu je pense que tu vas aimer Barcelone parce que ++ à l'u- à l'universit- euh pff oui + hm donc c'est vous_ + vous allez à l'université vous_ vous allez dans les cours euh comme +
B: euh /0, ou, où/
A: non + comme les autres étudiants espa<gnols ou> <B: euh> comment ça marche

DRAFT