Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation

Pascual Pérez-Paredes

Department of English, Northern Arizona University, Flagstaff, US
Departamento de Filología Inglesa, Universidad de Murcia, Spain

María Sánchez-Tornel

Departamento de Filología Inglesa, Universidad de Murcia, Murcia, Spain

Jose María Alcaraz Calero

Communications and Information Engineering Department, Universidad de Murcia, Murcia, Spain

Pilar Aguado Jiménez

Departamento de Filología Inglesa, Universidad de Murcia, Murcia, Spain

Abstract

Much of the research into the language learners' use of corpus resources has been conducted by means of indirect observation methodologies, like questionnaires or self-reports. While this type of study provides an excellent opportunity to reflect on the benefits and limitations of using corpora to teach and learn language, the use of indirect observation methodologies may confine the scope of research to the learners' opinions about the benefits of using corpora for language learning and their self-perceived difficulties in consulting them.

This article proposes and discusses the use of logs to research learners' actual use of corpus based resources, analyzing the number of events or actions performed by each individual, the total number of different web services used, the number of activities completed, the number of searches performed on the BNC and, last, the number of words or wildcards per BNC search. Our research used these parameters to investigate whether learner interaction with corpus-based resources differed under different corpus consultation conditions: guided vs. non-guided consultation.

Our findings show that the individuals in the two research conditions behaved differently in two of the parameters analysed: the number of different web services used during the completion of the tasks and the number of BNC searches. Our results corroborate empirically the suggestions found in the literature that skills and guidance are necessary when teachers take a corpus to the classroom. Similarly, we offer evidence that user tracking is essential to claim research and results validity.

Keywords: corpus linguistics; corpus-based resources; research methodology; tracking; activity monitoring;

1. Introduction

The use of corpora in the language classroom can no longer be considered a novel methodological phenomenon, at least in tertiary levels. In particular, the use of corpora for the teaching of specialized languages (Flowerdew, 2004; Gavioli, 2005; Thurstun and Candlin, 1998) and translation (Beeby, Rodríguez Inés & Sánchez-Gijón, 2009;

Bernardini, 2004; Frankenberg-Garcia, 2004; Santos and Frankenberg-Garcia, 2007) has become a widely accepted practice which has attracted the attention of researchers for almost two decades now. Nevertheless, much of the research dealing with the uses of corpora in the language classroom has not addressed the users' actual interaction with the resources themselves by means of computer-tracking, and certainly this has not been done in connection with different corpus consultation environments. Hafner and Candlin (2007:304) have stressed the lack of studies that provide "direct evidence of students' self-directed use of corpus tools", which may be attributed to the preference of indirect observation methodologies over direct data extraction.

In this research, we set out to gain insight into advanced English as a Foreign Language (EFL) learners' actual interaction with corpus-based resources while completing different focus-on-form activities under two research-conditions: guided and non-guided corpus consultation. The design of the tasks was exactly the same, except for the fact that participants in the experimental group had access to explicit information and guidelines on the use of the British National Corpus (BNC) web interface.

In the second part of this paper we will review previous research in the use of computer logs to investigate students' interaction with computers in the field of CALL, Data Driven Leaning (DDL) and the role of guidance in corpus use. Sections 3 and 4 are devoted to the research question and research methodology used in our research. Section 5 offers the result of our experiment, while section 6 discusses the implications of our findings in the context of language education and applied linguistics. Finally, in section 7 of this paper we offer conclusions and future research directions.

2. Literature review

2.1. Using computer-generated logs to research student-computer interaction

While investigations on how corpus-based resources can be integrated in language pedagogy are widely reported in several volumes (Aston, Bernardini & Stewart, 2004; Bernardini, 2000a; Braun, Kohn & Mukherjee, 2006; Hidalgo, Quereda & Santana, 2007; Sinclair, 2004) as well as in numerous papers (Boulton, 2009a; Chambers, 2005; Gilmore, 2009; Kennedy & Miceli, 2001, 2002a, 2002b, 2010; Lee & Swales, 2006; Sun, 2003; Sun & Wang, 2003; Vannestal & Lindquist, 2007; Varley, 2009), most of these studies have traditionally focused on learners' opinions and learning outcomes, failing to give an account of students' interaction with the resources.

The concern for keeping and examining a detailed record of students' actions while completing computer-based tasks is gaining importance in the research agenda within

and beyond general CALL and data-driven learning (DDL). Abdel Latif (2008) offers a thorough account of software employed by numerous researchers to obtain computer-generated protocols, and also describes how these protocols have been used for different analytical purposes in researching computer based writing: temporal aspects of the writing process, analysing revision, stimulation of writers' retrospection. Similar tools have been used for vocabulary (Desmarais, Duquette, Renié & Laurier[1], 1998) and grammar (Chapelle & Mizuno, 1989) in CALL. This approach to data collection has been used in corpus consultation studies by Cobb (1997), Chan and Liou (2005) and Gaskell and Cobb (2004).

The preference for direct data extraction techniques over other options such as questionnaires, verbal protocols or interviews is better understood given the unobtrusive nature of the software employed to track students' behaviour. Computer-based tracking techniques can contribute to overcoming one of the biggest problems posed by indirect tools: the fact that what students report to be doing or what we assume they are doing when we observe them might be quite distant from what they are actually doing (Chapelle & Mizuno, 1989; Fischer, 2007). Despite these enormous advantages, tracking techniques are sometimes criticized on the grounds that it is almost impossible to access learners' cognitive processes (Goodfellow & Laurillard, 1994; Abdel Latif, 2008). To minimise this limitation computer logs are often used in combination with introspective or retrospective techniques which can unveil cognitive processing. This can be achieved by recording both types of data independently using different tools (Goodfellow & Laurillard, 1994; Desmarais et al., 1998) or by using software which supports keystroke logging as well as speech recognition (Leijten & Van Waes, 2005, in Abdel Latif, 2008).

2.2 Using computer-generated logs to research corpus consultation

Despite the amount of existing research on DDL, and although many researchers have pointed up the importance of recording active exploitation of corpus tools (Johns, 1997; Horst, Cobb & Nicolae, 2005; Chambers, 2007; Hafner & Candlin, 2007), the studies reflecting the resources consulted by learners, exact queries, search results or even discoveries are still very few in number. Depending on the data extraction method, we find two major groups of studies: those in which students record their activity manually (Ma, 1994; Chambers & O'Sullivan, 2004; Frankenberg-García, 2005; O'Sullivan & Chambers, 2006; Varley, 2009) and those in which a computer program generates log files automatically (Cobb, 1997; Johns, 1997; Gaskell & Cobb, 2004; Chan & Liou, 2005; Hafner & Candlin, 2007; Yoon, 2008).

---

[1] Although they tracked students' performance they did not use these data in their analysis because of the unsufficient reliability of the tracking program.

Manual logs are diverse types of forms in which students note down search words, search results, their conclusions, or the resources they use. Chambers and O'Sullivan (2004) and O'Sullivan and Chambers (2006) have used them to determine the types of changes students make when they are prompted to use a concordancing program to improve learner writing. Gaskell and Cobb (2004) employed this type of log in order to check whether learners used concordances to correct errors. In this study, however, manual logs were used in combination with Internet Protocol (IP) logs to double-check that participants did not just guess the answers or relied on other sources, assuming that concordancing had actually taken place. Ma (1994), Frankenberg-García (2005) and Varley (2009) also used reflective logs to research learners' strategies in analysing corpus data.

Activity logging software is gradually making its way into language learning research and the field of DDL is no exception. Since Johns's (1997) and Cobb's (1997) early attempts to obtain computer-generated logs, several other investigations have implemented similar approaches to gain insight into user behaviour when interacting with concordancers or other corpus-based resources. At the very beginning, logs were hoped to eventually "provide a useful source of evidence for research into student learning behaviour in a CALL environment, and how that behaviour may change over time" (Johns, 1997:108). Earlier approaches were rather "rudimentary" (Gaskell & Cobb, 2004:316) because of the impossibility of linking learners and behaviours and the authors acknowledge the need for a finer-grained tracking method that can offer more precise information. This concern goes hand in hand with Horst et al.'s (2005:107) suggestion that "we need to look more closely, as a step toward tying resource use to learning outcomes".

More recently, other authors seem to have incorporated this suggestion and have collected individualized evidence on corpus consultation to avoid regarding "learners as a monolithic group rather than as idiosyncratic individuals" (Yoon, 2008:32). The use of more refined tracking methods offers the possibility of registering exact queries, the time spent on each activity, the corpus or sub-corpus searched, or answers. Chan and Liou (2005), in a study investigating the effects of concordancing for the learning of verb-noun collocations, kept a record of the words participants looked up in the concordancer, the number of times each word was searched and also the answers to the exercises. Probably, the best-known example of tracking learner activity in corpus consultation is Hafner and Candlin's (2007) study on the use of online concordancing by a group of students enrolled in a legal writing course. For each participant, they collected information on: the date and time they accessed the resources, the referring page, the exact search query and the corpus or sub-corpus searched in an attempt to overcome the criticisms made to studies which do not provide evidence of direct corpus consultation or on those which offer pooled results instead of individualized ones. Similarly, Yoon (2008) collected corpus search logs reflecting patterns of corpus

use of six students in order to examine how corpus use affects their approach to L2 writing.

2.3 Guidance during corpus consultation

Unlike indirect pedagogical corpus applications, where corpora are accessed by expert linguists or materials designers, direct applications (DDL) entail teachers and learners using corpora in the language classroom (Römer, 2008:113). Due to the high degree of induction demanded by direct applications, metaphorical references to learners as researchers (Mauranen, 2004), detectives (Johns, 1997) or, in a wider sense, travellers (Bernardini, 2000a) are frequent.

References to the need for guidance are constant in empirical studies involving students searching and analysing corpus data. Assistance is most often offered in the form of lessons or seminars on the basics of Corpus Linguistics or concordancing prior to hands-on work with the tools, and only in a minority of studies are learners guided, either by the teacher or by contextual aids while they perform corpus consultation. Introductory lessons are either short sessions designed *ad hoc* for the experiment (Sun, 2003; Gilmore, 2009) or longer seminars on diverse aspects of Corpus Linguistics which were already part of the official curriculum (Bernardini, 2000b, 2002; Cheng, Warren and Xun-feng, 2003; Chambers & O'Sullivan, 2004, Yoon & Hirvela, 2004; Lavid, 2007; Varley, 2009; Kennedy & Miceli, 2010). In both cases, the seminars almost always comprise a theoretical overview on the nature and functions of linguistic corpora and a more practical session in which the functionalities of a concordancer or other tools for corpus analysis are demonstrated.

Instances of guidance which is provided immediately before students come into contact with the tools or at the same time resource use is taking place are less frequently found. A working example is the study conducted by Sun (2003) in which the subjects were given an individualized 30 minute tutorial including theoretical as well as practical aspects of concordancing followed by a hands-on corpus consultation session with teacher intervention. In a similar fashion, Lavid (2007) also provided her students with guidance while they were searching parallel and comparable corpora to find lexico-grammatical patterns in Spanish and English in the context of a contrastive linguistics course

Contextual aids can also be used to guide participants while corpus consultation is going on. This is the case when little or no previous training on corpora or concordancing techniques has been offered. The present study and Götz and Mukherjee (2006) employ written guidelines on what a corpus is or how to conduct searches and analysees which are inserted in the task medium or which are delivered together with the assignment as a set of instructions. Götz and Mukherjee (2006) opted

for the latter approach which they combined with incidental personal guidance at the request of participants.

On some other occasions, students become familiar with the tools and techniques progressively, by first completing highly controlled tasks in which the teacher provides search strings followed by less controlled tasks including independent corpus consultation once they have reached a certain degree of competence. Vannestål and Lindquist (2007) or Charles (2007) are advocates of this approach. Charles (2007:297) maintains that, given that time constraints do not allow for students developing corpus skills in class, they should be directed at the beginning to "searches known to be of value for the matter under investigation". This, she argues, does not come into conflict with discovery learning and can help students gain a general understanding of the techniques and methods of analysis behind concordancing and serve as a "jumping off point" (*ibid*). Kennedy and Miceli (2010) and numerous specialists cited by them (Bernardini, 2004; Chambers, 2005; Cheng et al., 2003; Yoon & Hirvela, 2004) are also in favour of initiating students in corpus consultation by means of teacher-guided activities.

3. Research question

This study addresses the actual use of corpus-based resources in two groups of university EFL learners while completing different focus-on-form activities under two research-conditions: guided and non-guided corpus consultation. Our research question is whether learner interaction with corpus-based resources differs under different corpus consultation conditions. To answer this question, we will examine the computer-tracked behaviour of learners in the guided and non-guided corpus consultation groups.

4. Method

4.1. Setting

The data were collected from two ELF groups of second year students in the degree of English Studies at a medium-sized NATIONALITY university during the spring semester of the 2008-2009 academic year. All the students involved in this research were enrolled in a compulsory EFL subject where training was provided to gain a C1 level of English, with a focus on use of English and all four skills. The classes met three times a week and lasted one hour each. Classes were taken during thirty weeks throughout the academic year.

4.2. Participants

39 learners took part in our research. However, despite completing the tasks, two of the learners did not make use of the corpus-based resources, so they were left out of our analysis. The informants in this study were 37 students of English as a foreign language (9 male and 28 female) whose average age was 20.65 years (SD = 1.4). All students had achieved a B1-to-B2 level of English, being 7.3 (out of 10) the average entry level mark (SD = 1.09). The majority of them were native-Spanish speakers, although there were 3 students whose mother tongue was not Spanish. Given the limitations which classroom-based language research imposes, and the academic restrictions enforced by pre-existing group arranging, the individuals could not be allocated randomly to the two research conditions. Consequently, it was the two groups that were randomly assigned to one of the two aforementioned research conditions. None of the learners had previously used corpora in this course or during the last semester.

Our control group (CG) consisted of 15 students whose average age was 20.27, while our experimental group (EG) was formed by 22 students with an average age of 20.91. The mean for the control group entry level mark was 7.8 out of 10 (SD = 1.17), while the mean for experimental group entry level mark was 7.07 out of 10 (SD = 1.00). The entry level mark distributions in the two groups did not differ significantly (Mann–Whitney U = 113, P < 0.11 two-tailed). Both groups were also homogeneous in terms of previous experience with corpus-based resources, as most participants declared being completely unfamiliar with linguistic corpora.

4.3. Operationalizations: corpus consultation

Whereas the control group completed a series of focus-on-form tasks without explicit directions on how to use a corpus, the experimental group was provided with corpus consultation guidance. Within the framework of this study, *corpus consultation guidance* is defined as a set of general guidelines which introduced students to working with a principled corpus by explaining them what a corpus is and how it can be used in class. These guidelines included a brief overview on the nature and uses of linguistic corpora, a description of the query interface of the BNC[2] and simple instructions on how to conduct basic searches, i.e, search for single words or strings of words, the use of tags to search for grammatical features (eg. ADJECTIVE + *of* + NOUN), and the use of wildcards to search for word combinations with elements in between (eg. *was ... by*). The more general guidelines were provided right above the tasks in the learning environment and the participants in the experimental group were asked to read them carefully before completing the tasks and to go back to them whenever they considered it necessary. Each "Search the Corpus" activitiy included more specific guidelines concerning the use of wildcards and tags. These specific

---

[2] We used the University of Leeds interface: http://corpus.leeds.ac.uk/protected/query.html

guidelines were embedded in the activity and were below 100 words of length. These guidelines followed the general principles outlined in Sinclair (2003) for the interpretation of concordance lines.

The students in the control group performed the tasks under *non-guidance* conditions. While the learning environment and the tasks were kept the same, they did not have access to the wider range of directions given to the experimental group. The tasks were delivered through the Moodle course management platform, where a course for each group of students was created. Figure 1 shows screenshots of both learning environments:
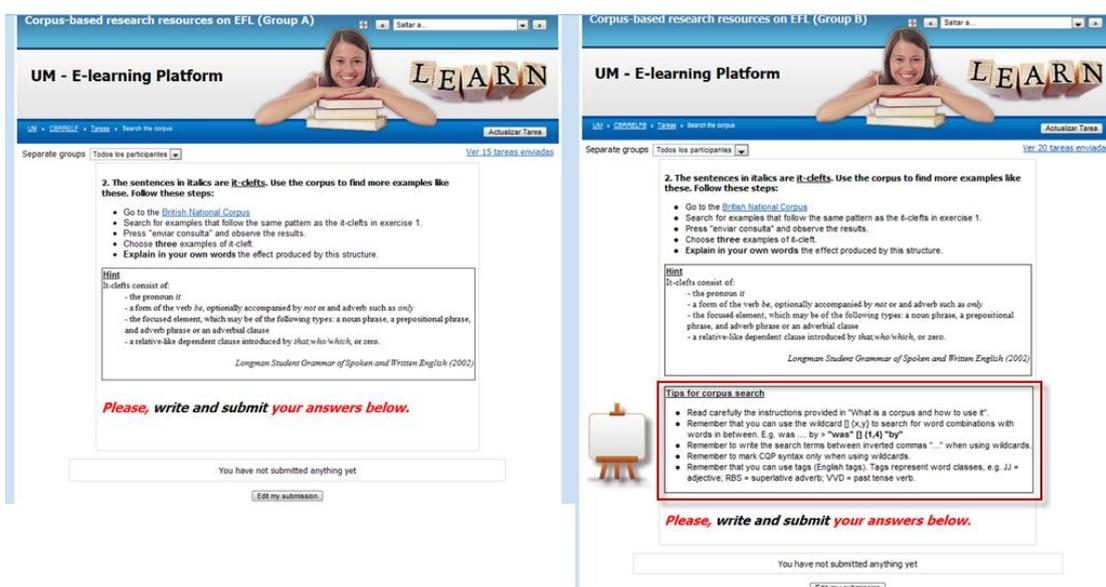

Figure 1. Experimental and control group's learning environment here.

4.4. Operationalizations: choice of tasks and procedures

The participants completed two focus-on-form tasks dealing with the use of emphatic structures based on a short story from their coursebook, *Objective CAE* (O'Dell & Broadhead, 2002). In each of the two Moodle courses there was a task dealing with *it*-cleft sentences and another addressing subject-verb inversion after negative/restrictive adverbial. The tasks gradually introduced the students to the form and use of these structures and presented the same configuration: an introductory activity (*Observe*), a hands-on activity using the BNC (*Search the corpus*), and a final activity (*Rewrite*) where students put into practice the structures under consideration. In the first activity of each task, students were presented with examples extracted from the BNC, but it was not until the second activity that they were asked to use the corpus themselves to find relevant sentences and infer patterns. This involves Sinclair (2003) *initiate* and *interpret* steps. In addition, interpreting the results in the form of concordance lines entailed vertical reading (Tognini-Bonelli, 2001) deducing patterns as well as formulating data-driven hypotheses (Sinclair, 2003), which were also new to them.

The *consolidate* and *report* steps (Sinclair, 2003) were confined to the final, rewriting activity of the task.

The experiment was conducted over a week and took three 60-minute sessions in total. Prior to the two data collection sessions in the computer lab, all the participants attended a regular lesson where they completed a series of listening and reading comprehension activities based on a short story adapted from their course book. In this session, the learners were exposed to contextualized uses of *it*-cleft sentences and inversion. Sessions 2 and 3 form the core of the study, as students were taken to the computer lab to perform the tasks described beforehand. During these sessions, students were explicitly asked to search the BNC through the link which was provided. Both groups were allowed to use whatever online resources (dictionaries, search engines, etc.) they deemed useful to complete the activities.

4.5 Operationalizations: computer-tracked behaviour of learners and log analysis

We tracked student-computer interaction by means of Fiddler[3] logs. Fiddler is a plug-in installed on the user computer which acts as a proxy intercepting all the communications between the user and the Internet. This tool for data collection assisted us in overcoming one of the limitations detected in corpus consultation studies pointed out by Hafner and Candlin (2007), i.e, the difficulty in linking individuals with their queries. Fiddler enables the tracking of web-browser actions and captures all the web pages visited, all the actions carried out on each web page and all the information typed in. We decided on a log-based tracking device as it offered us the chance to observe the learners' behaviour unobtrusively (Fischer, 2007:413). One of these logs in shown in Figure 2:

---

[3] Fiddler Web Debugging Proxy: http://www.fiddler2.com/fiddler2/
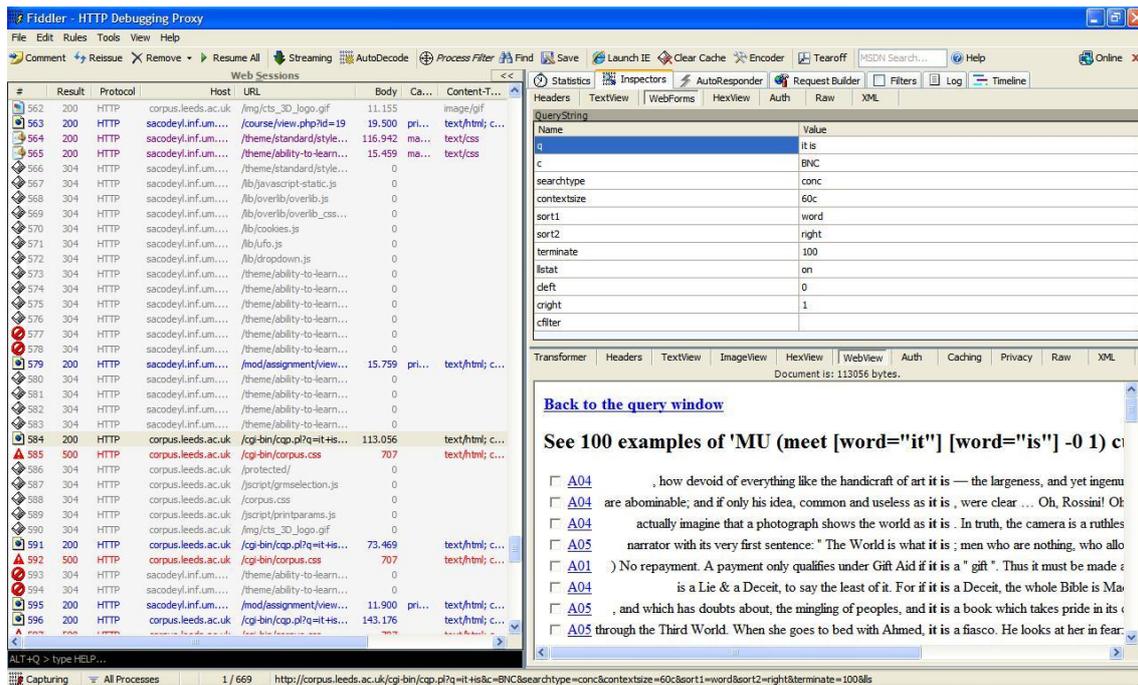
Figure 2. Log of student's behaviour here.

This huge amount of data was eventually processed by means of a tool that was developed *ad-hoc*. For example, web ads and server-generated scripts like CSS and JavaScript were wiped away and did not count as user actions. A key aspect in this process was determining the different queries performed by the learners while approaching the tasks in general, and consulting the BNC in particular.

4.5.1. Learner behaviour tracking parameters

The learner behaviour that was analyzed included (1) the number of events or actions performed by each individual, (2) the total of different web services used, (3) the total of Moodle activities completed, (4) the number of searches performed on the BNC and, last, (5) the number of words or wildcards per BNC search. Figure 3 summarizes our approach:
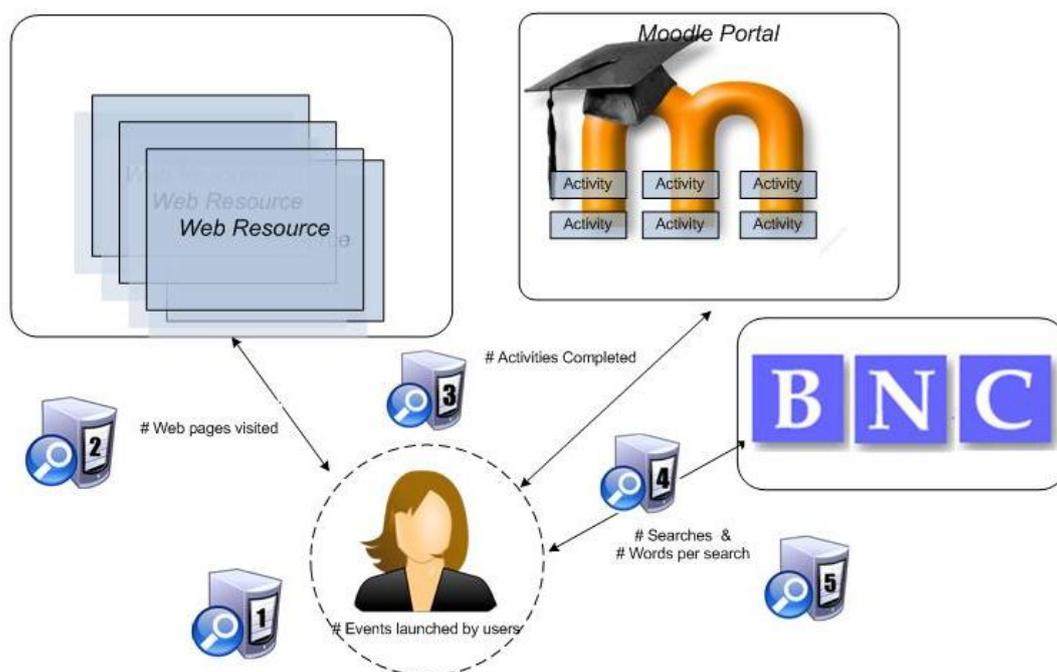
Figure 3. Parameters tracked here.

These parameters are what Yoon (2008) and Horst et al. (2005) consider precise descriptions of corpus consultation, as, first, they enable researchers to treat participants as individual units and not as a group, and, second, it is possible to link resource use to learner's performance.

5. Results

5.1. Log analysis: tracking the informants' behaviour
Table 1 shows the means and standard deviations of the means for the parameters that were tracked down in our study: number of browser events launched by the user, number of different web sites visited, the number of activities completed, number of BNC searches and, finally, number of items (words, wildcards and tags) per BNC search.

	Events launched		Web Services Visited		Activities completed		BNC Searches		Words per Search	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CG	39.1	12	4.7	1.9	4.3	1.45	3.5	1.8	2.2	0.81
EG	47.6	23	4.8	1.8	3.8	1.3	5.4	2.8	2.6	0.89

Table 1. Mean and standard deviation of parameters tracked

As the populations do not have a normal distribution in any of the five parameters, we decided on the use of non-parametric tests to check significance (Braun, 2007; Yeh et al., 2007).

5.1.1. Number of browser events launched

The learners in the control group, CG, the non-guided consultation condition, launched, on average, 8.5 browser events fewer than their counterparts in the experimental condition, EG, who launched 47.6 browser events on average during the completion of the two tasks. Figure 4 shows the mean and standard deviation of this parameter in both groups:
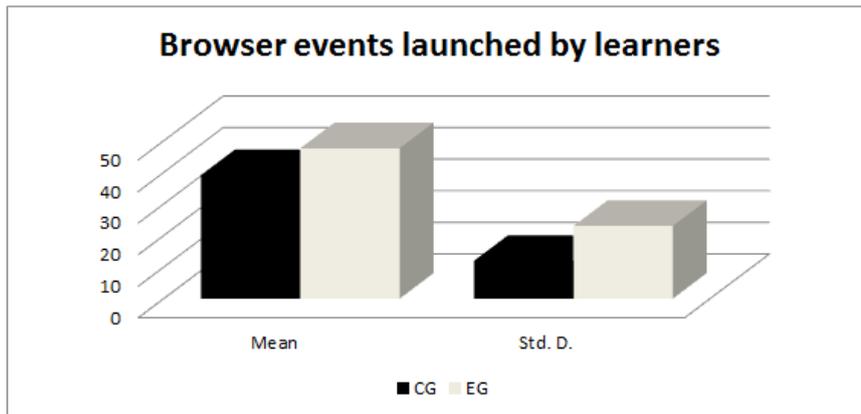

Figure 4. Browser events launched by learners here.

Despite being more packed with information, the guided-consultation condition did not impair productivity measured in terms of browser events launched by the learners. The distribution of the number of browser events parameter in the two groups did not differ significantly (Mann–Whitney $U = 192$, $P < 0.21$ two-tailed).

5.1.2. Different websites visited

Although the number of different web services used per informant does not apparently seem to be affected by the grouping condition (4.7, SD 1.9 for CG vs. 4.8, SD 1.8 for EG), the distribution of this parameter in the two groups did differ significantly (Mann–Whitney $U = 162$, $P < 0.04$ two-tailed). Figure 5 shows the mean and standard deviation of this parameter in both groups:
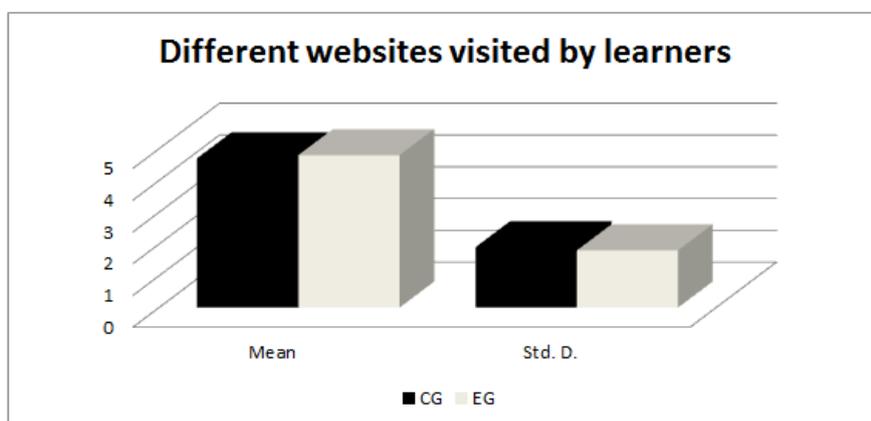

Figure 5. Different websites visited by learners here.

Individuals in the guided-consultation condition visited slightly more different websites than those in the non-guided condition. As for the websites themselves, the range of websites that were visited by more than one informant, 10, is limited. Figure 6 shows the normalized frequency (per 100 individuals) of visitors to these websites for both groups:
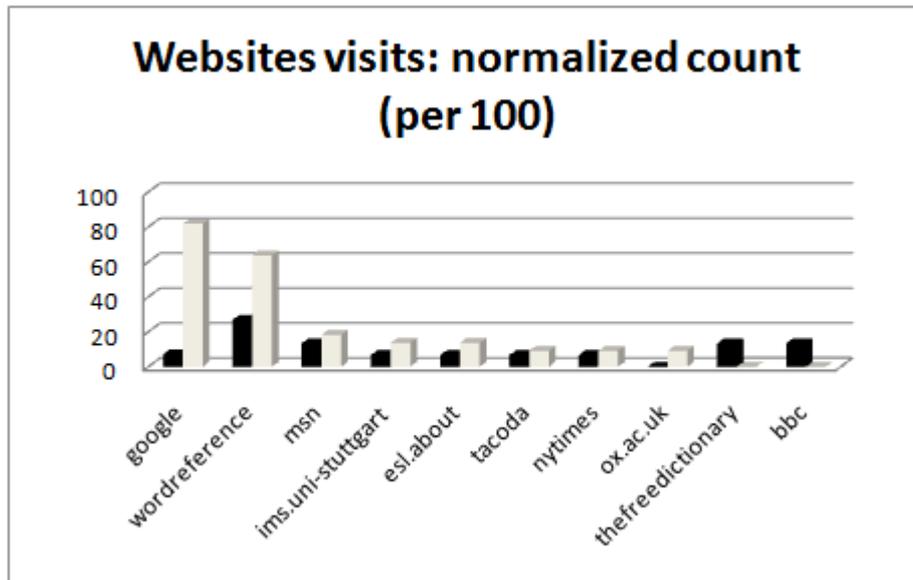

Figure 6 Websites visits: normalized count (per 100) here.

Google ranks up in EG with 82 informants, with wordreference.com coming up second with 64 and msn.com in third place with 19 individuals visiting this website while completing the tasks. In CG, wordreference.com is visited by 1 in 4 individuals, msn.com by 1 in 7 and google services by 1 in 16. These results clearly show that those in EG resorted more frequently to web resources than their counterparts in CG.

5.1.3. Activities completed

The total numbers of activities completed (4.35, SD 1.45 for CG vs. 3.8, SD 1.3 for EG) is higher in the non guided-consultation condition. The participants were asked to complete two focus-on-form tasks which were each split into an introductory activity (Observe), a hands-on activity using the BNC (Search the corpus), and a final activity (Rewrite), totaling 6 activities. Figure 7 shows the mean and standard deviation of this parameter in both groups:
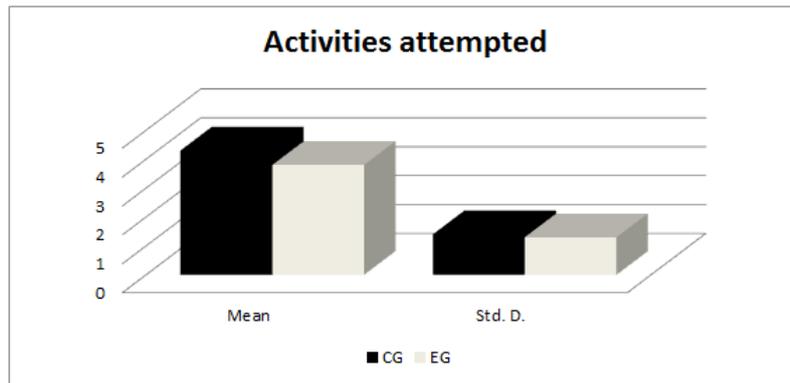
Figure 7. Activities attempted here.

The individuals in both groups found it hard to complete the 6 activities. As seen in 4.4, the initiate, interpret, consolidate and report steps (Sinclair, 2003) involved highly demanding cognitive processes that posed important challenges to the individuals. The very similar standard deviation in both groups points to the presence of a common pattern which seems not be influenced by the exposure to guided-consultation guidelines in EG. As in the case of the number of browser events launched by the learners, the distribution of this parameter in the two groups did not differ significantly (Mann–Whitney U = 194, P < 0.21 two-tailed).

5.1.4. Number of BNC searches

The total number of BNC searches is higher in the guided-consultation condition (3.5, SD 1.8 for CG vs. 5.4, SD 2.8 for EG). Figure 8 shows the mean and standard deviation of this parameter in both groups:
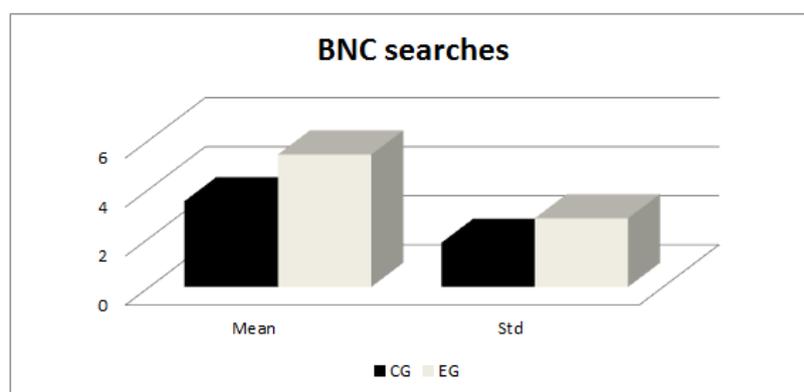

Figure 8. BNC searches here.

The informants in EG performed 5.4 BNC searches, while those in the non-guided condition did on average nearly 2 searches (1.9) fewer. This finding shows that the individuals in the guided-condition used the BNC more extensively than those in CG. This finding turned out to be statistically significant as the distribution of the number

of BNC searches in the two groups differed significantly (Mann–Whitney U = 150.5, P < 0.02 two-tailed).

5.1.5. Number of words, wildcards and tags per search

The number of items per search in both groups was very similar (2.2, SD 0.8 for CG vs. 2.6, SD 0.9 for EG). No wildcards or tags were used at all. Contrary to the mean number of searches, the distribution of this parameter in the two groups did not differ significantly (Mann–Whitney U = 184, P < 0.15 two-tailed). Figure 9 shows the mean of words used per BNC search:
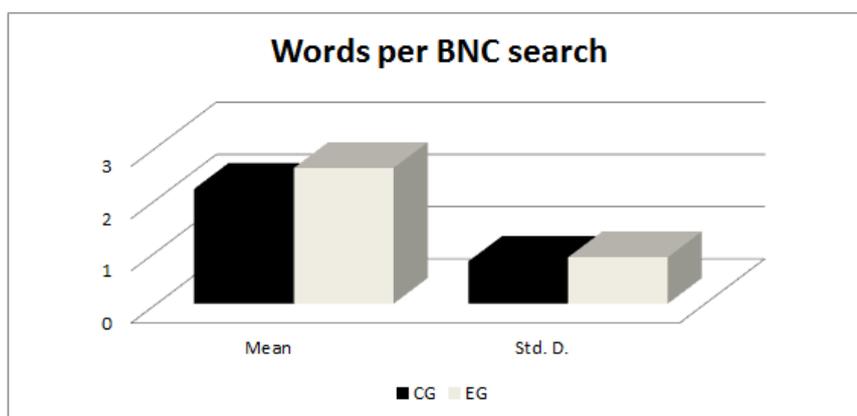

Figure 9. Words per BNC search here

The learners used less than 3 words each time they searched the BNC.
The words that were used in the search are of great interest. In the case of *it*-cleft sentences the prompts given to the learners,

(a) *It is policewomen who deal with both victims and offenders.*
(b) *It is through details that everything can be shown.*

motivated searches like

(c) *it is (CG)*
(d) *it was (CG)*
(e) *it is the (EG),*

which point to a lack of refined corpus skills as the learners failed to understand that using this input would return a massive amount of examples to go through. Searches of 3 or more words typically include

(f) *it was who (EG)*
(g) *it was only who (CG)*
(h) *it was which (CG),*

which shows that even when a wh-word is included in the pattern, there still persists a reluctance to using wildcards or tags. In the case of (f), the learner found 68 results where, more often than not, emerges a pattern consisting of a post subject pronoun that

is, in turn, post modified by the *it*-clause, like in "He it was who premièred Symphonies Nos. 5, 6, 8, 9 and 10"[4]. Other searches included fixed patterns like

(i) *it was you that (CG)*,

which returns 5 hits which are good examples of the use of an *it*-cleft sentence. In the case of inversions, the prompts given to the learners,

(j) *no sooner are the terrorists under control than there is a second [...]*

(k) *hardly had any pope become established and negotiations opened, than set up the World Council for [..]*

motivated searches like

(l) *no sooner (CG),*

(m) *not only (EG),*

which returned 242 and 14245 hits, respectively. (l) and (m) represent two interesting, yet opposing, cases where learners resort to the most expeditious search items with opposite results. While in (l) learners can relate immediately to the inversion pattern, in (m) learners will be mostly diverted by the not only…but pattern. Other searches like

(n) *hardly can (CG)*

produced 3 hits and no inversion pattern; or

(o) *little were (EG)*,

which returned no matching records at all.

6. Discussion

The results of our research provide learner data on two areas of relevance to the scientific community. The first area is research on the use of corpus-based resources in EFL, and, in particular, on the use of principled corpora in university foreign language learning contexts. The second addresses technical issues of concern in the use of direct research methodologies: implementation, use and challenges for future research.

6.1. Use of corpus-based resources in foreign language learning: using direct observation methodologies

Analysing the learners' interaction with corpus tools is a step forward in both understanding the way individuals use these resources (Chapelle, 2003) and assisting researchers in problematising work with corpora in the language classroom (Boulton, 2009b).

The authors that make use of indirect observation methodology have mainly researched learners' attitudes toward corpus use in the L2 classroom. Some of this research claims to analyze "students' corpus use behaviour" (Yoon & Hirvela, 2004:257), equating the analysis of *behaviour* with the information gathered mainly

---

[4] BNC source: Gramophone. Harrow: General Gramophone Pubcs Ltd, 1992.

through questionnaires. While these studies provide an excellent opportunity to reflect on the benefits and limitations of using corpora to teach and learn language (Bernardini, 2002; Chambers, 2005; Götz & Mukherjee, 2006; Kennedy & Miceli, 2001; Ma, 1994), the use of indirect observation methodologies may confine the scope of research to the learners' opinions about the benefits of using corpora for language learning, their self-perceived difficulties in consulting corpora and feelings as well as their own evaluation of corpus use during the task (Lee & Swales, 2006; Yoon & Hirvela, 2004; Cheng, et al., 2003). However, it is necessary to go beyond the opinions of learners to enrich our understanding of how corpus-based resources can be used to enhance language learning and teaching.

Our research offers learners' data which is based on unobtrusive, direct observation of their interaction with corpora. We found that 14 out of 15 learners in the guided-consultation group actually accessed the guidelines provided by the researchers, something of tremendous importance to claim experiment validity. Similarly, the initial 39 informants were trimmed down to 37 as two of the learners, despite completing the tasks, did not search the BNC whatsoever. It may be argued that the percentage of students who did not use the BNC, 6%, is low, but this evidence shows that, even when a learner completes tasks and fills in a questionnaire giving an opinion on his experience with the corpus-resources, researchers should be cautious as to the extent to which actual interaction with the corpus has taken place at all. Log analysis actually gives researchers the chance to sort out the learners that consult the corpus and, going further still, tell those who actually perform meaningful, outcome-oriented searches on the corpus. Researchers should be aware of the methodological shortcomings (Fischer, 2007; Chapelle & Mizuno, 1989) that their analysis faces if indirect observation is used exclusively. Based on the certainty that the individuals in our research did actually access the consultation guidelines (EG) and searched the corpus (CG & EG), our results provide data-driven insight into the way learners interact with corpus-based resources.

Did learner interaction with corpus-based resources differ under different corpus consultation conditions? Individuals in EG launched a similar amount of browser events as their counterparts in CG. While the data itself cannot be interpreted at this exact stage of CALL research either as conveying a positive or a negative bias in qualitative terms, it seems to us worth noting that the experimental condition of EG did not adversely affect the number of browser events that were launched by the learners. In other words, exposure to a greater amount of information in EG did not affect negatively this measure. This was corroborated by the Mann–Whitney test, which shows that the distribution of this parameter in the two groups did not differ significantly. Although more in-depth research will be necessary to evaluate the relationship between browser events and productivity in the context of CALL, this finding is of great interest in the process of integration of guided-corpus consultation

within the framework of formal education since it suggests that guided-consultation does not detract hands-on, activity time. In both groups the ratio of browser events per activity completed is high: 9.09 in CG and 12.5 in EG. This quasi-frenzied activity can be explained on the account of familiarity of the informants with digital media. Not surprisingly, these individuals have spent their youth in direct contact with digital media and online communication, a pervasive part of their lives (Ito et al., 2009).

In the context of our experience, learners were given the chance to use web sites in order to complete their tasks. Did they resort to these resources, or did they just limit themselves to using the BNC link provided? Every single learner, irrespective of the research condition, used some online resource while completing the tasks. This fact corroborates accounts of digital nativeness: "These kids […] study, work, write, and interact with each other in ways that are very different from the ways that you did growing up" (Palfrey & Gasser, 2008: 2). The question immediately arises whether corpus consultation is to be understood in a new learning context which goes beyond the strict boundaries of the corpus being queried, the BNC in our case. Palfrey and Gasser (2008: 4) maintain that for most kids born in Western societies after 1980 part of their identity is integrated by practices which include a "tendency to multitask" and an online "pattern of using the technologies to access and use information". We can no longer expect that learners which are "constantly connected" (Palfrey & Gasser, 2008: 5) refrain from seeking information on the Internet. Our research corroborates this tendency.

The integration of these "other" online resources may be one more step in the process of normalization which Bax (2003:24) put forward. For this author, CALL will be normalized "when [it] will not be the centre of any lesson but [it] will play a part in almost all […] other aspects of classroom life, alongside coursebooks, teachers and notepads […] will almost go unnoticed". The omnipresence of highly-linguistic loaded interfaces in principled corpora is not only a trace of the possibilities scenario (Pérez-Paredes & Alcaraz, 2009; Pérez-Paredes, 2010), in other words, the use of research-oriented lexical databases in pedagogic contexts, but also the trace of the overreliance on stand-alone corpora for classroom-language learning, which may be jeopardizing the normalization of these resources in mainstream education (Boulton, 2009a; Pérez-Paredes, 2010). This normalized vision of the uses of corpora contrasts with problems reported by students and researchers in numerous studies (Bernardini, 2000b; Chambers, 2005; Chambers & O'Sullivan, 2004; Cheng et al., 2003; Kennedy & Miceli, 2001, 2002a; Lavid, 2007; Vannestål and Lindquist, 2007; Yoon & Hirvela, 2004; Yoon, 2008) and by the need expressed by Mauranen (2004:99) to favour an integration of corpus resources which causes no "extra hassle" in the classroom.

By tracking the websites visited by the learners, different patterns of use in the two research conditions were unveiled. First, CG used these online resources to a lesser extent than EG. For CG individuals, wordreference.com was their favourite

destination, with 26[5] in 100 individuals choosing this resource. This site offers both monolingual and bilingual dictionaries as well as forums where users can refine their searches or ask for help in case they are looking at extended units which are not recorded in the dictionaries. Msn.com[6] came second, with 13 in 100 individuals looking up information here. Google came third, only 7 in 100 individuals made use of this site. The situation in EG was very different: 82 individuals in 100 used google as a resource, 63 used wordreference.com and 18 msn.com. EG learners diversified and used more resources simultaneously than CG individuals. Many of these searches on wordreference.com were simple word look-ups: *cleft*, *emphasis*, *corpus*, while others, especially those on google, were more elaborated: *emphatic effects*, *relative clauses* or *it is emphasis*.

These results point to the existence of different types of learners' approaches to corpus use (Hafner & Candlin, 2007). Future research may explore these findings based on learners' beliefs and direct observation methodologies. Although from the sample sizes and the research design cannot be concluded that the different research conditions prompted different approaches to online resources, the Mann–Whitney test shows that the distribution of this parameter in the two groups differed significantly. Future research should examine in detail the relationships between corpus work and the range of different web services used by students.

The completion of the activities involved putting into practice corpus-consultation sub-skills, namely, the initiate, interpret, consolidate and report steps suggested by Sinclair (2003). According to the Mann–Whitney test, the distribution of the number of activities in the two groups did not differ significantly, so we cannot claim that the groups behaved distinctively under the two research conditions. The individuals in the guided consultation condition had to access, and actually did access, extra information[7] which illustrated best-practices in consulting corpora. This extra time did not have an impact on the number of activities completed. This finding suggests that, despite the extra time devoted to reading and processing the guidelines, guided-consultation of corpus resources does not impair productivity as measured by the total number of activities completed by the learners. It can be assumed that if the experiment would have continued in time, EG would have caught up and eventually outperformed CG. Future research should address this issue throughout a longer period in order to cast more conclusive evidence.

Despite completing fewer activities, the learners in the guided-consultation condition searched the BNC almost three times more than the individuals in the non-guided

---

[5] Normalized counts.
[6] At the time, bing.com had not already been launched.
[7] A 391-word text which included 4 figures with self-explanatory how-to captions plus a 98-word text with clear guidelines to use wildcards and tags.

consultation. This behaviour can explain the lower rate of activity completion in EG. The informants in EG performed a mean of 5.4 BNC searches, while those in the non-guided condition searched the BNC 3.5 times. According to the Mann–Whitney test, the number of BNC searches parameter is statistically significant. The ratio which expresses the relationship between the BNC searches and the number of activities completed is revealing. For CG this ratio is 0.81, that is, individuals in CG used the BNC 0.81 times per activity. For EG, this ratio is 1.42, in other words, the learners in the guided-consultation condition used the BNC 1.42 times per activity completed. These apparently low rates corroborate the conclusions of previous researchers who found that students found it difficult to choose which terms to use in their searches (Cheng et al., 2003; Lavid, 2007) and the way to approach the use of corpora in general (Bernardini, 2000b; Chambers, 2005; Chambers & O'Sullivan, 2004; Cheng et al., 2003; Kennedy & Miceli, 2001, 2002a; Lavid, 2007; Vannestål and Lindquist, 2007; Yoon & Hirvela, 2004; Yoon, 2008).

The fifth parameter which was analyzed in our study was the number of words and wildcards per BNC search. The learners in both groups used less than 3 words and/or wildcards when searching the BNC, and according to the Mann–Whitney test, the distribution of this parameter in the two groups did not differ significantly.

Despite the lack of statistical significance, the learners in the guided-consultation condition used 0.4 words more per search, which seems more in tune with the demands of the proposed activities in our research, namely, *it*-cleft sentences and the use of inversion for emphasis. Nevertheless, this will require further examination as learners, despite the information provided in both research conditions, may rely on corpus consultation following a traditional dictionary look-up approach rather than one more driven by the discovery of patterns (Sinclair, 2003) and the need to clarify the meaningful units of language (Sinclair, 2004).

We believe that based on the type of evidence provided in this paper, future research will benefit greatly from the tracking methodology advocated here.

6.2. Implementation, use and challenges of computer behaviour tracking in corpus linguisctics

The empirical evidence of the actual use that learners make of corpus-based resources reported in this paper can be used to shed light on different areas of CALL applications (Fisher, 2007) and, more specifically, on the use of corpus resources in the context of language education. The use of unobtrusive research methodology proves relevant in this context, although the problems and challenges which this use entitles are not to be overlooked by researchers in the field. As put by Gaskel and Cobb (2004:317):

"It is often noted that the various educational uses of concordancing are more talked about than tested with real learners. […] The choices at this point would seem to […] work on a database of approaches, interfaces, and learner behaviours within a research-and-development perspective".

However, our account would be incomplete if we did not highlight the difficulties that using a tracking methodology presents. In our case, these implied the installation, testing and maintenance of Fiddler (c.f. 4.5) in a computer lab and the *ex novo* creation of an information processing tool that could transform these logs into meaningful data for researchers. While our research shows that it is possible to implement tracking methodologies, our experience demonstrates that fine-grained analyses are costly and time-consuming.

7. Conclusion

Our research shows that the two groups of learners behaved in a different way on two of the parameters analysed: the number of different websites visited and the number of BNC searches performed. While these results are to be read very cautiously given the number of individuals involved (N=37), the role of corpus consultation guidance needs to be further addressed taking into account the types of direct observation methodologies advocated here.

Tracking corpus consultation is beneficial from different perspectives, ranging from examining learners' actual use of the resources for mere work assessment purposes to identifying sources of difficulty that can subsequently inform the creation of training materials, or even the redesign of corpus-based software or web services founded on authentic observed behaviour (Johns, 1997; Bloch, 2009). Our results corroborate empirically the suggestions found in the literature that skills and guidance are necessary when teachers take a corpus to the classroom (Mauranen, 2004; Cheng, et al., 2003). Our results also corroborate that user tracking is essential to claim research and results validity.

Independently of the type of guidance provided, specialists in the field agree that autonomous corpus consultation is a long-term achievement which requires training (Cheng et al., 2003; Chambers & O'Sullivan, 2004; Kaur & Hegelheimer, 2005; O'Sullivan, 2007; Vannestål and Lindquist, 2007). Götz and Mukherjee (2006:59) call for more research in the delimitation of corpus literacy and argues that it "seems to be *the* (emphasis in the original) most central prerequisite for a successful implementation of DDL activities in the English language classroom". We expect that computer tracking will play a key role in this process. Future research should strive to integrate direct and indirect observation methodologies and thus take full advantage of the affordances that both methodologies can offer in the field of CALL and DDL.

Pascual Pérez-Paredes completed his PhD in Applied Linguistics in 1999 after a research stay at The Univesity of Texas at Austin. His main interests are quantitative research of register variation, the compilation and use of language corpora and the implementation of Information and Communication Technologies in foreign language teaching/learning. He has been project coordinator of a MINERVA initiative funded by the European Commission: SACODEYL.

María Sánchez-Tornel holds an MA in Applied Linguistics from the University of Murcia. She is currently completing her doctoral thesis on the integration of corpus-based resources and corpus consultation in the EFL classroom. She has participated in several national and European research projects related to the field of CALL. Her main research interests are data-driven learning, corpus literacy, learning strategies, and computer-mediated communication.

Jose María Alcaraz Calero received the Computer Engineering and Master degrees with honours at the University of Murcia in 2004, where he has been working since then in several European and International projects while completing his PhD. Currently, he is a research fellow at Hewlett Packard Laboratories, Bristol, UK. His research areas include knowledge representation, semantic web, corpus-based systems, information retrieval and e-learning platforms.

Pilar Aguado took her degree in English Studies at the University of Murcia (1983/88), and joined the English Department at the University of Murcia in 1990. She taught Business English at the Campus of Cartagena Business School (1990/99) and General English at the Faculty of Arts (1990/-). She took her PhD in 1997. Other posts and offices include: Visiting Scholar, UCSD (1994); Editorial Secretary, Cuadernos de Filología Inglesa, now IJES (1993/95); Research Scholar, University of Alicante (1998/99); Language Advisor, CAGE Panel, Cambridge University Press (2003/05). Her main current areas of research are Teaching English as a Foreign Language, Translation, Materials Design and ESP.

References

Abdel Latif, M.M. (2008) A state-of-the-art review of the real-time computer-aided study of the writing process. *International Journal of English Studies, 8*(1), 29-50. Retrieved from http://revistas.um.es/index.php/ijes/article/view/49081

Aston, G., Bernardini, S., & Stewart, D. (Eds.). (2004). *Corpora and language learners*. Amsterdam: John Benjamins.

Bax, S. (2003). CALL—past, present and future. *System, 31*, 13-28.

Beeby, A., Rodríguez Inés, P., & Sánchez-Gijón, P. (Eds.). (2009). *Corpus use and translating*. Amsterdam: John Benjamins.

Bernardini, S. (2000a). *Competence, capacity, corpora: A study in corpus-aided language learning*. Bologna: CLUEB.

Bernardini, S. (2000b). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard, & T. McEnery (Eds.). *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora* (pp. 183-190). Frankfurt: Peter Lang.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Ketteman, & G. Marko (Eds.). *Teaching and learning by doing corpus analysis* (pp. 165-182.). Amsterdam: Rodopi. Retrieved from http://docserver.ingentaconnect.com/deliver/connect/rodopi/09215034/v42n1/s14.pdf?expires=1274809284&id=56951697&titleid=1000&accname=Guest+User&checksum=F32B3255DD49CE1741E60D9BBC313EDC

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: John Benjamins.

Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Leaning and Technology, 13*(1), 59-78. Retrieved from http://llt.msu.edu/vol13num1/bloch.pdf.

Boulton, A. (2009a). Testing the limits of data-driven learning: language proficiency and training. *ReCALL, 21*, 37-54.

Boulton, A. (2009b). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics, 35*, 81-106.

Braun, S., Kohn, K., & Mukherjee, J. (Eds.). (2006). *Corpus technology and language pedagogy*. Frankfurt: Peter Lang.

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*, 307-328.

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom. Selected papers from TaLC 2004* (pp. 3-16). Amsterdam: Rodopi.

Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL, 16*, 158-172.
Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology, 9*(2), 111-125. Retrieved from http://llt.msu.edu/vol9num2/chambers/default.html

Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, *18*, 231-250.

Chapelle, C. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. Philadelphia: John Benjamins Publishing Company.

Chapelle, C., & Mizuno, S. (1989). Students' strategies with learner-controlled CALL. *CALICO Journal*, *7*, 25-47.

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*, 289-302.

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

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

Desmarais, L., Duquette, L., Renié, D., & Laurier, M. (1998). Evaluating learning and interactions in a multimedia environment. *Computers and the Humanities*, *31*, 327-349.

Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behaviour in CALL. *Computer Assisted Language Learning*, *20*, 409-442.

Flowerdew, L. (2004). The argument for using English specialized corpora. In U. Connor, & A. Upton (Eds.). *Discourse in the professions:Perspectives from corpus linguistics* (pp. 11-33.). Amsterdam: John Benjamins.

Frankenberg-García, A. (2004). Lost in Parallel Concordances. In G. Aston, S. Bernardini, & D. Stewart (Eds.). *Corpora and language learners* (pp. 213-229.). Amsterdam: John Benjamins.

Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography, 18*, 335-355.

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

Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, *63*, 363-372.

Goodfellow, R., & Laurillard, D. (1994). Modeling learning processes in lexical CALL. *CALICO Journal*, *11*, 19-46. Retrieved from https://calico.org/html/article_571.pdf

Götz, S., & Mukherjee, J. (2006). Evaluation of data-driven learning in university teaching: a project report. In S. Braun, K. Kohn, & J. Mukherjee (Eds.). *Corpus technology and language pedagogy* (pp. 69-86.). Frankfurt: Peter Lang.

Hafner, C.A., & Candlin, C.N. (2007). Corpus tools as an affordance to learning in professional legal education'. *Journal of English for Academic Purposes,* 6, 303-318.

Hidalgo, E., Quereda, L., & Santana, J. (Eds.). (2007). *Corpora in the foreign language classroom*. Amsterdam: Rodopi.

Horst, M., Cobb, T. & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning & Technology, 9*(2), 90-110. Retrieved from http://llt.msu.edu/vol9num2/pdf/horst.pdf

Ito, M., Baumer, S., Bittanti, M., Boyd, D., Cody, R., Herr-Stephenson, B., … Tripp, L. (Eds.). (2009). *Hanging out, messing around, and geeking out: living and learning with new media*. Cambridge, MA: The MIT Press.

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. 110-115.). Harlow: Addison Wesley Longman.

Kaur, J. & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: an exploratory study. *Computer Assisted Language Learning, 18*, 287-310.

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

Kennedy, C., & Miceli, T. (2002a). The CWIC project: Developing and using a corpus for intermediate Italian students. In B. Ketteman, & G. Marko (Eds.). *Teaching and learning by doing corpus analysis* (pp. 183-192.). Amsterdam: Rodopi. Retrieved from http://docserver.ingentaconnect.com/deliver/connect/rodopi/09215034/v42n1/s15.pdf?expires=1274820527&id=56954272&titleid=1000&accname=Guest+User&checksum=0DC2674F857F67A1CC41850CAF29956C

Kennedy, C., & Miceli, T. (2002b). An apprenticeship with the CWIC corpus: A tool for learner writers in Italian. In C. Kennedy (Ed.), *Innovations in Italian Teaching Workshop* (pp-83-94). Brisbane: Griffith University Press

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate italian learners to a corpus as a reference resource. *Language Learning and Technology, 14(1)*, 28-44. Retrieved from http://llt.msu.edu/vol14num1/kennedymiceli.pdf

Lavid, J. (2007). Contrastive patterns of mental transitivity in English and Spanish: a student-centred corpus-based study. In E. Hidalgo, L. Quereda, & J. Santana (Eds.). *Corpora in the foreign language classroom* (pp. 237-252.). Amsterdam: Rodopi.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*, 56-75.

Leijten, M. & Van Waes, L. (2005). Inputlog: An input logging tool for the research of writing processes. Research paper, University of Antwerp.

Ma, K. C., (1994). Learning strategies in ESP classroom concordancing: An initial investigation into data-driven learning. In J. Flowerdew & A. Tong (Eds.), *Entering texts* (pp. 197-214). Hong Kong: Language Centre, The Hong Kong University of Science and Technology.

Mauranen, Anna. 2004. Spoken - general: Spoken corpus for an ordinary learner. In John McH.Sinclair (ed.) *How to Use Corpora in Language Teaching* (89–105). Amsterdam: John Benjamins.

O'Dell, F., & Broadhead, A. (2002). *Objective CAE. Students' Book Self-Study*. 2nd Edition. Cambridge: Cambridge University Press.

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*, 269-286

O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, *15*, 49-68.

Palfrey, J., & Gasser, U. (2008). *Born digital. Understanding the first generation of digital natives*. New York: Basic Books.

Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: appropriation and the possibilities scenario. In T. Harris & M. Moreno Jaén (Eds.), Corpus Linguistics in Language Teaching. Bern: Peter Lang.

Pérez-Paredes, P. y Alcaraz-Calero, J.M. (2009). Developing annotation solutions for online data driven learning. *ReCALL, 21*, 55-75.

Römer, U. (2008). Corpora and language teaching. In A. Lüdeling, & M. Kytö (Eds.). *Corpus Linguistics. An International Handbook (Volume 1)* (pp. 112-130.). Berlin: Mouton de Gruyter.

Santos, D., & Frankenberg-García, A. (2007). The corpus, its users and their needs: A user-oriented evaluation of COMPARA. *International Journal of Corpus Linguistics, 12*, 335-374.

Sinclair, J. (2003). *Reading Concordances: An Introduction*. Harlow: Longman.

Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. M. Sinclair (Ed.). *How to use corpora in language teaching* (pp. 271-299.). Amsterdam: John Benjamins.

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

Sun, Y., & Wang, L. (2003). Concordances in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning, 16*, 83-94.

Thurstun, J., & Candlin, C. (1998). Concordancing and the Teaching of the Vocabulary of Academic English. *English for Specific Purposes, 17*, 267-280

Tognini-bonelli, E. (2001). *Corpus Linguistics at work*. Amsterdam: John Benjamins.

Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL, 19*, 329-350.

Varley, S. (2009). I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning, 22*, 133-152.

Yeh, Y., Liou, H. & Li, Y. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning, 20*, 131-152.

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology. 12*, 31-49. Retrieved from http://llt.msu.edu/vol12num2/yoon.pdf

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