Learners' search patterns during corpus-based focus-on-form activities

A study on hands-on concordancing

Pascual Pérez-Paredes, María Sánchez-Tornel &
Jose M. Alcaraz Calero
Universidad de Murcia, Campus Mare Nostrum / Hewlett Packard Laboratories Bristol

Our research explores the search behaviour of EFL learners (n=24) by tracking their interaction with corpus-based materials during focus-on-form activities (*Observe*, *Search the corpus*, *Rewriting*). One set of learners made no use of web services other than the BNC during the central *Search the corpus* activity while the other set resorted to other web services and/or consultation guidelines. The performance of the second group was higher, the learners' formulation of corpus queries on the BNC was unsophisticated and the students tended to use the BNC search interface to a great extent in the same way as they used Google or similar services. Our findings suggest that careful consideration should be given to the cognitive aspects concerning the initiation of corpus searches, the role of computer search interfaces, as well as the implementation of corpus-based language learning. Our study offers a taxonomy of learner searches that may be of interest in future research.

Keywords: Data Driven Learning, computer-assisted learner tracking, learner search behaviour, information retrieval

1. Introduction

The uses of language corpora in foreign language learning have been widely discussed over the last quarter of a century. In particular, the wealth of monograph volumes addressing the use of corpora in language teaching during the last decade (Sinclair 2004, Aston et al. 2004, Aijmer 2009, Braun et al. 2006, Hidalgo et al. 2007, Campoy et al. 2010, Bellés-Fortuño et al. 2010, Moreno et al. 2010) shows the relevance of the topic and, most significantly, the benefits of corpora for the

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

acquisition of languages. However, research in this area has neglected the study of hands-on uses of corpora in the language classroom in conjunction with other online services such as Google and dictionaries.

Linguists and educators who have promoted the use of corpus linguistics in the language classroom have drawn from Johns (1986) and Tribble & Jones (1997) the rationale for the use of Key Word in Context (KWIC) and concordance lines. Pérez-Paredes (2010) maintains that, in this tradition, the methods of research in corpus linguistics have been transferred to the language classroom, turning linguists' analytical procedures into a pedagogically-relevant tool for the use of authentic corpus data that can increase both learners' sensitivity to collocation-related phenomena and language learning strategies. In this context, much emphasis has been laid on the procedures for the compilation of language data, the uses of concordance lines, and the potential of corpus-based materials to unveil patterns of language use which had received little attention in the structuralist and Chomskyan traditions. Although it was in Johns (1986) where language instructors were introduced to the possibilities of Data-Driven Learning (DDL), the standard method for corpus-based research was further developed by Sinclair (1991) and Sinclair (2003: xiv–xv) and has been used steadily for over a quarter of a century now. Over a decade ago, Conrad (1999: 2) summarized research that already stressed the usefulness of concordancing "for vocabulary and grammar development" on the grounds that it promotes "the use of authentic language, makes students more active and independent analysers of language, and provides empirical evidence about language use".

In different ways, the use of concordancers, concordance lines and language corpora promoted the use of information technologies and active search strategies, which was perceived as a valuable asset. Recently, Johansson (2009: 41) claimed that more systematic studies are needed in order to test the benefits of DDL and that it is necessary to discuss "students' problems with corpus investigation" so that specialists can "suggest how [learners] could be better equipped to be corpus researchers". Nowadays, these problems and circumstances in university learning cannot be fully understood without the use of other online technologies and approaches which have become ubiquitous, at least in Western and Asian societies (Schroeder et al. 2010). What is more, learners now in the early years of the second decade of the 21st century have grown used to technologies and study habits which simply did not exist when most of the previous research on corpus use was carried out. Palfrey & Gasser's (2008) popular book on the new generation of digital natives points out how traditional institutions are still struggling with the integration of new technologies into everyday teaching and learning. The following quote is self-explanatory (Palfrey & Gasser 2008: 239): "The Internet is changing the way that children — and college students — gather and process information

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

in all aspects of their lives. For Digital Natives, "research" is more likely to mean a Google search than a trip to the library. They are more likely to check in with the Wikipedia community or to turn to another online friend, than they are to ask a reference librarian for help". Given the incontrovertible appeal of DDL in university contexts (Boulton 2010a), it remains to be seen how corpus resources co-exist with online services like Google and online dictionaries and how the learners' search habits behave in both contexts.

Using the same research methodology in Pérez-Paredes et al. (2011), this paper explores the initial stage in the process of learners' corpus consultation outlined in Sinclair (2003), so as to gain a better understanding of the direct, hands-on applications of language corpora to FLT contexts by tracking learner interaction with the learning resources. The aim of our research is to analyse learners' search behaviour when using the British National Corpus (BNC) and/or other web sites in an online learning environment, i.e. a physical classroom with open access to web services, from an evidence-based qualitative perspective. Section 2 reviews previous research into the hands-on uses of corpus-based language materials, in particular learners' corpus query and consultation. Section 3 deals with the methodological aspects of our research, including our operationalization of the concepts of corpus/online services query and the tracking of the learners' searches. In Section 4 we present the results of our research, in particular the sequential search patterns observed during the *Search the corpus* activity, the learners' activity completion and performance, as well as the observed search types and criteria. In the final section of this paper we offer a discussion of the learners' search patterns and search criteria, as well as of the different query types in the light of the use of DDL in tertiary education contexts.

2. The hands-on approach to corpus-based language learning

Direct applications of corpora in language teaching aim at bringing corpora closer to the classroom. Römer (2008) argues that, within this approach, corpora can be exploited either by the teacher, thus acting as a mediator between the corpus and the learners, or by the students themselves. The latter, which Gabrielatos (2005: 11) terms "the hard version", has grown in popularity over the last few years on the grounds that making students operate with corpora and corpus software, i.e. data-driven learning (DDL), can be beneficial in diverse ways. Researchers such as Chambers & O'Sullivan (2004), Mauranen (2004), Braun (2007), O'Sullivan (2007), Boulton (2008a, 2008b, 2011a), and Frankenberg-García (2010) have all established a link between DDL and relevant aspects in current pedagogy such as learner-centred teaching, autonomy, discovery learning and lifelong learning. Bernardini

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

(2000a) states that hands-on corpus exploitation can favour better retention and recall, as learning through data analysis involves a high degree of task involvement.

Despite the interest in DDL, very few studies so far have considered the actual, observable interaction between the learner and corpus-based tools. Flowerdew (1996: 112, in Kennedy & Miceli 2001) draws attention to "the paucity of critical perspectives in a perhaps over enthusiastic concordancing literature". Similarly, Kennedy & Miceli (2001: 80, emphasis in the original) stress that despite the amount of research conducted on *what* can be done with corpora in the foreign language learning classroom, there is still "relatively little in the literature on *how* students actually do this, and especially on how they fare on their own". Adopting the critical perspective advocated by Flowerdew (1996) and focusing on human-computer interaction will provide valuable insights into how students cope with concordancers or similar resources at hand during corpus consultation.

Leaving aside the use of questionnaires and other forms of indirect observation, two main tracking procedures have been employed to document corpus exploitation through direct observation: manual logs collected by the students (Ma 1994, Chambers & O'Sullivan 2004, Frankenberg-Garcia 2005, O'Sullivan & Chambers 2006, Varley 2009) and computer-generated logs (Cobb 1997, Johns 1997, Gaskell & Cobb 2004, Chan & Liou 2005, Hafner & Candlin 2007, Yoon 2008, Pérez-Paredes et al. 2011). These logs document the resources consulted, the search words employed, the results of corpus consultation, and the time spent on the activities, among other aspects.[1] Direct observation methods can provide more detailed information of all the steps involved in the process of corpus consultation than indirect observation methods such as interviews and post-task questionnaires, and computer-generated logs in particular present important advantages over manual logs, as these cannot guarantee the truthfulness of the data gathered. In particular, Pérez-Paredes et al. (2011) discussed the use of logs to understand learners' actual use of corpus-based resources in terms of the number of events or computer actions performed by each individual, the total number of different web services used, the number of activities completed as well as the number of searches performed on the BNC. In the following subsections, we offer an account of direct applications of corpora in language teaching that focuses on the process of querying the corpus, the challenges it poses to the learners and, finally, the suggestions to overcome such problems.

2.1 Querying the corpus

The use of corpora for research purposes, either linguistic or applied, as defined by Sinclair (1991) and Sinclair (2003: xiv–xv), comprises the observation of lexical patterning from an "unbiased selection of lines from the whole concordance".

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

After a refined search the corpus user is supposed to find the "main lines of the patterning of the node word or phrase" and then "use all the knowledge and intuition about the language that he or she has available, and […] move […] to some generalisations that group the data into sets which seem to show identical or sufficiently similar patterning […]". The user is advised that an active role is to be adopted and that pattern discovery must be pursued.

Sinclair (2003) summarized the seven procedural steps which must be taken to accomplish the kind of corpus-based analysis outlined above. The first step, *Initiate*, involves observing the words to the left and to the right of the node to decide on the strongest pattern. In the next step, *Interpret*, corpus users need to form a hypothesis that links all the words in the pattern together. This hypothesis is tested in the next step, *Consolidate*, where users will examine their hypothesis moving away from the closest words to the node and will try to find patterns which may go beyond the phrase boundary. This is a new opportunity to re-examine the original hypothesis and it is very likely that the corpus user will need to go back to the *Interpret* step. The next step, *Report*, requires that the observer writes down a version of the hypothesis so that in the following step, *Recycle*, the user can concentrate on other patterns which may be found in the concordance lines. The final steps of the procedure, *Result* and *Repeat*, involve the formulation of all the hypotheses on the node which was used as a starting point and the application of these results to a new set of data from the corpus.

Students exploiting corpora hands-on have often been metaphorically equated to detectives (Johns 1997) or researchers (Kennedy & Miceli 2001, Mauranen 2004). In a DDL setting, the researcher/learner needs to locate the appropriate data source, the corpus, and employ the right data collection tool, usually a concordancer, to obtain relevant data such as concordance lines, frequency lists, lists of collocates, etc., upon which to apply a relevant data analysis procedure (vertical reading, frequency analysis, etc.) in order to identify patterns, draw conclusions and, in sum, answer a research question. Boulton's comprehensive analysis of empirical studies outlines those studies which involved either the "soft" (teacher-mediated) or the "hard" (student-driven) versions of DDL.[2] At the time of writing this paper, out of 76 studies, only 9 involved non-tertiary education students, which goes to show that hands-on corpus exploitation has been mostly confined to undergraduate or postgraduate students majoring in Language, Linguistics or Translation or in other disciplines (Engineering, Law, Architecture, Economics, etc.) where Language for Specific Purposes (LSP) was part of the curriculum.

As regards the types of corpora used in DDL studies, an analysis of the literature in the field suggests that large, general corpora such as the BNC (Bernardini 2000b, 2002; Boulton 2009, 2011b; Kaur & Hegelheimer 2005; Gilmore 2009; Liu & Jiang 2009), the COBUILD/Bank of English (Estling Vannestal & Lindquist

2007, Yoon 2008, Gilmore 2009), the Corpus del Español (Davies 2004), the ICE (Cheng et al. 2003), as well as smaller, specialized or self-compiled corpora (Ma 1994, Kennedy & Miceli 2001, Kennedy & Miceli 2010, Bernardini 2002, Bloch 2009, Lee & Swales 2006, O'Sullivan & Chambers 2006) enjoy a similar degree of popularity among applied corpus linguists.

The query interface is the means through which learners retrieve information from the corpus and, therefore, their success in obtaining relevant results will largely depend on its features, functionalities and ease of use. Multiple desktop and online tools for querying corpora are available, some being more appropriate for language learning than others. In some cases, researchers report on the use of software that was designed by the same team as the corpus, as is the case of SARA, a highly specialized software developed for the BNC and employed by Bernardini in her 2002 study; the COBUILD Concordancer and Collocation Sampler, used to search the Bank of English online and employed by Estling Vannestal & Lindquist (2007), Yoon (2008), and Gilmore (2009), among others; and the concordancing interface incorporated in the Corpus del Español (Davies 2004). Access to large corpora via third-party interfaces is also popular among DDL practitioners. Three cases in point are the BYU interface to the BNC devised by Mark Davies (Boulton 2011b, Boulton 2009, Liu & Jiang 2009), Tom Cobb's online suite of tools *Compleat Lexical Tutor*, which retrieves data from the BNC and the Brown corpus among others (Kaur & Hegelheimer 2005), and the interface developed at the University of Leeds' Centre for Translation Studies, which also offers access to the BNC, the Brown corpus and other corpora (Pérez-Paredes et al. 2011). A third option is using commercial corpus query tools such as Mike Scott's (2008) Wordsmith Tools (Bowker 1998, Bernardini 2002, Cheng et al. 2003, Lee & Swales 2006, O'Sullivan & Chambers 2006, Varley 2009), Tim Johns' (1986) MicroConcord (Aston 1997, Granath 2009), and the no-longer-available Longman Mini Concordancer employed by Ma (1994). A different option often reported in empirical studies is to develop one's own corpus query interface, usually web-based, so that its features are perfectly suited to the students' needs (Kennedy & Miceli 2001, Kennedy & Miceli 2010, Miceli & Kennedy 2002, Sun 2003, Breyer 2006, Kaszubski 2006). Another possibility is to make the most of the World Wide Web by using web-as-corpus search engines like WebCorp or by simply using regular search engines such as Google and Yahoo (Acar et al. 2011) in spite of the limitations of this approach such as the unedited nature of the data or the instability of the results.

Numerous studies to date report on the learners' experiences with DDL, but there is a certain overreliance on indirect observation mainly based on questionnaires and interviews (Ma 1994, Kennedy & Miceli 2001, Bernardini 2002, Cheng et al. 2003, Yoon & Hirvela 2004, Chambers 2005, Götz & Mukherjee 2006, Lee & Swales 2006) which reflect not so much what learners "do" with the corpus

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

but their attitudes towards the resources and the methodology, their evaluation of the activity, and their self-perceived difficulties. Several researchers (Chapelle & Mizuno 1989, Fischer 2007) have pointed out the risks of assuming that what students report they are doing or what we believe they are (or should be) doing reveals their actual behaviour. There can be, therefore, a gap between what we assume the students are doing and what they are actually doing, a gap that must be filled if we want to identify and cope with potential sources of difficulty linked to the direct exploitation of corpora for language teaching. Even though the importance of tracking learners' use of corpus-based resources has been repeatedly stressed (Johns 1997, Horst et al. 2005, Chambers 2007, Hafner & Candlin 2007), we often find that the actual search strings are not collected (Gaskell & Cobb 2004) or are collected but not included as part of the study (Frankenberg-García 2005, Hafner & Candlin 2007).

Despite the fact that studies reflecting actual queries are still few in number in comparison with the amount of existing research on DDL, an analysis of students' searches suggests that these are, in most cases, somewhat simple. When learners are confronted with the search interface, they normally limit themselves to introducing a search string consisting of one or two words. With a few exceptions (Ma 1994, Bowker 1998, Chambers 2005), advanced searches are usually neglected, i.e. the use of wildcards and POS tags is absent from their queries. On some occasions, the authors mention specific cases where students missed the chance to extract relevant results only because they failed to use the wildcard '*', even if it can be assumed that performing advanced searches had been part of their training. Chambers & O'Sullivan (2004) report on a student who could have easily found several examples of *s'oppos* à* if he had searched for the lemma *oppos** and not for the less successful term *opposer*. Kennedy & Miceli (2001) report on a case in which a student could have spotted the mobility of the adjective *estremo* had she conducted a search for *estrem** instead of using the more restrictive feminine form *estrema.* Bernardini (2002) describes a similar situation regarding the use of variants by her students. Ma (1994) reports that, even though his students did take advantage of the '*' character and "tended to overuse the wildcard by putting it in every search" (Ma 1994: 11), they almost always placed it at the end of the search string and hardly ever in the middle or at the beginning. Moreover, the wildcard '?' was completely ignored, although it had been explained to those learners.

2.2 Challenges of corpus consultation and recommendations

Despite the fact that "the overwhelming majority of studies [on learning outcomes from corpus consultation] produce encouraging results, even if they are not always statistically significant on all research questions" (Boulton 2010b: 143), and

that students' evaluation of DDL is often positive (Bernardini 2000b, Chambers & O'Sullivan 2004, Lee & Swales 2006, Lavid 2007, Farr 2008), corpus consultation poses several challenges to the learners, according to the feedback they give in post-task questionnaires or interviews and to the search behaviour unveiled through direct observation methodologies. Ädel (2010) offers an interesting account of the challenges of using corpora in writing, including the availability of corpora, the interpretation of data, the danger of drowning in data, and the decontextualized nature of corpus data, among other things.

Two main sources of difficulty stand out in the literature: obtaining relevant results, i.e. querying the corpus, and analysing the data. These processes are so closely interwoven that the inability to cope with them can lead to a snowball effect that spreads over the subsequent stages of corpus consultation, resulting in a feeling of general dissatisfaction with DDL. Formulating corpus queries tends to pose problems for students (Ma 1994, Bernardini 2000b, Kennedy & Miceli 2001, Miceli & Kennedy 2002, Cheng et al. 2003, Sun 2003, O'Sullivan & Chambers 2006, Estling Vannestal & Lindquist 2007, Hafner & Candlin 2007) and is usually linked to aspects such as insufficient training and difficulty in understanding the query syntax or in grasping the functionalities of the search interface. Bernardini (2000b) stresses that learners sometimes use either very rare or very common words in their queries, which causes the program to retrieve a small number of concordance lines or too many results, respectively. Hafner & Candlin (2007) found that the concordancer was being used as a search engine, as they detected searches for full documents while their students were involved in legal research.

The second main obstacle identified in the literature, the analysis of results, is intricately linked to the difficulty in querying the corpus. This problem inevitably hinders the progression through the first two stages of corpus-based research outlined by Sinclair (2003): the *Initiate* and *Interpret* stages. Analysing concordancer output is often considered a very complex task, according to the students' answers to questionnaires and interviews with their instructors (Ma 1994, Bowker 1998, Cheng et al. 2003, Kennedy & Miceli 2001, Miceli & Kennedy 2002, Sun 2003, Yoon & Hirvela 2004, O'Sullivan & Chambers 2006, Estling Vannestal & Lindquist 2007, Lavid 2007, Johns et al. 2008, Varley 2009, Liu & Jiang 2009, Boulton 2009). This problem tends to be associated with the amount of data presented (an evident consequence of their insufficient query skills), the nature of corpus data, the display of results, the students' level of L2 proficiency and the difficulty in coping with unknown words, as well as their inadequate reasoning and observation skills. Students often express feelings of frustration (Lavid 2007) or feel drowned in data (Ädel 2010). Words such as "overwhelming" and "overwhelmed" (Johns et al. 2008, Liu & Jiang 2009, Kennedy & Miceli 2010) are frequently found in their descriptions of hands-on corpus exploitation. Pérez-Paredes et al. (2011) tracked

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

the behaviour of learners while using corpora and found that the learners in the guided-consultation condition searched the BNC almost three times more than the individuals in the non-guided consultation group. This finding was statistically significant and points to the relevance of appropriate learning environments in DDL.

The most frequent recommendation in overcoming the kind of difficulties cited above is to provide learners with extensive and detailed training on the nature of corpus data and on the functionalities of corpus-based resources. From the analysis of results of a study involving corpus consultation after a training period, Kennedy & Miceli (2001) concluded that students were not "adequately equipped" (Kennedy & Miceli 2001: 81) as corpus researchers and thus stressed that training should follow an appropriate sequence (Miceli & Kennedy 2002). Along the same lines, numerous researchers argue that hands-on work with corpora should follow a period of teacher demonstration (Cheng et al. 2003, Yoon & Hirvela 2004, Chambers 2005, O'Sullivan & Chambers 2006, Estling Vannestal & Lindquist 2007). Liu & Jiang (2009) suggest presenting students with modelling, group work and deductive activities before putting them to work with corpora individually on inductive or exploratory tasks, while Ädel (2010: 46) states that "teacher-guided settings, clearly defined tasks and smarter tools would all contribute to making it easier to find ways out of the maze". Several scholars recommend paying special attention to the need for interfaces which are "appealing and easy to use" (Römer 2008: 123) and tailored to the needs of the students, as it is often the case that they operate with tools which have been designed for research purposes (Breyer 2006). Similarly, pedagogic mediation (Pérez-Paredes & Alcaraz 2009, Pérez-Paredes 2010) has been suggested to play a role in the design of a more user-friendly corpus consultation setting. Wible et al. (2002), Breyer (2006), Kaszubski (2006), Kennedy & Miceli (2010), and Bloch (2009), just to cite a few, adopted this approach and devised search interfaces that meet their students' needs and abilities. This paper sets out to contribute to the previous body of research in the pedagogic uses of hands-on concordancing by examining the initial stages of learners' corpus consultation, and in particular, learners' searches on the BNC and other possible online web sites.

3. Research questions and methodology

3.1 Research rationale

This study examines the searches of university EFL learners during their use of corpus-based resources and other online services while completing focus-on-form activities. Our research has been designed (i) to gain further insight into the

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

search behaviour and patterns of learners in DDL, and, specifically, (ii) to explore the initial stages of the process of corpus consultation outlined in Sinclair (2003).

3.2 Setting

Our data was collected from one EFL class of second-year students studying for an English degree at a medium-sized Spanish university. All the students involved in this research were enrolled in a compulsory EFL subject where language instruction was provided to gain a B2-to-C1 level of English. The classes met three times a week and lasted one hour each, over a course of thirty weeks during the academic year. The experiment was conducted over a week toward the end of the semester, taking two 60-minute sessions in total. In class session 1, the students were exposed to a text that included contextualized uses of *it*-cleft sentences. During class session 2, students had to complete some focus-on-form activities which involved English cleft sentences. To do this, they were instructed to use the BNC and other online services of their own choice which they deemed necessary or useful. A set of explicit guidelines introduced students to working with corpora by explaining to them what a corpus is and how it can be used in class. These electronic guidelines included a simple overview of the uses of corpora, a description of the query interface of the BNC used in this experiment, and simple screen-shot instructions on how to search a corpus. These guidelines can be found in Appendix A. The students were asked to read the guidelines carefully before completing the tasks and to consult them as much as necessary. In the *Search the corpus* activity, students were explicitly told about specific search tips, wildcards and tags. These specific guidelines were intentionally embedded in the activity. Appendix B shows the *Search the corpus* activity.

3.3 Participants

The original pool of informants included a total of 43 EFL university students. Of those learners, only 24 (9 male and 15 female) with an average age of 20.61 years (SD = 1.1) participated in the entire experiment. All students had achieved a B2 level of English, 6.7 (out of 10) being the average entry level mark (SD = 1.03) for this group. The majority of them were native Spanish speakers, although two students had a different mother tongue. They had all used the English Corpus Concordance website *Compleat Lexical Tutor* the year before and were familiar with the notion of corpus and concordance, and, in particular, with the interface of this resource, as well as with terms such as 'keywords', BNC and other English corpora, 'sorting' and 'collocation'.[3] This contact with corpora took place during three tutor-led sessions as part of a course which dealt with the basics of linguistics

where the students examined the behaviour of selected words as part of their training. The *Compleat Lexical Tutor* includes a description of the origin and the components of corpora such as BNC, Brown and many others.

3.4 Task and activities

Following Granath (2009), we decided that the design of the task should be integrated as much as possible into the students' regular course. The participants completed a focus-on-form task dealing with the use of English cleft sentences based on a short story from their coursebook, *Objective CAE* (O'Dell & Broadhead 2002). As in Pérez-Paredes et al. (2011), the task introduced the students to the form and uses of *it*-clefts and presented the following structure: 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, the students were offered examples extracted from the BNC; in the second activity, they were asked to use the University of Leeds gateway to the BNC to find relevant examples of *it*-clefts and infer patterns.[4] This involves Sinclair's (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.

3.5 Tracking learners' searches and corpus/online services queries categorization

Student-computer interactions were tracked by means of Fiddler logs.[5] Fiddler acts as a proxy intercepting all the communications between the user and the Internet, assisting researchers in linking individuals with their queries (Fischer 2007, Hafner & Candlin 2007). Fiddler tracks web-browser actions, captures all the web pages visited and all the information typed in. Used in combination with Moodle activity reports, the logs generated by Fiddler allowed us to track the whole task resolution process, which facilitated the extraction of behaviour patterns. Apart from accessing and compiling student behaviour unobtrusively, this data collection procedure overcomes the difficulty of linking each student with their queries, an element which is often absent from similar studies, as emphasized by Hafner & Candlin (2007: 304).

In the context of this research, a search action is any query performed in any type of online resource, whether corpus-based (the BNC or a different corpus) or not (Google, WordReference, etc.). For the purposes of this study, the searches performed by the learners were analysed from three perspectives. First, the

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

sequential search pattern followed by the learner, that is, the order and number of BNC searches and/or other online services. The emerging patterns were correlated with the learners' activity performance. Second, a learner search analysis gave us the total number and types of searches. Analysing the number and type of query performed in each resource helped us gain insight into the motivation behind each search action. Finally, the search criteria used by the learners were examined, that is, the exact strings of words that were submitted to the BNC, or other services, during the completion of the activities, in particular during the *Search the corpus* activity.

Our taxonomy of queries is shown in Figure 1. For the development of our taxonomy of queries, we drew from Broder's (2002) classification of web searches into "navigational", "informational" and "transactional", to which we added a fourth category ("undetermined searches"), as well as three subcategories under informational searches ("explanation", "pattern" and "undefined").

Broder (2002) defines navigational queries as those whose motivation is to locate the website of a specific web resource whose existence is known, or assumed, by the user, as could be the case of the search for the website of a particular dictionary, e.g. the string "WordReference". In the case of informational searches, on the contrary, no further interaction other than reading is implied, since the drive

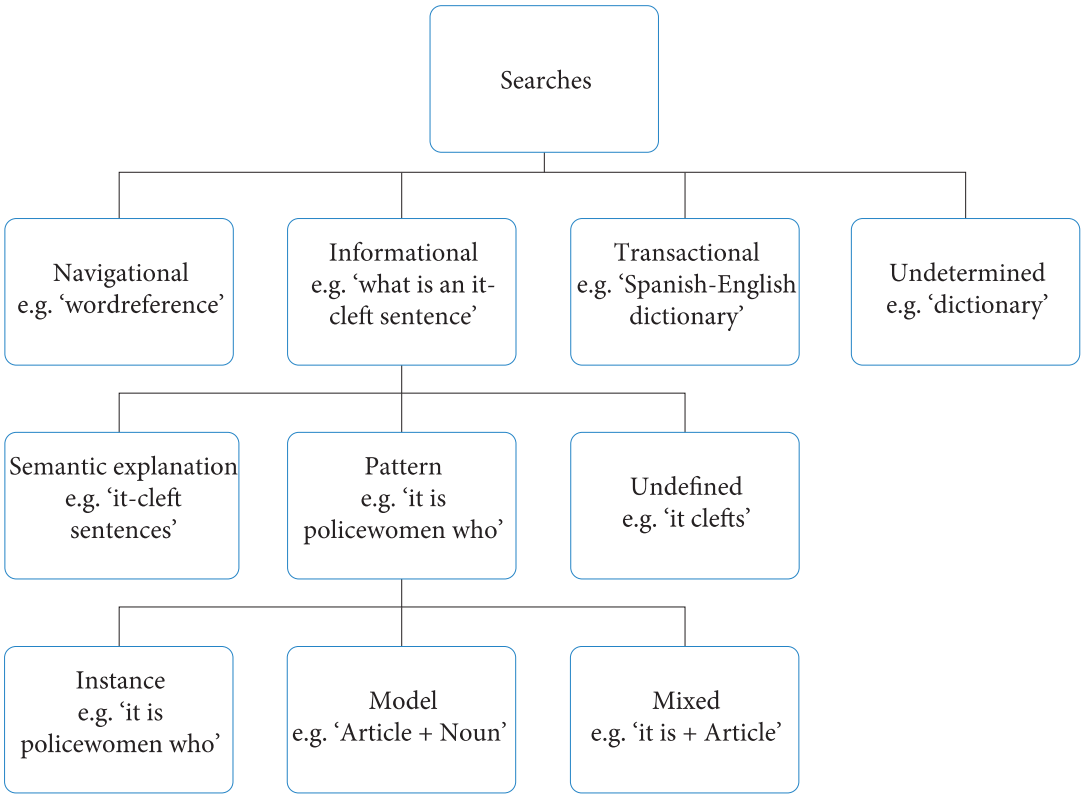


Figure 1. Taxonomy of queries

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

behind this type of search is to find information that is assumed to be present on the Web, for instance "what is an it-cleft sentence". A transactional query aims at finding a site "where further interaction will happen" (Broder 2002: 6). This means that the user needs to find a website where a specific action can be performed, such as checking the meaning of a word. A search of the transactional type would be, for instance, that in which the user enters the string "Spanish-English dictionary" into a search engine. Finally, undetermined queries are those which cannot be associated to any other type of search due to the lack of enough contextual information and its associated ambiguity. For example, the search criteria could be provided with the aim of retrieving the explanation of the term "dictionary" (informational), with the aim of retrieving a web resource in particular (Dictionary.com) or with the aim of finding a list of dictionaries from which to select the most appropriate. Note that the search criteria are ambiguous enough so as to not to determine clearly what the intention of the user is.

For the purposes of our analysis of queries in corpus-based and online resources, we refined the *informational* category by adding further subcategories. Thus, *semantic explanation* searches aim at locating web resources that provide an explanation of the concept specified in the search string. For example, a search for "*it*-cleft sentences" or "emphatic effects" will have as a result a list of web pages in which the definition of such a term is provided to the reader. *Pattern* searches, in contrast, are launched with the intention of retrieving sites matching the patterns provided in the given search criteria. The search "it is policewomen who" aims to retrieve all the web pages which contain the search pattern provided. These kinds of searches can be enclosed in inverted commas in order to let the search engine know that the search pattern has to match the text found in the web page exactly, or can alternatively be entered without inverted commas in case the user wants the search engine to find results which match the search string only partially. This would be the case of web pages containing just "any" of the words provided in the search criteria. Such searches can be subdivided into *instance* searches, *model* searches and *mixed* searches. Instance searches only provide word forms in the search criteria (e.g. "it is policewomen who"), against model searches, in which the pattern contains a model representing categories of elements, as would be the case of a corpus search containing the tags "Article + Noun". In the case of mixed searches, the query combines instance searches and model searches. This implies that the search criteria contain not only actual word forms but also word classes. The string "it is + Noun" is a clear example of a mixed search. Finally, those searches that were carried out on an inappropriate resource (e.g. searching for "*it*-cleft sentences" on the BNC as if it was a search engine), or those in which the search terms are too vague to fall neatly into any of the above categories, were labelled as *undefined* searches (e.g. "*it* clefts" could refer to an *instance* search trying to retrieve all the resources containing "it"

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

followed by "cleft" or to a *semantic explanation* search trying to retrieve resources explaining cleft sentences).

4. Results

The learners' activity logs were processed and analysed to account for the students' search behaviour during corpus consultation. The analysis focused first on the sequential search patterns elicited by the activity logs, which enabled us to extract the exact steps that each student took during the completion of the activities. Second, we examined the rate of activity completion, in addition to students' performance in relation to the number of search engine or corpus searches carried out. We then examined the type and number of search sessions that participants launched to solve the activity, *Search the corpus*. Finally, an analysis of the search criteria employed by the students, that is, the exact search strings, was carried out.

4.1 Sequential search patterns during the *Search the corpus* activity

The number of BNC search sessions during the completion of the *Search the corpus* activity ranged from 1 to 5. A search session is defined as the search or searches performed by an individual before exiting a given web service, i.e. BNC, Google, etc. in order to do something else. Ten students, 41.7%, searched the BNC on one single session, eleven learners, 45.8%, searched the BNC during three or more different query sessions before completing the activity, while three, 12.5%, needed two. Taking Activity 2 (*Search the corpus*) as a pivotal reference, we have found that the learners' search behaviour followed three different sequential patterns.

Pattern A: BNC search > Activity completion
All of the learners who queried the BNC during only one session answered and submitted the activity straight away. This is a unidirectional pattern where one and only one BNC search session is enough for the *Initiate* and *Interpret* stages in Sinclair (2003). Pattern A was followed by 10 of the 24 students, that is, 41.7% of the informants.

Pattern B: BNC search > Google or similar > BNC search > Activity completion
The few students who queried the BNC during two query sessions used Google invariably after their first query: two for navigational purposes and one for informational purposes. The navigational searches were motivated by the need to obtain the URL of WordReference.com and, accordingly, query this online dictionary service. For all the students querying the BNC during two search sessions, this

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

is a reiterative pattern where the first BNC query is followed regularly by queries on other services and subsequently followed by a new search on the BNC and the completion of the activity. Pattern B was followed by 3 of the 24 students, that is, 12.5 % of the informants.

Pattern C: BNC search > BNC guidelines |BNC search | Google or similar > BNC search > Activity completion

This search behaviour pattern represents the students who used the BNC resource at least three or more times. Its description above contains a vertical bar "|" indicating the logical operator "OR" as well as underlined text indicating that the underlined step is compulsory, i.e. all the students under pattern C did search the BNC in the stage following the first BNC search and preceding the second BNC search prior to activity completion. This pattern, therefore, represents those students who first turn to the BNC, then either read the BNC guidelines or use Google and search the BNC and, finally, visit the BNC once more before completing the activity. What all the students following this pattern have in common is that, prior to completing the activity, they have searched the BNC at least three times, having used other resources between the first and the last BNC search. Not all of the students who queried the BNC during three sessions or more resorted to other online services. However, most of them checked in with the electronic BNC search guidelines provided by the researchers. When learners query the BNC on at least three occasions, the pattern is clearly reiterative, coming back to a new BNC search after checking in with resources other than the BNC interface used in our experiment. Those students who queried the BNC during 4 (3 students, that is 12.5%) and 5 (3 students, that is 12.5%) sessions invariably used a reiterative pattern which involved an informational use of Google, going back to the BNC search guidelines or, in 5 of the 6 cases, both. Pattern C was followed by 11 of the 24 students, that is, 45.8% of the informants.

4.2 Activity completion and performance

All three activities (*Observe*, *Search the corpus* and *Rewrite*) required that the learners submit their answers to their instructor through an online application. All 24 participants completed Activity 1 (*Observe*), Activity 2 (*Search the corpus*) and Activity 3 (*Rewrite*) (see Appendix B). As for the number of attempts per activity, we found that participants did not normally need more than one try. Five individuals needed two attempts to complete Activity 1, three students completed Activity 2 after two attempts, and five students needed 2 or 3 attempts in Activity 3. Considering the learners' performance assessment on the *Search the corpus* activity, which was the one that specifically asked the participants to search the BNC, those individuals who followed Pattern A performed below the standard of those who followed Pattern C.

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

Table 1. Students' performance/Search patterns

Search the corpus Performance	A	B	C
0–2	62.5%	20%	32.8%
3	12.5%	-	-
4–5	25%	80%	67.2%

Table 1 shows the grades and the students' search patterns. Performance is rated from 0 to 5, 0 being the lowest mark possible and 5 the top score. In Table 1, students were divided into three different groups according to their performance: those whose performance was below the passing mark (0, 1 or 2 marks), those who obtained 3 marks and those whose performance was good or very good (4 or 5 marks).

In fact, learners following Patterns B and C outperformed learners following Pattern A, which indicates that, for this group of students, reliance on online services and multiple BNC search sessions worked well.

4.3 Search types

Table 2 presents an overview of the total number and type of searches conducted by the participants. Although it was only in Activity 2 (*Search the corpus*) that students were explicitly required to search the BNC, they were free to use whatever online resources they considered necessary during the completion of the other two activities. Activity 2 is, for that reason, the one in which searching was more intense (164 actions), followed by Activity 1 (*Observe*), where students performed 31 searches, and Activity 3 (*Rewrite*), with only 12 search actions. Regarding the type of online resource consulted, participants resorted to the language corpora available from the interface they were instructed to use (BNC, *British News Corpus*, *Internet Creative Commons Corpus*, *Reuters Corpus*), search engines (Google), and dictionaries (WordReference). These online services were mostly used to solve Activities 1 and 3, while the use of corpora was mainly restricted to Activity 2, which demanded students' extraction of *it*-clefts from a corpus.

As illustrated in Table 2, pattern instance searches are the largest in number across all activities and in all resources. The vast majority of searches carried out in the available corpora fall within this category (121 out of 142), as do some of the searches conducted on Google when it was, consciously or unconsciously, used as a corpus (7 out of 37). Informational searches for explanation, model patterns and undefined subtypes occur less frequently in search engines, dictionaries and corpora alike. We then find informational pattern-mixed searches, which combine an exact instance with a model or concept, e.g. "it was + noun", "it was + emphasizing", and which were only performed in the BNC. The last subtype of

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

Table 2. Type and number of searches per activity

Activity	Resource and total number of searches	Searches conducted		
		Type	Subtype (Sub-subtype)	Number
1. Observe	BNC (5)	Informational	*Pattern (Instance)*	5
	Google (14)	Informational	*Pattern (Instance)*	4
			Undefined	1
		Navigational	-	4
		Transactional	-	1
		Undetermined	-	4
	WordReference (12)	Informational	*Explanation*	12
		Total 31		
2. Search the corpus	BNC (128)	Informational	*Pattern (Instance)*	109
			Pattern (Mixed)	7
			Explanation	1
			Undefined	11
	British News Corpus (5)	Informational	*Pattern (Instance)*	5
	Internet (Creative Commons) Corpus (3)	Informational	*Pattern (Instance)*	1
		Informational	*Explanation*	2
	Reuters Corpus (1)	Informational	*Pattern (Instance)*	1
	Google (13)	Informational	*Pattern (Instance)*	1
		Informational	*Explanation*	5
		Informational	*Undefined*	4
		Navigational	-	3
	Google Books (1)	Informational	*Undefined*	1
	WordReference (12)	Informational	*Explanation*	12
	MSN (1)	Navigational	-	1
		Total 164		
3. Rewrite	Google (10)	Informational	*Pattern (Instance)*	2
		Informational	*Explanation*	2
		Informational	*Undefined*	4
		Navigational	-	2
	WordReference (2)	Informational	*Explanation*	2
		Total 12		
			Total number of searches across all 3 tasks: 207	

informational searches is undefined searches. This subcategory comprises queries that were ill-formed, e.g. a search for "*it*-clefts" in the BNC, or very vague, e.g. a search for "affirmative" in Google.

In comparison to informational searches, navigational, transactional and undetermined searches are considerably less frequent. Out of the 207 requests

registered, only 10 were of the navigational type, those aimed at finding a resource (e.g. "English-Spanish dictionary"), and all of them took place in the search engines Google and MSN. Only one student carried out a transactional search, "diccionario inglés-español" (the Spanish for "English-Spanish dictionary"), which aimed at locating a resource where he could carry out a specific action, in this case, find a translation or definition of a term.

4.4 Search criteria

Table 3 provides a record of the search strings entered more than once in any of the three types of resources under consideration in each activity. A first look at the table reveals that the searches which were performed more than once were generally conducted on the BNC. This resource was most frequently employed to locate sample strings containing the exact words typed in, although it was sometimes misused to find definitions of concepts, as seems to be the case of search strings such as "*it*-clefts", which was the grammar point dealt with in all three activities, and "hint", which was one of the words that appeared in the rubric of Activity 2. Overall, students did not use any wildcards or tags when searching the BNC and the most frequent search strings ("it is", "it is the", "it was", "it was only") were generally rather too short and vague to locate relevant results containing examples of *it*-clefts. The most recurrent queries registered in Google were carried out in Activity 1 (*Observe*) and represent searches of the informational pattern instance subtype, that is, similar to those conducted in the BNC; the informational explanation subtype, aimed at finding an explanation to concepts such as "emphatic effects"; and the navigational type, aimed at locating a specific resource (WordReference). With regard to WordReference, it was used in Activity 1 (*Observe*) only to find term definitions or English equivalents.

If we now compare the search strings of the informational pattern instance subtype entered in the BNC to those of the same kind entered in Google, we observe that they do not differ substantially, as students tend to replicate the examples given in the activity ("it is policewomen who") or enter very similar strings to the one provided ("it is the man who"). In neither case did students use wildcards or tags.

5. Discussion

The transfer of corpus linguistics methods to the language classroom (Pérez-Paredes 2010) has promoted renewed interest in the analytical skills of learners (Sinclair 1991, Sinclair 2003), at least on a purely theoretical basis. Unfortunately, the fact is that very few studies have examined the actual interaction of end users

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

Table 3. Search criteria (excluding hapax legomena)

Search Resource	Activity	Search string	Number of repeated searches
BNC	1. Observe	it is	2
BNC	2. Search the corpus	Hint	4
BNC	2. Search the corpus	if-clefts	2
BNC	2. Search the corpus	It	3
BNC	2. Search the corpus	it clefts	5
BNC	2. Search the corpus	it is	23
BNC	2. Search the corpus	it is human who	2
BNC	2. Search the corpus	it is human who think	2
BNC	2. Search the corpus	it is noun	2
BNC	2. Search the corpus	It is the	10
BNC	2. Search the corpus	it is the man who	5
BNC	2. Search the corpus	it is through	3
BNC	2. Search the corpus	it was	8
BNC	2. Search the corpus	it was not	2
BNC	2. Search the corpus	it was only	8
BNC	2. Search the corpus	it'd be that	2
BNC	2. Search the corpus	it'd be who	2
British News Corpus	2. Search the corpus	it is	2
Google	1. Observe	It is policewoman who	2
Google	1. Observe	It is policewomen who	2
Google	1. Observe	a collection of English corpora	2
Google	1. Observe	WordReference	3
Google	2. Search the corpus	emphatic effects	2
Google	2. Search the corpus	it cleft	2
Google	3. Rewrite	it was emphasising	2
WordReference	1. Observe	Deal	2
WordReference	1. Observe	Resaltar	2

with corpus-based materials, which may prevent experts from gaining more detailed knowledge of how corpora can be effectively integrated into language learning. Thus, documenting the learners' search behaviour during corpus consultation can only help us gain insight into the search habits of corpus users, which, in turn, can shed light on what goes on during the different stages of corpus analysis outlined by Sinclair (2003) and can help teachers in the design of appropriate training and guidance activities.

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

5.1 Search patterns, search criteria and activity performance

Two predominant patterns of search behaviour emerge from the analysis of the learners' logs. Pattern A was followed by 41,7% of our informants. For these students, one corpus query session was enough to complete the *Search the corpus* activity and there was no need to resort to other resources such as Google or online dictionaries. Pattern C was followed by 45.8% of our informants. For these students, at least three corpus query sessions were needed before completing the activity. These learners resorted to the corpus consultation guidelines provided and/or to online web services. In every case, new BNC query sessions were launched before completing the activity. The number of learners that either relied on a strategy based on one single query session with no further consultation of other services or consultation guidelines, or one based on a more extensive use of corpora and combined resources, including services like Google and/or the consultation of the guidelines provided by the instructors, was almost identical, which suggests that, at least for this class, the first stage of corpus consultation, the *Initiate* stage, was not as undemanding as the literature seems to suggest given the little attention it has received.

Despite the growing evidence on the existence of a new type of learner, that is a sophisticated user of information-related web services (Schroeder et al. 2010, Palfrey & Gasser 2008), our results suggest that, irrespective of the search pattern, the use of online services such as Google is not so pervasive when students face work with corpora, as shown by the fact that learners following Pattern A did not feel the need to do any kind of research outside the corpus query interface or the need to re-check the corpus consultation guidelines provided. These results differ from those in Pérez-Paredes et al. (2011) where, over a longer period of corpus interaction involving more activities, all students used the web in some way to locate information. It would appear that Pattern A was followed by learners who had made an extremely accurate choice as to the word, or string of words, they wanted to use in their corpus query. However, the logs reveal that the choice was ineffective, as the search criteria show (Table 3). Students following Pattern C behaved more like researchers as they exited the corpus query interface, looked for information which could help them with their search, and always came back to the corpus to formulate a new query before completing the activity. Given the fact that all of these learners had already been exposed to some focus on forms on *it*-clefts during the task (*Observe*), almost half of them queried the corpus relying, for the most part, on the language samples provided in the first (*Observe*) and in the second (*Search the corpus*) activities.

When querying the BNC, the learners that followed Pattern C showed search behaviour similar to that of the learners following Pattern A. They narrowed down

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

their search criteria once they realized that the exact string which was being used was not present in the corpus or that it was the only hit returned. So the question now is whether Pattern C followers actually benefited from their strategy, that is, whether resorting to resources outside the corpus was more efficient at all. This is a hard question to answer, especially if we do not try to respond from different angles. Let us examine in more detail two of the learners that performed mixed searches which combined instance, "it is", with meta-information, "noun", which obviously returned no hits. The learners then checked with the corpus consultation guidelines and/or with Google, mainly for explanation. One of these learners searched the BNC in three different search sessions using the search criteria illustrated in Figure 2.

It took four search sessions, two on the BNC and two on Google, to realize that "it is" is an essential component of the language pattern at stake, which, interestingly, was part of the very first search this learner performed. Had the learner read the guidelines more carefully, he would have been 100% successful. A classmate followed a similar path and reached a similar conclusion, illustrated in Figure 3.

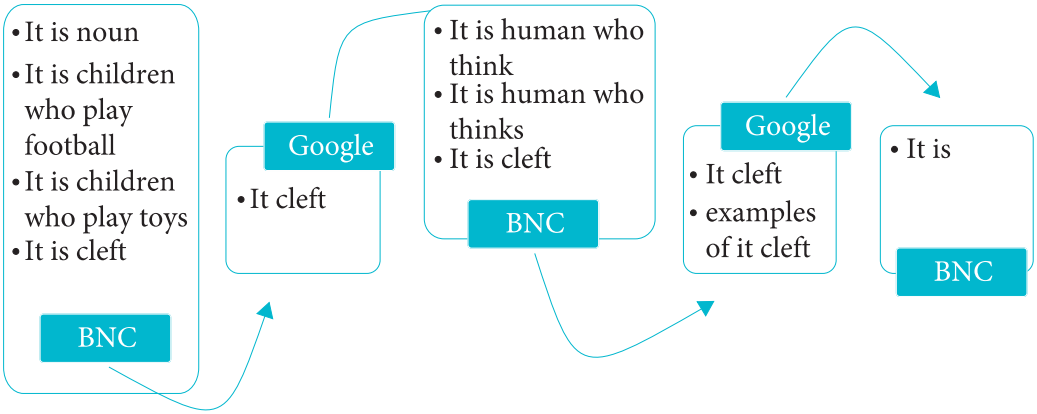


Figure 2. A sample of mixed searches

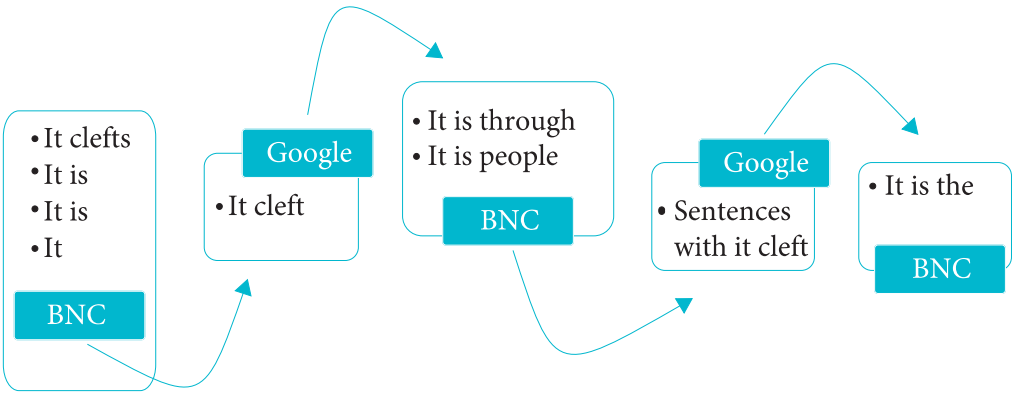


Figure 3. Another sample of mixed searches

On a formal basis the learner is carrying out informational pattern instance searches on the BNC, but at the end of the process the student ends up with a simple search string that will return 6,199 hits which will, inevitably, produce lots of noise totally unrelated to *it*-clefts. In this respect, our findings are in line with those of Kennedy & Miceli (2001), Chambers & O'Sullivan (2004) and Pérez-Paredes et al. (2011). All of our informants following Patterns A and C performed unsophisticated searches which did not make use of the wildcards or POS tags in the guidelines provided. During the activity, learners were told about the regex "it" [] {1,4} "who" which would have saved them lots of time and unfruitful searching. In general terms, all of our informants showed a common behaviour pattern during their interaction with the corpus. They usually started their query with one of the language samples provided in the first activity (*Observe*), i.e. "It is policewomen who", and went on to reduce the number of words, i.e. "it is policewomen", and then "it is". Others queried the corpus only once, e.g. "it was", while another started with an explanation search, "it clefts", and then moved on to "policewomen" and "it was only". This process shows that learners are aware of the need to refine their searches, although it similarly reflects their failure to grasp the potential of corpora for language learning as their search is mostly restricted to informational pattern instances, which excludes the combinatory power of language to present both syntagmatic and paradigmatic variation. The fact that the string "it is policewomen" was actually taken from the BNC may have encouraged learners to believe that a corpus is likely to contain almost any kind of language realization imaginable.

The above suggests that, despite both the exposure to explicit guidelines on how to use language corpora and using the *Compleat Lexical Tutor* concordancer the year before, the students that took part in our study might not be ready to perform corpus searches autonomously, as their query skills seem rather limited in many cases. Not being specifically trained can obviously be problematic as it makes it difficult for learners to find the balance between very broad and very narrow or restrictive searches. It seems the multiple and diverse corpus interfaces available, however, can play havoc with any attempt to teach general corpus skills. In the case of broad searches, in particular when the search string consists of one or two words or when the terms are very common, a large number of the concordance lines retrieved will be unlikely to represent the required language point or feature, as the results may be "contaminated" by irrelevant examples even if they match the search terms, and will accordingly not produce a satisfactory number of examples, causing the extraction of patterns to be inaccurate or even impossible. This could, in turn, make students feel they are swamped by data (Ädel 2010) and could possibly make them think that DDL is too laborious and time-consuming. Our data show that while most of the learner searches contained between three and four words, the lack of refinement and sophistication actually forced them

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

to use 2-word queries such as "it is" which would not be of any help in trying to understand the uses of *it*-clefts. This was the case in the non-guided experimental condition in Pérez-Paredes et al. (2011).

As for performance differences between learners following different search patterns, those following Pattern C achieved better results, with 67.2% obtaining a 4 or a 5 grade, than their peers following Pattern A, where only 25% achieved this standard and 62.5% failed, compared with 32.8% following Pattern C. While these results may suggest that learners following Pattern C were better prepared to make the most of corpus-based language learning, the number of individuals taking part in the study and, in particular, the number of learners in each of the groups, 10 vs. 11, prevents us from attributing significance to these findings. These results, however, confirm claims in the literature about the relationship between the role of corpora as input-focusing tools (Johansson 2009) and suggest that the stages prior to analysing concordance lines can be improved if the learners have access to familiar information services on the Web. This was corroborated by Pérez-Paredes et al. (2011) where learners in the guided-consultation experimental condition not only searched the BNC more intensively than those in the non-guided condition, but also used more complex strings of words. Despite the similarities in the simplicity of the search criteria of both groups, the performance of learners following Pattern C was more positively assessed than that of their classroom peers following Pattern A. Future research should examine in more detail whether these patterns emerge in larger populations and whether Pattern C followers systematically outperform learners that prefer search Pattern A.

5.2 Query types

Our results show that informational searches of the pattern instance subtype are the most frequent type of search across all three activities. They aim to locate exact instances of the information typed in and therefore represent the simplest and most prototypical kind of query that is usually performed on a corpus. Two examples are "it is policewomen who" and "it was only". Informational searches of the explanation subtype were carried out when a concept clarification or definition was needed and took place mainly in WordReference (24 out of 36) and less commonly in Google. "Emphatic effects" and "*it*-clefts" are examples of explanation searches registered in search engines, while requests such as "deal" and "resaltar" (the Spanish word for "highlight") were searched for in the online dictionary. This type of search was uncommon in the BNC and more frequent in Google during Activity 2 (*Search the corpus*), which suggests that students were well aware of the nature of the information resources being queried. This is supported by the fact that during Activity 1 (*Observe*), this type of search was more common than

the other informational subtypes. During Activity 3 (*Rewriting*), no BNC searches were performed and Google was the most frequent site used by the learners who resorted to web services. Navigational searches are scarce.

Possibly the most interesting finding which emerges from our results is that while learners do not use Google or other services to perform instance searches during Activity 2, that is, they do not query Google as if it was a corpus, they certainly tend to use the BNC search interface in the same way as if they were using Google. This finding corroborates Hafner & Candlin's (2007) conclusion regarding learners' use of corpora as search engines. The learners' approach to instance searches on the BNC ignores the complexities of POS tags and regexes and, as we can tell from Table 2, is based on the expectancy that the corpus interface will compensate for the lack of textual refinement or contextualization. It is interesting that far from having a problem with the very concept of corpora, what learners are missing here is a search interface which does what Google can do on an everyday basis. It appears that the instructors' hints provided in every activity (see Appendix B), together with the guidelines which they were supposed to read and use, were simply beyond their interest, which was probably more focused on task completion. While Google's invisible technology compensates users for the lack of precision and the lack of context, the kind of gateway that is offered to learners that use corpora such as the BNC ignores that modern learners have grown used to this back-door to finding relevant information. Inadvertently, when searching on the BNC, learners are querying a database without a full grasp of its structure. Google, on the contrary, is proud of one of its mottos: "let Google fill in the blanks". The number of undefined searches in our results is indicative of the learners' lack of accurate criteria and their struggle to come to terms with finding the right criteria to start using language corpora.

Our findings suggest that Palfrey & Gasser's (2008) digital natives are right now sitting in our classrooms consulting corpora which were devised over two decades ago, and who seem to have very little interest in building complex queries, even if they are told how to do so. Even wildcards easily interpreted by Google (i.e. "*") are not used by the students that took part in our research. Our direct, unobtrusive research methodology was intended to give a snapshot of what students actually do when interacting with corpora. Our findings corroborate previous concerns about the reliability of indirect methodologies (Chapelle & Mizuno 1989, Fischer 2007) and the lack of attested data regarding the early stages of corpus use in language classrooms. Boulton (2010a) has pointed out that some of the obstacles of DDL are more concerned with the implementation of DDL rather than with its nature, and this seems to be the case with our informants.

Existing corpus query interfaces have been designed with a research paradigm in mind (Breyer 2006, Pérez-Paredes 2010) which, as our data show, cannot be

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

easily transferred to the language classroom, even with advanced learners. The interpretation of language data (Ädel 2010) gets more difficult if the very first step in the process is not appropriately addressed by language experts and corpus linguists. On the other hand, our results confirm previous experiences with standard corpora where it was concluded that familiarity with corpus linguistic methods was key to implementing corpora successfully in the language classroom (Cheng et al. 2003, Yoon & Hirvela 2004, Chambers 2005, O'Sullivan & Chambers 2006, Estling Vannestal & Lindquist 2007). It seems sensible to think the representation of data itself (Barlow 2011) is one area where linguists will have to put extra effort if we want to maximize the opportunities for learner engagement with DDL.

6. Conclusions

While it is sensible to think that a more extensive period of training would have improved the corpus skills of learners, it must be stressed that in our research we adopted Liu & Jiang's (2009) suggestions to present modelling and warm-up activities that helped contextualize corpus work (see Appendix A and B). In this way, it cannot be claimed that our informants were unaware of the work with corpora and/or practical ways to compensate for a presumed lack of familiarity. Despite this awareness, and their previous experience with online concordancing, our learners performed unsophisticated searches that did not take advantage of the full potential of corpora in language learning.

Future research should strive to confirm whether this situation is similar in other learning situations and/or L1 contexts and gain deeper knowledge into the factors that may affect the *initiation* stage of learner corpus consultation, including cognitive, interface design as well as corpus suitability issues. Similarly, a combination of quantitative and qualitative methods (Figures 2 and 3) may prove fruitful in understanding the relationship between learners' search behaviour and the completion of focus-on-form activities.

Notes

1. For a more detailed account of the use of tracking devices to record student-computer interaction during corpus consultation see Pérez-Paredes et al. (2011).

2. Boulton, A. *Empirical research in data-driven learning — a summary*. Web supplement to Boulton (2010). Available at: http://arche.univ-nancy2.fr/file.php/967/DDL_empirical_survey_2011_April.pdf [accessed April 2011]

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

3. Compleat Lexical Tutor: http://www.lextutor.ca/concordancers/concord_e.html
4. University of Leeds' suite of corpora: http://corpus.leeds.ac.uk/protected/query.html
5. Fiddler Web Debugging Proxy: http://www.fiddler2.com/fiddler2/

References

Acar, A., Geluso, J. & Shiki, T. 2011. "How can search engines improve your writing?". *CALL-EJ*, 12 (1), 1–10.

Ädel, A. 2010. "Using corpora to teach academic writing: Challenges for the direct approach". In M. C. Campoy-Cubillo, B. Belles-Fortuño & M. L. Gea-Valor (Eds.), *Corpus-Based Approaches to ELT*. London: Continuum, 39–55.

Aijmer, K. (Ed.) 2009. *Corpora and Language Teaching.* Amsterdam/Philadelphia: John Benjamins.

Aston, G. 1997. "Involving learners in developing learning methods: Exploiting text corpora in self-access". In P. Benson & P. Voller (Eds.), *Autonomy and Independence in Language Learning*. London: Longman, 204–214.

Aston, G., Bernardini, S. & Stewart, D. (Eds.) 2004. *Corpora and Language Learners.* Amsterdam/Philadelphia: John Benjamins.

Barlow, M. 2011. "Corpus linguistics and theoretical linguistics". *International Journal of Corpus Linguistics*, 16 (1), 3–44.

Bellés-Fortuño, B., Campoy, M. C. & Gea-Valor, M. L. (Eds.) 2010. *Exploring Corpus-Based Research in English Language Teaching.* Castellón de la Plana: Publicacions de la Universitat Jaume I.

Bernardini, S. 2000a. *Competence, Capacity, Corpora. A Study in Corpus-Aided Language Learning*. Bolonia: 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*. Frankfurt: Peter Lang, 225–234.

Bernardini, S. 2002. "Exploring new directions for discovery learning". In B. Ketteman & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 165–182. Also available at: http://docserver.ingentaconnect.com/deliver/connect/rodopi/09215034/v42n1/s14.pdf?expires=1274809284&id=56951697&titleid=1000&accname=Guest+User&checksum=F32B3255DD49CE1741E60D9BBC313EDC

Bloch, J. 2009. "The design of an online concordancing program for teaching about reporting verbs". *Language Leaning and Technology*, 13 (2), 59–78. Available at: http://llt.msu.edu/vol13num1/bloch.pdf

Boulton, A. 2008a. "DDL: Reaching the parts other teaching can't reach?". In A. Frankenberg-Garcia (Ed.), *Proceedings of the 8th Teaching and Language Corpora Conference.* Lisbon: Associação de Estudos e de Investigação Científica do ISLA-Lisboa, 38–44.

Boulton, A. 2008b. "Evaluating corpus use in language learning: State of play and future directions". *American Association of Corpus Linguistics. Provo, UT (USA), 13–15 March*. Available at: corpus.byu.edu/aacl2008/ppt/6.ppt

Boulton, A. 2009. "Corpora for all? Learning styles and data-driven learning". In M. Mahlberg, V. González-Díaz & C. Smith (Eds.), *Proceedings of the 5th Corpus Linguistics Conference*. Liverpool: UCREL.

Boulton, A. 2010a. "Data-driven learning: Taking the computer out of the equation". *Language Learning*, 60 (3), 534–572.

Boulton, A. 2010b. "Learning outcomes from corpus consultation". In M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (Eds.), *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*. London: Equinox, 129–144.

Boulton, A. 2011a. "Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies". In N. Kübler (Ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang, 69–96.

Boulton, A. 2011b. "Language awareness and medium-term benefits of corpus consultation". In A. Gimeno Sanz (Ed.), *New Trends in CALL — Working Together*. Madrid: Macmillan ELT, 39–46.

Bowker, Y. 1998. "Using specialized monolingual native-language corpora as a translation resource: A pilot study". *Meta*, 43 (4), 631–651.

Braun, S. 2007. "Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora". *ReCALL*, 19 (3), 307–328.

Braun, S., Kohn, K. & Mukherjee, J. (Eds.) 2006. *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods.* Frankfurt: Peter Lang.

Breyer, Y. 2006. "My concordancer: Tailor-made software for language learners and teachers". In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt: Peter Lang, 157–176.

Broder, A. 2002. "A taxonomy of web search". *SIGIR Forum*, 36 (2), 3–10.

Campoy, M. C., Belles-Fortuno, B. & Gea-Valor, M. L. (Eds.) 2010. *Corpus-Based Approaches to English Language Teaching*. London: Continuum.

Chambers, A. 2005. "Integrating corpus consultation in language studies". *Language Learning and Technology*, 9 (2), 111–125. Available at: http://llt.msu.edu/vol9num2/chambers/default.html

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*. Amsterdam: Rodopi, 3–16.

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

Chan, T. & Liou, H. 2005. "Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations". *Computer Assisted Language Learning*, 18 (3), 231–250.

Chapelle, C. & Mizuno, S. 1989. "Students' strategies with learner-controlled CALL". *CALICO Journal*, 7 (1), 25–47.

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

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

Conrad, S. 1999. "The importance of corpus-based research for language teachers". *System*, 27 (1), 1–18.

Davies, M. 2004. "Student use of large corpora to investigate language change". In B. Ketteman & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis.* Amsterdam:

Rodopi, 165–182. Also available at: http://www.ingentaconnect.com/content/rodopi/lang/2004/00000052/00000001/art00012

Estling Vannestål, M. & Lindquist, H. 2007. "Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course". *ReCALL,* 19 (3), 329–350.

Farr, F. 2008. "Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users". *Language Awareness*, 17 (1), 25–43.

Fischer, R. 2007. "How do we know what students are actually doing? Monitoring students' behaviour in CALL". *Computer Assisted Language Learning*, 20 (5), 409–442.

Flowerdew, J. 1996. "Concordancing in language learning". In M. Pennington (Ed.), *The Power of CALL*. Houston, TX: Athelstan, 97–113.

Frankenberg-García, A. 2005. "A peek into what today's language learners as researchers actually do". *International Journal of Lexicography*, 18 (3), 335–355.

Frankenberg-García, A. 2010. "Raising teachers' awareness to corpora". *Language Teaching*. FirstView Articles, 45 (4), 1–15.

Gabrielatos, C. 2005. "Corpora and language teaching: Just a fling, or wedding bells?". *TESL-EJ*, 8 (1), 1–35.

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

Gilmore, A. 2009. "Using online corpora to develop students' writing skills". *ELT Journal,* 63 (4), 363–372.

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*. Frankfurt: Peter Lang, 69–86.

Granath, S. 2009. "Who benefits from learning how to use corpora?". In K. Aijmer (Ed.), *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins, 47–65.

Hafner, C. & Candlin, C. 2007. "Corpus tools as an affordance to learning in professional legal education". *Journal of English for Academic Purposes*, 6 (4), 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 online database". *Language Learning & Technology*, 9 (2), 90–110. Available at: http://llt.msu.edu/vol9num2/pdf/horst.pdf

Johansson, S. 2009. "Some thoughts on corpora and second language acquisition". In K. Aijmer (Ed.), *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins, 33–44.

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

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*. Harlow: Addison Wesley Longman, 110–115.

Johns, T., Lee, H. C. & Wang, L. 2008. "Integrating corpus-based CALL programs in teaching English through children's literature". *Computer Assisted Language Learning,* 21 (5), 483–506.

Kaszubski, P. 2006. "Web-based concordancing and ESAP writing". *Poznań Studies in Contemporary Linguistics*, 41, 161–193.

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 (4), 287–310.

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

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. Available at: 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*. Amsterdam: Rodopi, 237–252.

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 (1), 56–75.

Liu, D. & Jiang, P. 2009. "Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts". *The Modern Language Journal*, 93 (1), 61–78.

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

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

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

Moreno Jaén, M., Serrano Valverde, F. & Calzada Pérez, M. (Eds.) 2010. *Exploring New Paths in Language Pedagogy. Lexis and Corpus-Based Language Teaching.* Londres: Equinox.

O'Dell, F. & Broadhead, A. 2002. *Objective CAE. Students' Book Self-Study*. 2nd ed. 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 (3), 269–286.

O'Sullivan, Í. & Chambers, A. 2006. "Learners' writing skills in French: Corpus consultation and learner evaluation". *Journal of Second Language Writing*, 15 (1), 49–68.

Palfrey, J. & Gasser, U. 2008. *Born Digital. Understanding the First Generation of Digital Natives.* New York: Basic Books.

Pérez-Paredes, P. 2010. "Appropriation and integration issues in corpus methods and mainstream language education". In T. Harris & M. Moreno Jaen (Eds.), *Corpus Linguistics in Language Teaching.* Bern/Frankfurt am Main: Peter Lang, 53–73.

Pérez-Paredes, P. & Alcaraz-Calero, J. M. 2009. "Developing annotation solutions for online data driven learning". *ReCALL*, 21 (1), 55–75.

Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J. M. & Aguado Jiménez, P. 2011. "Tracking learners' actual uses of corpora: Guided vs. non-guided corpus consultation". *Computer Assisted Language Learning*, 24 (3), 233–253.

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

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

Schroeder, A., Minocha, S. & Schneider, C. 2010. "The strengths, weaknesses, opportunities and threats of using social software in higher and further education teaching and learning". *Journal of Computer Assisted Learning*, 26 (3), 159–174.

Scott, M. 2008. "Developing WordSmith". *International Journal of English Studies*, 8 (1), 153–172.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. 2003. *Reading Concordances: An Introduction.* London: Longman.

Sinclair, J. (Ed.) 2004. *How to use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins.

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

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.

Tribble, C. & Jones, G. 1997. *Concordances in the Classroom: A Resource Book for Teachers*. Houston: Athelstan Press.

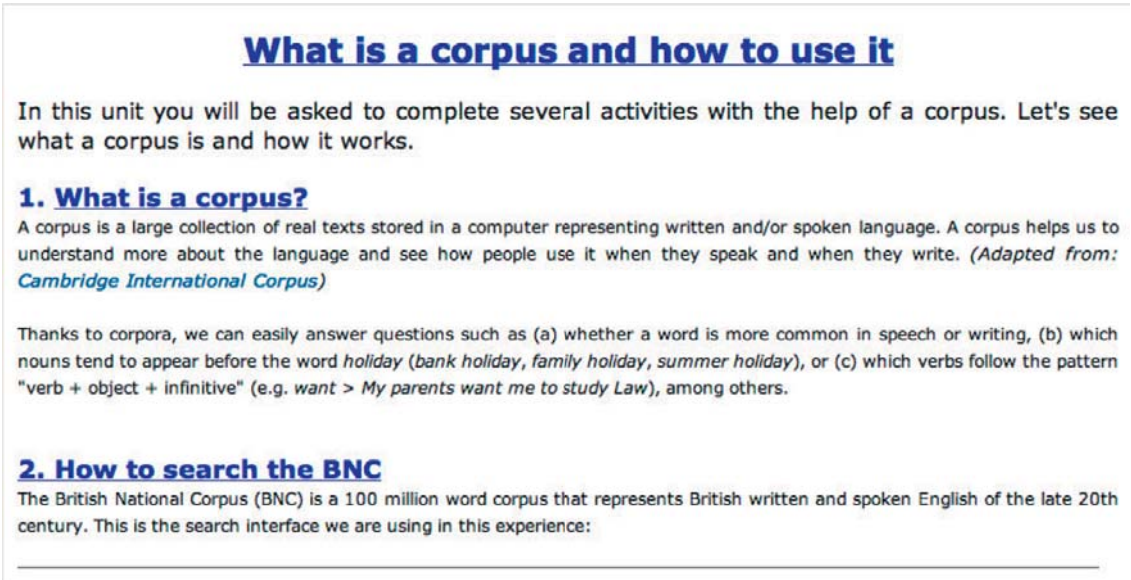
Varley, S. 2009. "I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment". *Computer Assisted Language Learning*, 22 (2), 133–152.

Wible, D., Kuo, C-H., Chien, F. & Wang, C. C. 2002. "Toward automating a personalized concordancer for data-driven learning: A lexical difficulty filter for language learners". In B. Ketteman & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 165–182.

Yoon, H. 2008. "More than a linguistic reference: The influence of corpus technology on L2 academic writing". *Language Learning & Technology*, 12 (2), 31–49. Available at: 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 (4), 257–283.

Appendix A



What is a corpus and how to use it

In this unit you will be asked to complete several activities with the help of a corpus. Let's see what a corpus is and how it works.

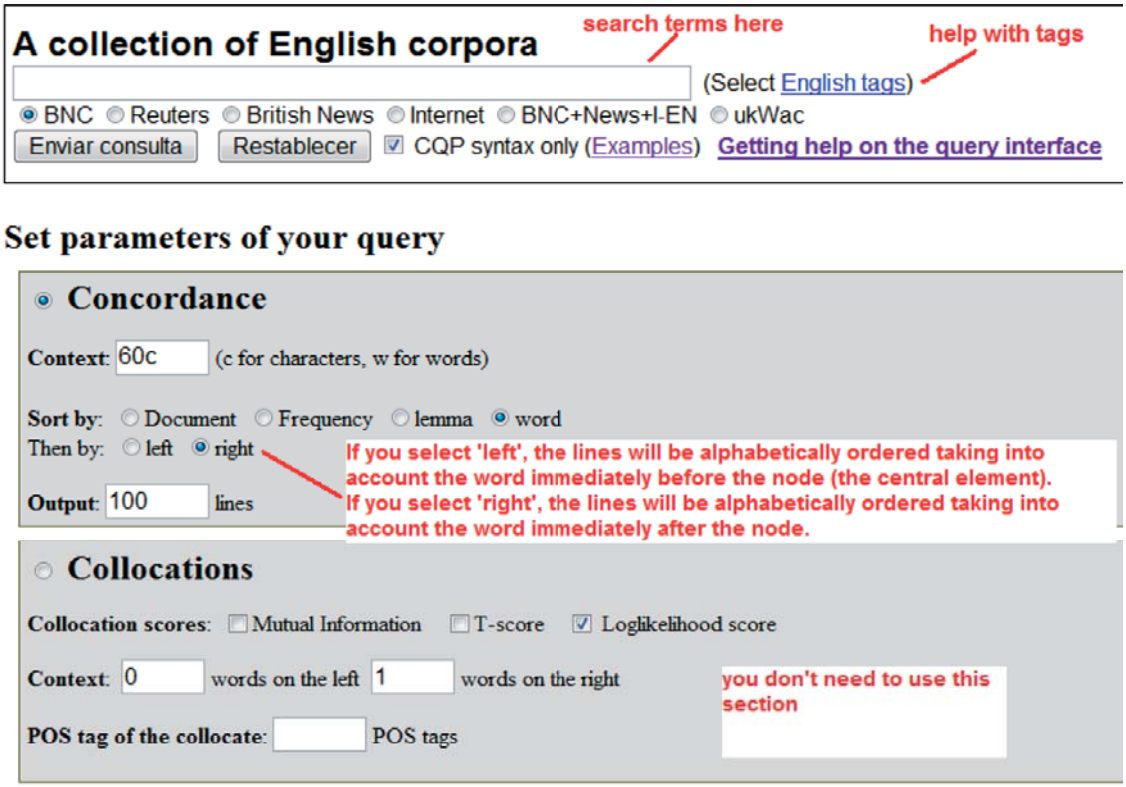
1. What is a corpus?

A corpus is a large collection of real texts stored in a computer representing written and/or spoken language. A corpus helps us to understand more about the language and see how people use it when they speak and when they write. *(Adapted from: Cambridge International Corpus)*

Thanks to corpora, we can easily answer questions such as (a) whether a word is more common in speech or writing, (b) which nouns tend to appear before the word *holiday* (*bank holiday*, *family holiday*, *summer holiday*), or (c) which verbs follow the pattern "verb + object + infinitive" (e.g. *want* > *My parents want me to study Law*), among others.

2. How to search the BNC

The British National Corpus (BNC) is a 100 million word corpus that represents British written and spoken English of the late 20th century. This is the search interface we are using in this experience:



To search the BNC you have to enter your search terms in the search box. Different search options are available:

To search the BNC you have to enter your search terms in the search box. Different search options are available:

1. You can enter a single word or word combinations. If you search for *holiday*, you will get this:

Dogs Today 10% discount voucher, redeemable against any **holiday** listed in the brochure and booked through Consort Central
largest concentration of easy routes, it is very crowded at **holiday** time. A growing menace in Verdon is the determination of
closely by May Day and not much later by the spring bank **holiday** . The NEDC Committee suggests that one of these could be
the children 's autumn half-term coincided with a bank **holiday** weekend for their parents, that would make sense both for
25% to 18%. In addition to the hotel income which a bank **holiday** encourages, the group was also aware of its importance in
. A much easier alternative is to come here on a summer bank **holiday** weekend, join the queue and be winched down in a bosun
an uninterrupted flow of production between the August bank **holiday** and Christmas. The fear is that if this is broken. Britain
Council (NEDC) points to the creation of a bank **holiday** in October as a way of extending the tourist season into
have such high-peak-season prices. " The October bank **holiday** proposal has received a muted response from Employment
always made things, " says Ned, " and while on a boating **holiday** on the Thames about 15 years ago I found a branch of very
, partly because of strong car sales and also credit card **holiday** bookings. BOOK REVIEW / The spy who lunched me: " The Ultra

this means 'bank' usually appears before 'holiday'

2. You can introduce grammatical information, that is, you can search for grammatical features. E.g. adjective + *of* + noun. To do this, you have to use the tags that appear next to the search box. If you want to search for combinations of adective + *of* + noun, you will have to enter the following search string: /JJ of /N.* and you will get this:

to the ground. Nor am I under the illusion that I alone am	**free of illusion**	. But that is the effect, he wrote, that is the effect I am
tax threshold. However, a gift to ACET is completely	**free of inheritance**	tax, due to the fact that ACET is a charity. Money that is
? At present, the first £128,000 of any estate is	**free of inheritance**	tax. Anything over this amount is, basically, liable to
way in his Peugeot from the coast, this one is n't "	**full of blood**	". But it is between coups, or unrests, and has lately been
's toes He sits Library books. Long books —	**full of quiet**	reading far from Mrs Raistrick 's ears, Books with
. The speech is not a lamenting " downer ", it 's	**full of irony**	, humour and compassion and honest indignation. She bursts

3. You can search for strings of words with several words in between. If you want to do this, you should use the wildcard [] {x,y}, where x stands for the minimum number of words in between and y stands for the maximum number of words in between. You will also have to click on CQP Syntax only (below the search box) and write the words between inverted commas "XX". Imagine you want to search for *was ..... by* combinations that have between 1 and 4 words in between. Your search string will be as follows: "was" [] {1,4} "by"

in England for the last eighty years. The starting point	**was the formation by**	Sir Herbert Tree of a training school at His Majesty
in 1985, was preceded by the Eliot biography of 1984, which	**was preceded by**	The Last Testament of Oscar Wilde of the year before, and
from the greetings card industry. The design itself	**was commissioned by**	Trading Officer Craig Methven and painted by Sheila Moxley,
or two books on the subject of AIDS. Initial sponsorship	**was provided by**	World in Need, a charity which also helped launch Action
burglaries and rapes resulting in the victim 's death,	**was approved by**	a Congressional Commission of the House of Deputies in
. One of the major accounts of the Renaissance in Italy	**was publisled in 1860 by**	a Swiss history professor at Basel University. Jacob
, however, this is a cogent and sensible account (which	**was constrained by**	a barbarous embargo on quotation). I do n't think
have it not, labour is in vain; genius is all in all. It	**was wittily said by**	a bright genius, who observed another to labour in the

Appendix B

Activity 1

1. Compare these sentences. What effect is produced by the sentences in *italics*?

- Policewomen deal with both victims and offenders. —> *It is policewomen who deal with both victims and offenders.*
- Everything can be shown through details. —> *It is through details that everything can be shown.*

Please, write and submit your answers below.

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.

Activity 2

2. The sentences in italics are it-clefts. Use the corpus to find more examples like these. Follow these steps:

- Go to the British National Corpus
- Search for examples that follow the same pattern as the it-clefts in exercise 1.
- Press "enviar consulta" and observe the results.
- Choose **three** examples of it-cleft.
- **Explain in your own words** the effect produced by this structure.

Hint

It-clefts consist of:

- the pronoun *it*
- a form of the verb *be*, optionally accompanied by *not* or and adverb such as *only*
- the focused element, which may be of the following types: a noun phrase, a prepositional phrase, and adverb phrase or an adverbial clause
- a relative-like dependent clause introduced by *that,who/which*, or zero.

Longman Student Grammar of Spoken and Written English (2002)

Tips for corpus search

- Read carefully the instructions provided in "What is a corpus and how to use it".
- Remember that you can use the wildcard [] {x,y} to search for word combinations with words in between. E.g. was .... by > **"was" [] {1,4} "by"**
- Remember to write the search terms between inverted commas "..." when using wildcards.
- Remember to mark CQP syntax only when using wildcards.
- Remember that you can use tags (English tags). Tags represent word classes, e.g. JJ = adjective; RBS = superlative adverb; VVD = past tense verb.

Activity 3

3. Rewrite the following sentences from "The open window". Emphasise the underlined word or words.

E.g. James bought a new car last week. It was a new car that James bought last week.
James bought a new car last week. It was James who bought a new car last week.

1 Poor aunt always thinks that they will come back some day.

Punto/s: --/1

Respuesta:

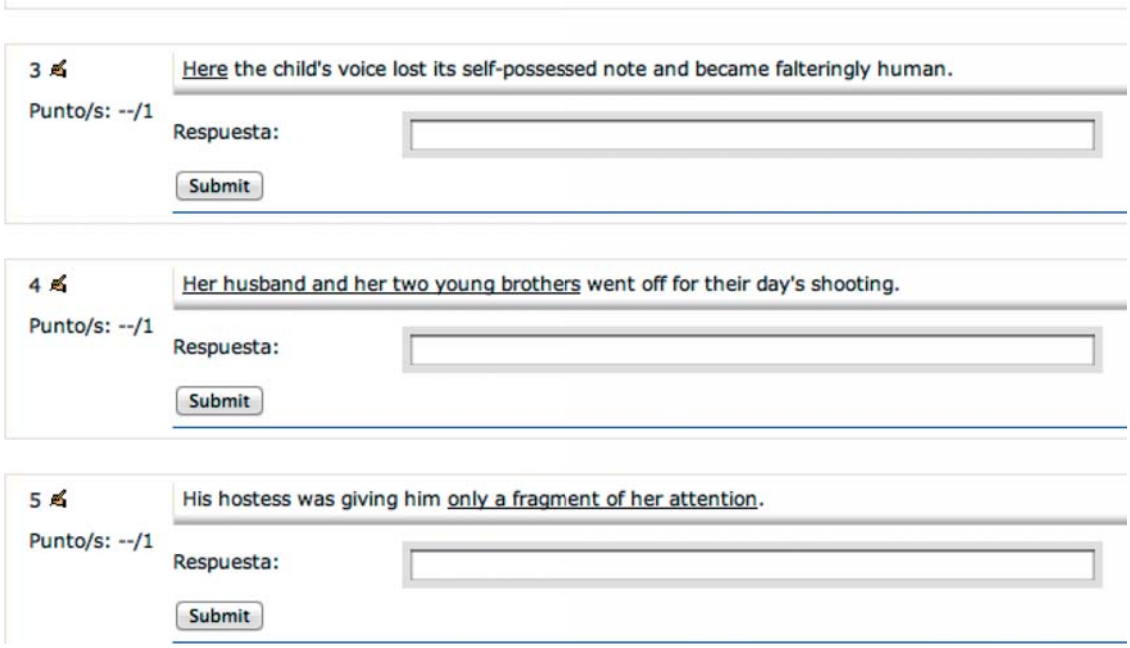
Submit

2 She broke off with a little shudder.

Punto/s: --/1

Respuesta:

Submit



Authors' addresses

Pascual Pérez-Paredes
Departamento de Filología Inglesa
Regional Campus of International Excellence
"Campus Mare Nostrum"
Universidad de Murcia
Facultad de Letras, Campus de La Merced
30071, Murcia
Spain

pascualf@um.es

María Sánchez-Tornel
Departamento de Filología Inglesa
Regional Campus of International Excellence
"Campus Mare Nostrum"
Universidad de Murcia
Facultad de Letras, Campus de La Merced
30071, Murcia
Spain

mstornel@um.es

Jose M. Alcaraz Calero
Cloud and Security Lab
Hewlett Packard Laboratories Bristol
Filton Rd, Stroke Gifford
BS34 8QZ Bristol
United Kingdom

jmalcaraz@um.es

Pérez-Paredes, P., Sánchez-Tornel, M. & Alcaraz Calero, J.M. (2012) Learners search patterns during corpus-based focus-on-form activities. A study on hands-on concordancing. International Journal of Corpus Linguistics, 17:4, pp. 482-515.