# Corpus Pedagogy: Analyzing Corpus Use in the Classroom and EFL Business Student Attitudes towards Corpora

John O'Donoghue\*

(Technical University of Applied Sciences Wildau)

**Chae Kwan Jung** 

(Korea Institute for Curriculum and Evaluation)

O'Donoghue, John, & Jung, Chae Kwan. (2013). Corpus pedagogy: Analyzing corpus use in the classroom and EFL business student attitudes towards corpora. *English Language Teaching*, 25(3), 51-74.

For over thirty years corpus linguists have been discussing the applicability of corpora to language teaching. While their use in compiling dictionaries and shaping teaching materials has been promoted and celebrated, using corpora directly in the classroom has been propagated by some but rejected, or even ignored, by many. Those who have conducted research into such direct use of corpora have tended to select postgraduate students who have an intrinsic interest in language or linguistics. The results of such research have invariably been rather positive. This study aims to examine the application of one corpus at undergraduate level with students whose focus is on the more general area of business and establish whether similarly positive results are to be observed. The results of which indicate that the majority of participants benefited from the experience and valued their corpus activity. There was, however, some criticism concerning navigating the British National Corpus (BNC) and the time-consuming nature of corpus work. The diverse nature of students' performances reflects the wide range of student responses to engaging with corpus. The student' attitudes indicate that an active guiding role by the teacher may significantly improve the quality of such learners' involvement with corpora.

#### I. INTRODUCTION

The idea of compiling a corpus has been an important aspect of linguistic research for many years. A number of corpora, for example the Bank of English (BOE), British National Corpus (BNC) and Corpus of Contemporary American English (COCA), have offered lexicographers, materials designers and teachers a new and exciting resource for accessing a considerable amount of authentic linguistic data. This has produced a wide range of innovative dictionaries and course books in the English as a Foreign Language (EFL) sphere in particular. The resources

\_

<sup>\*</sup> First author: John O'Donoghue; \*\*Corresponding author: Chae Kwan Jung

that language corpora comprise and the goals that language pedagogy pursues would seem to complement each other. The former offers a huge amount of authentic linguistic data that can be assessed by language learners, whose aim is ostensibly to attain a command of such data in terms of lexis and grammar primarily. And yet the relationship between corpus and language learning is not simply that of solution to problem, rather it appears that there has been continuing debate about the way in which large amounts of linguistic data can be applied both inside and outside the classroom. The initial enthusiasm of simply allowing students to access data as researchers, for example using Data-Driven Learning (Johns, 1991), seems to have succeeded by a more mediated approach to corpus application (Bernardini, 1998; Hong & Oh, 2008; Kennedy & Miceli, 2001; Tribble & Jones, 1997). At the same time it has become clear that for many teachers the whole area of corpus linguistics is one that hardly plays a role in their classroom experience (Kim & Chun, 2009; Seidhofer, 2002; Tribble, 2000).

Much of the research that has focused on corpora and language learners has selected advanced learners and/or those concentrating on languages or the humanities (e.g., Bernardini, 2000; Chambers, 2005; Chambers & O'Sullivan, 2004; Cheng, Warren, & Xun-feng, 2003; Chun, 2012). On the other hand, there have been studies which have dealt with more general types of learners (Boulton, 2007; Chambers, 2005; Mukherjee, 2004). One tentative explanation for the fact that many studies belong to the former group of more specialized graduates is that researchers may have had rather convenient access to such students at their own institutes. This study, in contrast, focuses on undergraduate students who appear to have little intrinsic interest in language as such, are not resident in an L2 speaking environment and may not intend to develop their competence to a native-speaker standard. Nevertheless, these students may be regarded as being representative of the majority of learners that teachers encounter in the classroom. Therefore, it seemed a worthwhile study to investigate how such students actually interact with the corpus, whether they were overwhelmed by the data or whether they were inspired to discover more about their target language. Students' feedback on their experience of corpus learning may assist teachers in adapting this tool to render corpus learning more effective in the classroom.

#### II. LITERATURE REVIEW

The Collins Birmingham University International language Database (COBUILD) project has been widely considered as one of the earliest and most important modern corpus studies using computers and has influenced many other corpus projects in the later 1990's (see more details in Carter, 1998). Table 1 shows some of the noteworthy corpus definitions suggested by different corpus researchers.

TABLE 1
Definitions of Corpus

Definitions of Corpus				
Researcher/Association	Defintion			
Sinclair (1991, p. 171)	"a corpus is a collection of naturally-occuring language text, chosen to characterize a state or variety of a language"			
Expert Advisory Group on Language Engineering Standards, EAGLES (1994, Section 2.1)	"a corpus is a collection of piece of language that are selected and ordered according to exploit linguistic criteria in order to be used as a sample of the language"			
McEnery & Wilson (1996, p. 177)	"corpus: (i) (loosely) any body of text, (ii) (most commonly) a body of machine-readable text, (iii) (more strictly) a finite collection of machine-readable text, sampled to be maximally representative of a language or variety"			
Baker, Hardie & McEnery (2006, p. 48)	"a corpus is a collection of texts (a 'body' of language) stored in an electronic database"			
O'Keeffe, McCarthy & Carter (2007, p. 1)	"a corpus is a collection of electronic texts, written or spoken, which is stored on a computer"			
Hunston (2008, CLARET Workshop, Birmingham, UK)	"a corpus is a collection of texts in any language(s) spoken (transcribed) or written usually naturally-occurring (not written specifically) stored and searchable electronically"			

The six corpus definitions in Table 1 overlap with the corpus definitions in the Concise Oxford English Dictionary (2004) which are 1) a collection of written texts, 2) a collection of written or spoken material in machine-readable form. However some of them have crucially important distinctions, for example 'naturally-occuring language' (Sinclair, 1991), 'electronic texts' (O'Keeff, McCarthy, & Carter, 2007) and 'stored and searchable electronically' (Hunston, 2008). Also, 'text' is more explicitly explained in later definitions. Hunston (2008) expands the meaning of 'text' to 'spoken' in which she includes transcribed texts in addition to written texts. Therefore, if we agree that the meaning of 'text' includes both 'spoken and written', then, a corpus can be interpreted as 'a collection of naturally occurring spoken or written texts which is searchable electronically' in this study (Jung, 2011).

As Hunston (2002) notes, "a corpus by itself can do nothing at all, being nothing other than a store of used language. Corpus access software, however, can re-arrange that store so that observations of various kinds of can made" (p. 3). The real benefit of corpora can only be achieved through the use of computational text analysis tools, for example, by investigating concordances or collocations (Baker, Hardie, & McEnery, 2006). By using advanced computer software packages, for example AntConc 3.2.4w (one of the most

well-known and free corpus analysis tools), we may store a corpus in a searchable database and carry out a wide range of queries. Computational text analysis can offer us powerful evidence-based arguments about linguistic features such as discourse structure (Biber, Connor, & Upton, 2007). It also helps us identify 'visual and textual clues' which can be a starting point when investigating a text (Hunston, 1994, 2002).

## 1. Corpus as a Description of Language

Leech (1992) notes that corpus linguistics is based on the assumption that analyzing recurrent patterns of real language helps us to describe language in a more realistic (or less unrealistic) and complete fashion; that observing language in use leads to the theory rather than the other way around. Consequently, learners will feel that what they are learning is a closer approximation to the target language with which they seek to interact. Corpus linguistics lends itself to an empirical approach to such an analysis.

Corpus linguistics, as has been pointed out, is not "a theory of language in competition with other theories of language such as transformational grammar" (Kennedy, 1998, p. 7), rather its theories of language are based on strong textual research evidence. Corpus work can be carried out with relatively old-fashioned teaching methods, serving to reinforce the status of the teacher and a rule-based approach to language acquisition (Sripicharn, 2004; Tribble, 1997). However, the nature of corpus analysis may indicate a different direction in language acquisition. As Partington (1998) points out when discussing the relation of corpus analysis to other resources, corpus analysis "not only constitutes a new technological device, but also provides a new philosophy for language description" (p. 1). A learner reading concordance lines can only benefit from this experience if she or he becomes aware of certain linguistic patterns and is able to identify similarities and contrasts in lexicogrammatical patterns.

## 2. Corpora, Language Learners, and Teachers

Consciousness-raising and language awareness have often been mentioned as the key effects that corpus analysis has on learners (Gavioli, 2005). In this regard, corpus-based teaching activities can be related to doing task-based teaching to some extent. Willis and Willis (2007) note that "learners have to apply their own initiative and work things out for themselves. This process is likely to increase their awareness of the language" (p. 125). Especially noticing has been regarded as preceding understanding and important for language production in conjunction with language awareness. Schmidt (1990) proposes 'Noticing Hypothesis,' arguing that features of language cannot be learned if they have not been noticed by language learners. Flowerdew (2012) also adds that learners will improve

language acquisition "if their attention is drawn to salient linguistic features" (p. 7).

In a similar vein, Ellis (2003) regards attention to input as a conscious process. He conceptualizes noticing and noticing-the-gap. He illustrates that noticing involves registering formal features in the output whereas noticing-the-gap involves identifying how the output to which the learner is exposed differs from the output the learner is able to generate. From a didactic point of view with particular emphasis on the Noticing Hypothesis, the value of corpus work lies in the huge range of output which the learner can be exposed to and is able to 'notice.' With careful teacher mediation, the gap between the learner's output and more native-speaker-like production can be bridged.

As we have reviewed, corpora allow us to gain a myriad of insights into the workings of language. Kennedy (1998) found the advantage of corpus analysis helpful mentioning "richness of the evidence and the confidence we can have in the generalizability of that evidence" (p. 8). The resource of concordance lines may, on the one hand, be a means of continued discovery through the rich complexity of authentic language but, on the other hand, it may serve to undermine the rule-based systematic approach to language acquisition that many teachers still adhere to. In the former case we may see the corpus as filling the gap where "the rule does not reflect the evidence of language use" (Tognini-Bonelli, 2001, p. 10). Partington (1998) also offers two approaches to using corpus materials for teachers: one is to utilize them for designing course materials, the other is to introduce the students to corpora in the classroom and teach them how to use this resource. In the former situation the teacher selects specific material (e.g. textbooks) to reinforce or demonstrate a particular teaching point. In the latter there is a choice between the teacher determining the path of investigation and students being allowed to search the corpus without restriction.

#### III. METHODS

#### 1. Participants

The research participants for this study were thirty-nine second-semester students of Business Administration taking Business English for the first time in their studies. The University is a Technical University of Applied Sciences in Brandenburg, founded in 1991 and situated just outside the capital Berlin. Such a university emphasizes a practical approach to its degree content, with the focus on equipping students with skills that are required by companies both regionally and internationally.

The degree course comprises six semesters and the students have sixty contact hours in their first semester of Business English. Time both in and outside the classroom was therefore a limited resource. In view of the fact that corpus has often been applied in areas such as English for Academic Purposes (EAP) or English for Specific Purposes (ESP), this Business English course could be regarded as being situated somewhere between the two since they were pursuing a Bachelor of Arts (BA) in Business Administration. Business English was one of the compulsory courses in the syllabus and English is the only foreign language in which most of them are familiar and their language level ranges from B1 to C1 in terms of the European Reference Framework of Reference for Languages (Council of Europe, 2001). German students would generally have had seven to eight years of English instruction in school prior to entering a university. The interviews were conducted in June 2012, each one lasting approximately fifteen minutes.

## 2. Working with the BNC in class

The students were introduced to corpus in the first week of term. Within the context of a traditional teaching framework, translating from L1 (German) to L2 (English), a particular lexical item profit was reproduced with fourteen concordance lines from the BNCweb (http://bncweb.lancs.ac.uk). The first question was designed to alert the students to the fact that the translation of profit as Gewinn in German had an obvious limitation in that the German equivalent covers the noun but not the verb in English. The students were then encouraged to revise their assumptions concerning one-to-one equivalence translation. The first encounter with concordance lines was therefore of a rather controlled nature, with specific questions that related to the reproduced concordance lines as shown in APPENDIX I. These questions concentrated both on collocational and colligational patterns as well as semantic prosody (Sinclair, 2003). Having conducted a controlled exercise in week one, week two was dedicated to introducing the students to the BNC. We chose the BNCweb because it appeared relatively user-friendly and was free of charge. Once all students had registered with the BNCweb, a whole class period was dedicated to familiarizing the students with the BNC using a guided approach (see APPENDIX II). The students were then asked to choose a key business lexical item and analyze it using the model demonstrated in class.

## 3. A Brief Questionnaire Survey and Interviews

Dornyei (2007) distinguishes the differences between quantitative and qualitative research methods in the area of Applied Linguistics and English language teaching. According to him, the central goal of quantitative research is to generalize research findings whereas qualitative research is to highlight situated or personal and multiple meaning of the findings. A questionnaire survey was conducted with thirty-nine students to gain a general

impression of students' experience of corpus (APPENDIX III). It was decided to use the qualitative approach using a semi-structured interview to clarify some of the issues arising from the results of the questionnaire. This is the central focus for this study as this study attempted to gain deeper insights into the comprehensive attitude of using corpora in the classroom by investigating the students' experience. Therefore a mixture of male and female students were invited to participate in the interviews who did so on a voluntary basis. Due to time restraints and the in-depth nature of the interviews, it was judged that six students would provide sufficient diversity of experience to yield relevant feedback on corpus learning. The basis for the interviews was a series of questions which focused on the students' experience of working with the BNC and dealing with issues such their individual use of the corpus, challenges they encountered, serendipity learning, language awareness and complementary use of bilingual dictionaries.

The interviews were recorded in the instructor's office and lasted approximately fifteen minutes, although some were longer depending on the length of the participants' answers. Most of the questions were identical in each interview but as the interviews had a conversational character there were some deviations to follow up students' individual opinions. By using the QSR NVivo 10 (one of the most well-known computer assisted qualitative data analysis software) and AntConc 3.2.4w in conjunction with Constituent Likelihood Automatic Word-tagging System (CLAWS) Part-of-speech (POS) C7 tagset, the interview transcripts were scrutinized to identify the main issues regarding the interviewees' experience (APPENDIX V). The main reason why we POS tagged all the interview scripts were to learn more about student attitudes towards corpora by looking at general adjectives (e.g. complicated, confusing, etc.), general comparative adjectives (e.g. clearer, faster, etc.) general superlative adjectives (e.g. easiest, best, etc.), and catenative adjectives (e.g. able to, etc.).

#### IV. RESULTS AND ANALYSIS

Overall, the interviewees were positive about using the BNC which enabled them to discover the context around the key word and to observe how key lexical items are used in different contexts both generally and in specific commercial/financial domains.

## 1. Questionnaire Survey

Following the completion of the task described above the students were given a questionnaire to complete in class. Thirty-nine students enrolled in the business administration degree course filled out the questionnaire. Table 2 and Table 3 represent the

answers that the student provided to questions.

TABLE 2	
Using the British National Corp	us

Category	Useful	Not Useful	Mean	SD
Usage of Vocabulary	79.5%	20.5%	4.20	1.16
Usage of Phrases	79.5%	17.9%	4.60	1.08
Usage of Grammar	61.5%	33.3%	3.83	1.19
Reading Skill	28.2%	64.1%	3.08	1.06
Writing Skill	46.2%	48.7%	3.54	1.30

The statements that the students reacted to most positively in this questionnaire were that they understood the purpose of using the BNC (87.2%) and they found the BNC helpful for learning the usage of phrases (79.5%). This response would fit in with the phraseological approach to language that corpus promotes, "the way in which meaning is sometimes associated with a whole phrase rather than a single word" (Sinclair, 2003, p. 10). The students also valued the BNC in terms of learning the meaning of vocabulary (79.5%). These findings are similar to those of Yoon and Hirvela (2004) whose students responded that corpus use was most helpful for learning the usage of vocabulary and phrases.

TABLE 3
Using the British National Corpus (cont.)

Category	Agree	Disagree	Mean	SD
Own Choice	17.9%	82.1%	2.41	1.27
Purpose	87.2%	12.8%	4.56	1.12
Future Writing	69.2%	28.2%	3.81	1.29
Useful Resource	59.0%	41.0%	3.79	1.30
Helpful Writing	74.4%	20.5%	4.34	1.34
Helpful Reading	15.4%	79.5%	2.32	1.18
Third level	59.0%	35.9%	3.89	1.50
Secondary	69.2%	25.6%	4.16	1.32
Improved Understanding	59.0%	38.5%	3.57	1.18
Cut-off Sentences	41.0%	35.9%	3.63	1.19
Real Texts	41.0%	59.0%	3.17	1.25
Unfamiliar Vocabulary	53.8%	30.3%	3.75	0.95

There was nearly a twenty percent drop when it comes to how useful the corpus is for learning grammar. Although prepositions played a major role in the introductory guide to the BNC and in the students' own analysis of key business terms, the students may associate grammar more with tenses in a more structured format and therefore have ignored the

grammatical element in favor of a lexical concentration. There is also a clear difference in the responses between corpus as a tool to be used for reading or for writing. Despite the fact that corpus can, of course, be used to aid reading comprehension as Gavioli (2005) observed above, it was valued essentially as a writing tool; 74.4% compared to 15.4%. As the interviews revealed later, most students found a dictionary to be a more effective tool for looking up unknown words. Apparently, when the students have to produce language (write) rather than understand it (read) they turn to the corpus. The difference may be that when reading a word there is often one specific meaning that needs to be identified. When writing and producing language the key word can be used in a myriad of ways that the corpus can help to demonstrate. Corpus analysis is seen both as a way of looking how the word works and as a source of ideas as one of the interviewees mentioned. In the following section, we will report the results of the in-depth individual interviews with analysis.

#### 2. Interviews

#### 1) Student 1 (S1)

S1 said that she enjoyed the inductive method the BNC offers, "to figure it out yourself' and thought this enhanced memory storage of such items in contrast to the rule-based approach experienced in school. S1 also mentioned that it helped her learn 'chunks,' something she found valuable based on her experience of time spent in America, where she obviously had difficulty comprehending certain fixed phrases or idioms. This may indicate a special need for such culturally determined prefabricated phrases to be explicitly taught. As this was the first time that any of them had used any corpus tool and many of them mentioned the time-consuming nature of the work, the issue whether it would be preferable to do more or less corpus work in class was discussed. S1 felt that we could have spent more time on corpus activities if other teachers (in secondary school) had introduced the resource and students were already familiar with it at third level. In other words, if students had already been familiar with corpus it would have been possible to work more extensively with it in class. S1 found the interface "a little difficult to use" but said that the more often she used it the easier it became. S1 also pointed to the vocabulary in the lines as a source of difficulty.

#### 2) Student 2 (S2)

The second student used the BNC for her term essay and gained ideas for it which fitted the general context. The BNC was useful in showing her how words work. S2 found navigating her way around the BNCweb rather challenging, confusing and time-consuming, for example, having to enter brackets for certain queries, although this did improve with

practice. Due to time pressure she did not experience any incidental learning but the commercial or financial domain of the BNC did provide her with some insights concerning business lexis.

## 3) Student 3 (S3)

S3 also used the BNC for her essay "to know in which context I can use the word". It also seems to have helped her to look for additional lexical items around the key word. This student seems to have fewer problems navigating the BNC than S2 although she reported that at the beginning it "was hard to look at what you can use." This student seems to have engaged with corpus activities quite considerably for her writing tasks and would consider using it to write her BA dissertation in English, although she never used it for reading. This attitude explains the fact that nearly 80% of students said that the corpus was not helpful with reading. Finally, this student thought that corpus should be used in schools, an opinion that seems to enjoy widespread support with agreement from two-thirds of students. Moreover, the feeling expressed by various students in classroom discussion was that if secondary-school pupils were introduced to corpus work, it would save considerable time at third-level education. As with language acquisition in general, starting with corpus earlier rather than later was regarded as more effective.

#### 4) Student 4 (S4)

The fourth student was chosen because he was both communicative and openly critical of the BNC. He found navigating around the BNC to be "not that easy to use and handle". He picked upon the term *query* as an example of a word in the BNC interface that many non-native speakers would not necessarily understand. He had considered using the BNC for writing but had decided for the online English-Germany dictionary *Leo* (www.dict.leo.org) instead as the BNC was just "too complicated". He also found the truncated nature of the concordance lines confusing. Additionally, the fact that online sources provide translation was regarded positively in contrast to the monolingual BNC. On the other hand, he found the business domain in the BNC useful for demonstrating the use of lexis and would consider using the corpus for his bachelor dissertation as more time would be available then. Overall though, it seems that for this student working with the BNC was simply too much effort.

## 5) Student 5 (S5)

The fifth student was motivated to work with the BNC but found the task quite

time-consuming. However, she compared the activity to mind-mapping and called it an "adventure", being able to discover so many different aspects of a lexical item even though at times she felt "overloaded" by the experience as the corpus appeared never-ending. She also emphasized the novelty value of corpus activity, which as we have seen is positive in the sense of learning-by-discovery and negative in the sense of demanding an initial effort, considerable time and a readiness to deal with a new, non-intuitive interface. This student thought it would be a good tool for parents to help their children in learning. She complained about the statistical evidence, which she said gave her "headaches" and that some financial vocabulary was unfamiliar to her. Similarly to the previous student, she reported that she used dictionaries with her essay because it was quicker, although it seems that she was suggesting that the BNC would be a good source for doing preliminary research for an essay rather than looking up words at a later stage.

#### 6) Student 6 (S6)

The final student returned to the issue of time and how time-consuming and "overwhelming" corpus research is, commenting that it takes two or three minutes to analyze a particular word. To observe how a particular word is used in various ways, a corpus is useful but simply to understand a particular word (in one specific, defined meaning) the dictionary was preferred. The translation element was an added attraction for this student. This student analyzed the word instance and found that it is used as a noun, adjective and verb. He seemed to enjoy that discovery experience as he could, "get to know a word, really, like you didn't know it before" but at the same time he felt overwhelmed by ten-thousand concordance lines that a query might return. He ended, however, on a positive note by saying that if he were to write his BA dissertation in English he would use the corpus, because in an unknown field it would provide a lot of information that he would not find in a dictionary. In that respect he held a similar view to the previous student, who saw the value of corpus use in becoming acquainted with a relatively unknown word and finding out more about the different ways in which a particular lexicogrammatical item behaves, which he referred to as the "real meaning."

In summary, the students had problems with the BNC interface, found the process of investigating words time-consuming and some vocabulary difficult. Some felt overwhelmed by the amount of data. For these reasons many preferred to use dictionaries in their writing. On the other hand, the students enjoyed the fact that they had to work things out for themselves (the inductive approach) and that the concordance lines provided greater context than dictionaries. It seems that when time is not a pressing factor, corpus is considered a positive alternative. Corpus may be viewed as an adventure that leads in many different directions; this is seemingly both attractive and unattractive for students.

#### V. DISCUSSIONS

As mentioned above, these interviews were conducted following a questionnaire completed by thirty-nine undergraduate students of Business Administration. The issues which the six students highlighted are reflected in the results of this questionnaire. In other words, 53.8% of the thirty-nine Business Administration undergraduate students found the unfamiliar vocabulary difficult and 41.0% responded that the cut-off sentences and the fact that the concordance lines were made up of real texts made it difficult for them to use the corpus. On the other hand, 79.5% indicated that they found the BNC useful for learning how vocabulary and phrases are used. 69.2% said that they would use corpus later in their studies and the same number thought that corpus should be taught in secondary schools in Germany (See Questions 31, 32, 33, 25, 26, 40, 46 in Section B, APPENDIX II respectively).

How does this compare with other studies dealing with corpus? Although Charles (2011) in her study was dealing with post-graduate students in the UK under arguably more favorable circumstances, 34% of her students agreed that concordance lines took too much time because there was a lot of data. Only 62% thought it was easy to perform searches and a similar number thought the language not to be difficult. An earlier study by Yoon and Hirvela (2004) reported that 75% of intermediate students had problems with the amount of time necessary to analyze data from corpora. 62% of their students found cut-off sentences problematic and 38% found unfamiliar vocabulary, real texts and the search technique difficult (Yoon & Hirvela, 2004, p. 270). While different studies invariably choose different subjects under different conditions it seems that a pattern of challenges is emerging regarding learners' problems when encountering corpora.

What lessons can be drawn from analyzing these corpus activities? Looking at the results from the questionnaires in this study and others it may be recommendable to equip the students with some linguistic framework or filter to improve their analytic skills and language awareness while performing corpus analysis. This should probably be done on a one-to-one/small group basis, as a general introduction to language analysis might be of little interest. For example, when looking at the word *benefit* in class, students listed many modal verbs although the term 'modal' itself was unfamiliar to them. The students were then encouraged to speculate on why the word *benefit* attracts so many modal verbs, and what it might tell us about this word in semantic or pragmatic terms.

Another possible lesson may be for teachers to allow learners to immediately find (or to direct them to) information relevant to their course of studies, without being distracted by information which the corpora can offer but which does little to immediately increase language awareness, for example statistical information on gender or class. For many EFL students who have little intrinsic interest in language as such, it is important to guide their corpus investigations in such a way that an immediate and tangible return on their time and energy invested is evident; as

Tribble (1997) has commented "the most useful corpus for learners ... is the one which offers a collection of expert performances in genres which have relevance to the needs and interests of the learners" (p. 3). This may mean ignoring some of the information that linguists find interesting in favor of concentrating on lexicogrammatical items that are clearly identifiable as a priority for language learners. It is the teacher's role to identify these relevant areas – e.g. for our students the finance/commerce domain in the BNC – and subsequently to facilitate learners in determining their own points of research. To use Aston's distinction between reference and browsing, working with corpora may serve to "clarify doubts on particular problems which had arisen in other language activities" in a back-up function, or, on the other hand "a corpus may be treated as a source of activity in itself" (Aston, 1997, p. 205). It was the intention of this study to attempt to link these two distinct activities in a step-by-step approach, so that learners would do more than just complement dictionaries and yet not be overawed by the complexity and unfamiliarity of corpora. If we compare this study to one carried out by Charles (2011) we can see that even at a higher level (post-graduate) language difficulty remains a challenge to learner.

Furthermore, learners should be encouraged not merely to imitate the instructor's approach but rather to reflect critically on their own learning experience and potential deficiencies as the basis for investigative research, for instance by asking themselves, "What is it I need to learn about this word (that I did not know before?)". Adopting the path metaphor of corpus investigation, it is important that learners are not distracted by minor offshoots which are potential dead ends. The teacher may be in a good position to recognize when a path is leading nowhere and gently guide the learner on to a more rewarding route of discovery. For the vast majority of the students who were the subject of this research study, the idea of giving them "direct access to the data so that the learner can take part in building up his or her own profiles of meaning and uses" (Johns, 1991, pp. 30-31) may be inspirational, but unfortunately not entirely realistic in terms of the linguistic awareness and limited skills that many intermediate learners of English possess.

The role of the teacher may be described as starting with providing learners with clear guidance and moving on to awarding them maximum independence as they grow in mastering corpus analysis. It may well be that this process requires more than a few classes of instruction and one set assignment, and that in the following semester subsequent corpus work may yield greater rewards.

#### VI. CONCLUSION

Corpus pedagogy in this study primarily takes a corpus-based approach. In other words, learners were encouraged to approach language by analyzing recurrent patterns of real language and observing language in use employing an inductive rather deductive approach to language

acquisition. As a result, learners feel that what they are learning is a closer approximation to the target language with which they seek to interact.

It should be remembered that this was the first time these students had been confronted with corpus activity and that the focus of their studies is Business Administration and not linguistics. Due to the time constraints discussed above, they were provided with a relatively brief introduction to the BNC. Taking these circumstances into account, it is encouraging that so many students found the activity a useful one. Although the statistics are not as positive as Charles's report, one can agree with the statement "it is still possible to incorporate hands-on concordancing into courses and, as the feedback in this study shows, this can be a positive experience for many students" (Charles, 2011, p. 39) The students seemed to see its value mainly in observing how particular lexicogrammatical items behave and becoming more aware of certain collocational features. Some of them actively used the corpus for their term essay while others relied on more conventional sources such as online dictionaries.

The main challenges for students using the BNC as indicated by the questionnaire are unknown vocabulary and the need to understand the BNC's cultural background, both of which render some lines difficult to interpret and authenticate. The BNC is rooted in British culture and some references are rather opaque to German students. As one student observed in the interview, the solution may simply be to skip one particularly difficult line and find a simpler instance below it. This is, however, something that other students need to learn and more proficient students may feel more confident about doing this than less proficient ones. As the results of the study seem to suggest, far from the role of the teacher being restricted to introducing the learners to the mechanisms of the corpus, his/her linguistic expertise is required in guiding the learners to filter the data and interpret their findings, to categorize them in a meaningful way and present them to other students in a useful manner. Whether discussing with individual students or in groups, or by leading class discussions, the teacher's role may be to focus learners' attention on both familiar and unfamiliar linguistic forms and patterns.

One recurring issue is that of time. It may be worth considering whether class time should be used to allow students to perform corpus activities. In this way the issues raised above; lack of time outside class, lexical challenges, truncated sentences, authentic content could be addressed by the teacher directly. This may be a method to minimize the sense of being overwhelmed by the data as the instructor guides groups to focus on salient points and to learn to disregard irrelevant distractions. It may be worth noting that in Yoon and Hirvela's study cited above, the intermediate group spent more class time on corpus than the advanced one and reported a greater comfort level.

Appropriately for students in third-level education, the academic skills of analysis and evaluation are essential to carrying out corpus work. It is the instructor's role to use his/her linguistic expertise to assist students in acquiring or developing such transferable skills. The hurdle of a foreign language is one that necessitates such intervention, as neither native

speaker-like intuition nor linguistic awareness can generally be assumed.

This study indicates that it is meaningful to introduce corpus work to intermediate level EFL students who are taking a degree course in a non-humanities field. One open question is whether a different choice of corpora may have made the experience more attractive and if a sharper focus on more relevant linguistic features would have had a more motivating effect on some students. Future corpus work will offer the opportunity of exploring different approaches using a range of corpora based upon the insights gained in the present study.

#### **REFERENCES**

- 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* (pp. 204-263). London: Longman.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Bernardini, S. (1998). Systematising serendipity: Proposals for large-corpora concordancing with language learners. *Proceedings of TACL98* (pp. 12-16). Oxford, UK: Seacourt Press
- Bernardini, S. (2000). Systematizing serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking pedagogy from a corpus perspective* (pp. 225-234). Bern: Peter Lang.
- Biber, D., Cornor, U., & Upton, A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.
- Boulton, A. (2007). But where's the proof? The need for empirical evidence for data-driven learning. *Proceedings of the BAAL Annual Conference 2007* (pp. 13-16). London: Scitsiugnil Press.
- Carter, R. (1998). Vocabulary: Applied linguistic perspectives (2nd ed.). London: Routledge.
- Chambers, A. (2005). Integrating corpus consultancy in language studies. *Language Learning and Technology*, 9(2), 111-125.
- Chambers, A., & O'Sullivan, I. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, *16*(1), 158-172.
- Charles, M. (2011). Using hands-on concordancing to teach rhetorical functions: Evaluation and implications for EAP writing classes. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 26-43). London: Continuum.
- Cheng, W., Warren, M., & Xun-fung, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System, 31*(2), 173-186.
- Chun, S. (2012). Effective concordance workshop for corpus learners. Modern English

- Education, 13(4), 67-88.
- Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.
- Dornyei, Z. (2007). Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies. Oxford: Oxford University Press.
- Ellis, R. (2003). Task-based language learning and teaching. Oxford: Oxford University Press.
- Expert Advisory Group of Language Engineering Standards (EAGLES). (1994). *The EAGLES Guidelines*. Retrieved March 31 2013, from the World Wide Web: http://www.ilc.cnr.it/EAGLES/browse.html.
- Flowerdew, L. (2012). Exploiting a corpus of business letters from a phraseological: Functional perspective. *ReCALL*, *24*, 152-168.
- Gavioli, L. (2005). Exploring corpora for ESP learning. Amsterdam: John Benjamins.
- Hong, Sun-Yi, & Oh, Sun-Young. (2008). The effects of corpus-based learning of vocabulary and grammar on Korean high-school students. *English Language Teaching*, 20(1), 261-283.
- Hunston, S. (1994). Evaluation and organization in a sample of written academic discourse. In M. Coulthard (Ed.), *Advance in wirtten text analysis* (pp. 167-183). London: Routledge.
- Hunston, S. (2002). Corpora in applied linguistics. Cambridge: Cambridge University Press.
- Hunston, S. (2008). Can corpus linguistics (not) be applied? Paper presented at Corpus Linguistics Advanced Research Education and Training (CLARET), Birmingham, UK.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, 4, 1-16.
- Jung, C. K. (2011). Understanding undergraduate engineering laboratory reports (UELR): A genre and corpus approach. Unpublished doctoral dissertation. University of Warwick, UK.
- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, *5*(3), 77-90.
- Kennedy, G. (1998). An introduction to corpus linguistics. London: Longman.
- Kim, Heyoung, & Chun, Sooin. (2009). The trend of corpus-based CALL research: A qualitative meta-analysis of seventy-one selected studies. *English Language Teaching*, 21(2), 283-310.
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *New directions in corpus linguistics: Proceedings of Nobel Symposium 82* (pp. 125-148). Berilin: Mouton de Gruyter.
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp.

- 239-250). New York, NY: Rodopi.
- McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). From corpus to classroom: Language use and language teaching. Cambridge: Cambridge University Press.
- Partington, A. (1998). *Patterns and meanings using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Liguistic*, 11, 129-158.
- Seidlhofer, B. (2002). Pedagogy and local learner corpora: working with learning-driven data. In S. Granger, J. Hung & B. S. Petch-Tyson (Eds.), Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (pp. 213-234). Amsterdam; Philadelphia. PA: John Benjamins.
- Sinclair, J. (1991). Corpus concordance collocation. Oxford: Oxford University Press.
- Sinclair, J. (2003). Reading concordances. Edinburgh: Pearson Longman.
- Sripicharn, P. (2004). Examining native speakers' and learners' investigation of the same concordance data and its implications for classroom concordancing with ELF learners. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 233-245). Amsterdam: John Benjamins.
- Tognini-Bonelli, E. (2001). Corpus linguistics at work. Amsterdam: John Benjamins.
- Tribble, C. (1997). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *International conference on practical applications in language corpora* (pp. 106-117). Lódz: Lódz University Press.
- Tribble, C. (2000). Practical uses for language corpora in ELT. In P. Brett & G. Motteram (Eds.), *A special interest in computers* (pp. 31-41). Whitstable: IATEFL.
- Tibble, C. & Jones, G. (1997). Concordances in the classroom: Using corpora in langage education. Houston, TX: Athelstan.
- Willis, D, & Willis, J. (2007). Doing task-based teaching. Oxford: Oxford University Press.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*, 257–283.

#### APPENDIX I

#### Week One in Class

**Corpus task**: A Corpus is a collection of millions of words. I have just selected some lines for you. Don't worry that the lines are not complete, just concentrate on the few words directly in front of and behind the key word. The following is the entry for profit in the *Collins Cobuild Learner's Dictionary* 

#### Concise Edition (2003).

**Profit**. Profits, profiting, profited. 1 A profit is an amount of money that you gain when you are paid more for something that it cost you to make, get, or do it. *The bank made pre-tax profits of 3.5 million*. 2 If you profit from something, you earn a profit from it. *The dealers profited shamefully at the expense of my family*. 3 If you profit from something, or it profits only you, you gain some advantages or benefit from it. *So far the French alliance had profited the rebels little* ... *Whom would it profit to terrify or to kill James Sinclair*? Also a noun. *The artist found more to his profit in the sculpture collections*.

1	Television has recouped much of its	profit	margins. Television has also
2	only fair that all states should	profit	from the fruits of victory, in this
3	never lost an opportunity to make a	profit.	While such practices were not
4	that the only people to stand to	profit	from the schemes were the Hamilton's
5	Why should wealthy speculators now	profit	from their hardship? If the
6	Always on the look out for a quick	profit.	Weens wrote to his publisher
7	They did this by selling land at a	profit	to settlers who had some capital
8	Improved land could be sold at a	profit	and the proceeds used to buy new
9	in St Louis, Missouri, made a net	profit	of \$307m on sales of more than \$12
10	Mrs Thatcher's replacement, gloom as	profit	falls and warnings rolled in.
11	Falling shares pushed	profit	forecasts lower still, since many
12	Bank is likely to report next to no	profit	for 1990. Banks are bracing for
13	managed luxury groups have made a	profit	thanks to a booming world economy.
14	of 1990, but managed to stay in	profit	for the year. The Hongkong and

In the text above you saw how profit was used as a noun. Of course, as with many other English words, profit can be used in different ways.

- 1) See how *profit* is used in other ways, not just as a noun.
- 2) Which verbs are used with *profit*? Which verb is used the most?
- 3) Do you notice any lines using the structure Verb + Preposition + profit?
- 4) Which prepositions are used after *profit*? What differences do you notice between the words or phrases following these prepositions?
- 5) Is the word *profit* something negative or positive in each case? Or is it neutral? Go through the lines and quickly mark them as: Neg, Pos or Neut.
- 6) Have you noticed anything about the word *profit* that your teacher has not asked about?
- 7) What is the difference between reading the dictionary entry and looking at concordance lines in the previous exercise?

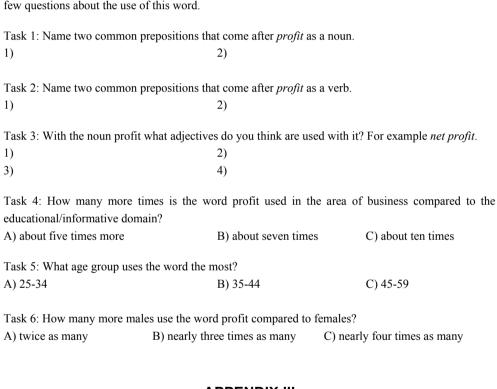
#### APPENDIX II

#### Week Two in Class

Objective: This guide is designed to help students make relatively simple investigations into words they find interesting using the **British National Corpus (BNC)**. Using this handout should enable you to analyze

specific words and evaluate how they are used (and not used) and what other words are normally used with them. For example, with a noun we can look at which verbs commonly go with this noun (and what prepositions are to be found with these verbs). With a noun we can also look at what other nouns come before and after it, not just one time but many times, even a hundred or more. We can also find out in what context (domain) the word is used, by whom it is used and to whom. This handout is divided into two sections; micro and macro so to say. At first we will look at the word and the immediate words around it, "the company the word keeps". In the same way that we can tell quite a lot about someone from the friends that a person has, we can also tell quite a lot about a specific word from the company of other words around it. Then we can consider the wider social field that our chosen word belongs to.

This introduction aims to examine the word *profit*. Before we use the BNC, please try and answer a few questions about the use of this word.



## **APPENDIX III**

Questionnaire about using The **British National Corpus** in language learning for Students of Business Administration. This information will be treated confidentially. No information will be given to any of the University's authorities nor will your name ever be mentioned in any research work connected with this questionnaire. The aim of this questionnaire is to gain an overview of how students react to working with corpus, in our case the BNC.

Section	<b>A</b> •	Background	information

1.	First Name:	Second

Name:	_				
2. Gender: Male				F	emale
3. Native language(s):					
4. How long have you ever spe	ent in an English-	speaking cour	ntry?		
Years Months	We	eks			
5. What was the grade that you	received in your	last English	examination:		
6. In general, do you like using	g computers? Yes	S		N	o
7. How often do you use the co	omputer for perso	nal purposes	(e.g., email/faceb	ook)?	
Several times a day	_		About once	a day	
Every two/three days	-		Once a wee	ek	
8. How often do you use the co	omputer for study	ing at the TH	Wildau?		
Several times a day	_		About once	a day	
Every two/three days	One	ce a week	Less than o	nce a week _	
9. When you use the computer					
language?					
German/Native language on	ly Eng	glish only	Both		
10. How much of your total co	mputer time is in	your native 1	anguage?		
Almost none About 2	25% Abou	ıt 50%	_ About 75%	All	
11. Do you have internet acces	ss at home? You	es	No		
12. Do you use a dictionary for	r learning/writing	English? Yes	S	N	o
13. If yes, what kind of diction	ary do you use? I	Please check a	all that apply.		
Bilingual (i.e. German/Eng	glish) Mor	nolingual (En	glish-English) _		
Paper dictionaryO	n-line dictionary	(e.g. Leo)	Electronic did	ctionary	
14. Had you heard about corpo	ora before this cla	ss this semest	er? Yes	N	o
15. Had you used corpora befo	ore this class this s	semester? Yes	S	N	о
16. If yes, which corpora did y	ou use?				
17. Do you use the BNC for globalization)?				essay (on f	or example
Writing first time C	hanging B	oth Ne	either		
18. Do you use the corpora on	ly for the Corpus	word or for o	ther activities too	).	
Only Corpus	Cor	pus and other	r activities		
Section B: Using the British	National Corpus				
The following questions conce the response that most closely (1: strongly disagree, 2: disag agree, N: no opinion)	resembles your o	pinion. Please	e be honest.		
19. The searching technique w	as easy to learn				
1 2 3	4	5	6	N	

20.	The practice	e session in	class/Medioth	nek helped for	learning the to	echnique	
	1	2	3	4	5	6	N
21.	It was easy	to do the C	orpus task (the	e one for Moo	dle)		
	1	2	3	4	5	6	N
22.	Doing the C	Corpus task	is a useful exp	perience for la	nguage learnir	ng	
	1	2	3	4	5	6	N
23.	The BNC is	more help	ful than a dicti	onary for my	English writin	g	
	1	2	3	4	5	6	N
24.	After using	the BNC I	use dictionarie	es in a more ci	ritical way		
	1	2	3	4	5	6	N
25.	Using the B	NC is help	ful for learning	g the meaning	of vocabulary		
	1	2	3	4	5	6	N
26.	Using the B	NC is help	ful for learning	g the usage of	phrases		
	1	2	3	4	5	6	N
27.	Using the B	NC is help	ful for learning	g grammar			
	1	2	3	4	5	6	N
28.	Using the B	NC improv	ed my Englisl	n reading skill			
	1	2	3	4	5	6	N
29.	Using the B	NC improv	ed my Englisl	n writing/essay	y skill		
	1	2	3	4	5	6	N
30.	I have some	difficulty i	in using the B	NC because of	f time and effo	ort spent on an	alyzing the data
	1	2	3	4	5	6	N
31.	I have	some diff	iculty in us	sing the BN	NC because	of unfamili	ar vocabulary on
	concordance	e/collocation	n output				
	1	2	3	4	5	6	N
32.	I have some	difficulty	in using the B	NC because of	f cut-off senter	nces in concor	dance output
	1	2	3	4	5	6	N
33.	I have som understand	e difficulty	in using the	BNC because	e the real texts	s in the BNC	are too difficult to
	1	2	3	4	5	6	N
34.	I often use t	he BNC by	my own choi	ce			
	1	2	3	4	5	6	N
35.	I understand	d the purpos	se of using the	corpus in this	course		
	1	2	3	4	5	6	N
36.	I have some	difficulty i	in using the B	NC because of	f limited acces	s to Internet/c	omputers
	1	2	3	4	5	6	N
37.	When I hav	e problems	in English wr	iting, I search	for help in the	BNC	
	1	2	3	4	5	6	N
38.	When I sear	ch for info	rmation in the	BNC, I usuall	ly get the infor	mation that I	need
	1	2	3	4	5	6	N
39.	The more I	use the BN	C, the more I	enjoy it			
	1	2	3	4	5	6	N

10	T: 11 41-	- DNC for	En aliab		_		
40	. I Will use th		my English w	•			
	1	2	3	4	5	6	N
41	. Learning ab	out the BN	IC has increas	ed my confide	ence about wri	ting in English	1
	1	2	3	4	5	6	N
42	. Overall the	BNC is a v	ery useful res	ource for my l	English writing	g/learning	
	1	2	3	4	5	6	N
43	. Corpus use	is more he	lpful for writin	ng than for rea	ding in Englis	h	
	1	2	3	4	5	6	N
44	. Corpus use	is more he	lpful for readi	ng than for wr	iting in Englis	h	
	1	2	3	4	5	6	N
45	. The corpus	should be	introduced to a	all English cla	ss at third leve	el (TH level)	
	1	2	3	4	5	6	N
46	. Corpus use	should be	introduced at s	secondary sch	ool level in Ge	rmany	
	1	2	3	4	5	6	N
47	. I would rec	commend t	he BNC to st	udents from o	ther degree co	ourses who wa	ant to improve their
	English						
	1	2	3	4	5	6	N
48	. My attitude	e to using	the corpus has	s become moi	re positive con	npared to wh	en I first used it (in
	class)						
	1	2	3	4	5	6	N
49	. My attitude class)	e to using	the corpus has	s become mor	e negative con	mpared to wh	en I first used it (in
	1	2	3	4	5	6	N
50	I Harring arress	_	-	•	-	· ·	
30.	_		-	_		-	words has improved
	1	2	3	4	5	6	N
51	. Finally, wh	at I enjoye	d most about u	sing the BNC	was		
52	. What I enjo	yed least a	bout using the	BNC was			
53	. It would be	easier if th	e BNC were .				

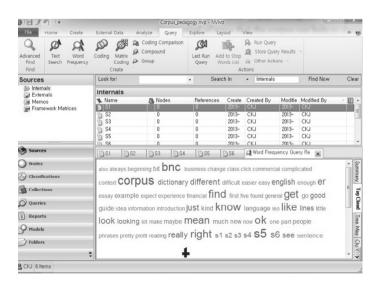
## **APPENDIX IV**

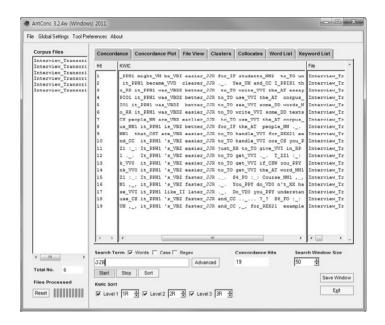
## Interview Questions

- 1. How did you find working with the British National Corpus?
- 2. Were some concordance lines difficult to understand because of the vocabulary?
- 3. Did you find entering the word a little bit complicated?
- 4. Did you experience the interface as easy to use?
- 5. How often did you use the BNC?
- 6. When you were looking up a key word, how many concordance lines did you look at?

- 7. When you were reading the lines, did you find a lot of vocabulary that you didn't know?
- 8. Would you have liked a bit of theory about how the corpus can be used and how it was created?
- 9. Did your attitude change, positive to negative, negative to positive when you were working with the BNC? Did you find it confusing at the beginning when you were on the corpus looking at these lines?
- 10. Should we have spent more time on the BNC?
- 11. I have a question about incidental learning. It really means did you find things on the BNC that you did not expect, did you learn things by clicking on different things? When you were dealing with your word, did you find things that you didn't expect?
- 12. Do you plan to use the corpus in the future?
- 13. If you were writing your Bachelor thesis, would you then use the BNC? Do you think it would be useful?
- 14. How did you use the BNC when writing your essay?
- 15. Do you feel now that it (working with the BNC) has made you more aware of different words and how words work and prepositions?
- 16. Do you use an English-English dictionary or German-English?
- 17. Did (looking at words in the commercial domain of the BNC) give you an insight into financial language?
- 18. What was the most useful aspect in the BNC?
- 19. (When you were writing your essay and looked up words in the BNC) What kind of information did you find?

## APPENDIX V A Screenshot of the QSR NVivo 10





#### A Screenshot of the AntConc 3.2.4w with CLAWS POS C7 tagset

Applicable levels: tertiary education

Key words: corpus pedagogy, corpus use in the classroom, attitudes towards corpora

John O'Donoghue (First author)

Faculty of Business, Computer Science and Law Technical University of Applied Sciences Wildau Bahnhofstrasse

15745 Wildau, Brandenburg, Germany

Tel: +49 (0)3375-508-363

Email: john.odonoghue@th-wildau.de

Received on June 30, 2013 Reviewed on August 28, 2013 Accepted on September 8, 2013 Chae Kwan Jung (Corresponding author) Division of National English Assessment Korea Institute for Curriculum and Evaluation 21-15(Jeongdong), Jeongdong-gil, Jung-gu, Seoul, 100-784, Korea

Tel: (02) 3704-5992, Cell: 010-5007-7481

Email: ckjung@kice.re.kr