RAEL: Revista Electrónica de Lingüística Aplicada
Vol./Núm.: 15/1. Año 2016
Enero-diciembre 2016
Páginas: 150-165
Artículo recibido: 17/11/2016
Artículo aceptado: 17/01/2017
Artículo publicado: 31/01/2017
Url: http://www.aesla.org.es/ojs/index.php/RAEL/article/view/297

Corpora and New Technologies in the Linguistics Classroom: A Pedagogical Use of a Clause Pattern Database

Corpus textuales y nuevas tecnologías en el aula de lingüística: uso pedagógico de una base de datos de patrones léxico-sintácticos

NATALIA JUDITH LASO MARTIN
ELISABET COMELLES PUJADAS
MARÍA LUZ CELAYA VILLANUEVA
MONTSERRAT FORCADELL GUINJOAN
UNIVERSITY OF BARCELONA

Several corpus-based studies have suggested that the observation of real instances of language production is a valuable methodological approach in the description of language. In this respect, the use of computer technology tools has recently become commonplace in the teaching and learning of second and foreign languages. This paper presents the pedagogical use of a clause pattern database which has been designed as a resource to create corpus-based teaching materials to be applied in the linguistics classroom. More specifically, the *Clause Pattern Database* (*CPDB*) provides a wide representative sample of clause patterns in context. The task derived from it and described in this paper illustrates the effectiveness of the *CPDB* as a methodological tool that contributes not only to improving students' understanding of the mechanisms of English syntax but also to integrating new information and communication technologies (NICTs) in the linguistics classroom.

Keywords: *corpus linguistics; lexico-grammatical study; clause pattern database; linguistics classroom; NICTs*

Varios estudios basados en corpus textuales han subrayado la relevancia de la observación de ejemplos de producción lingüística real como herramienta metodológica al servicio de la descripción lingüística. En este sentido, en el campo de la enseñanza y aprendizaje de SL y LE el uso de herramientas computacionales ha aumentado recientemente. Este artículo presenta el uso pedagógico de una base de datos de patrones léxico-sintácticos, la *Clause Pattern Database* (*CPDB*), diseñada como recurso para la creación de materiales didácticos, basados en corpus, para la enseñanza de la lingüística. La *CPDB* proporciona una muestra representativa de patrones léxico-sintácFticos que permite visualizar la complementación de un verbo léxico en un contexto real de aparición. La tarea que presentamos ha sido diseñada a partir de esta base de datos e ilustra la efectividad de la *CPDB* como herramienta metodológica que contribuye no sólo a mejorar la comprensión de los mecanismos que rigen la sintaxis inglesa, sino también a integrar nueva información y NTICs en el aula de lingüística.

Palabras clave: *lingüística de corpus*; *estudio léxico-gramatical; base de datos de patrones léxico-sintácticos; aula de lingüística; NTICs*

1. INTRODUCTION

The use of corpus data in applied linguistics has contributed significantly to both the description and the analysis of language. The observation of real instances of language through the analysis of concordance lines and the inference of various lexical and morphological processes have proven to be of special relevance in first and second language learning and teaching as well as in the teaching of English for Specific Purposes (ESP), among other areas, as will be seen below (see section 2). The present study is framed within these two strands, since it focuses on the study of specific linguistic features by university learners of English linguistics who are non-native speakers of English.

Firstly, the study reported in this paper aims at presenting a database of clause patterns in English, the *Clause Pattern Database* (*CPDB*), which has been specifically designed as a corpus-based resource to create materials to be used in the course of "Descriptive Grammar of English II" (DGE II) as part of the English Studies degree at a Catalan university (see section 3.1). Secondly, the pedagogical applications of this database are explored by highlighting its effectiveness as a learning tool that contributes not only to improving students' understanding of the mechanisms of English syntax but also to integrating new technologies in the linguistics classroom. Finally, some conclusions regarding the appropriateness of integrating corpus-based learning tools in the linguistics classroom are also drawn.

2. REVIEW OF THE LITERATURE

Over the past decades, there has been an increasing interest in data-oriented language research. Corpus linguistics is currently seen as a methodological tool, rather than as a separate linguistic discipline (Meyer, 2002), whose object of inquiry is a collection of texts upon which different linguistic analyses can be conducted. For their corpus-based studies, corpus linguists depend on actual instances of language production rather than on made-up examples. Therefore, the observation of naturally occurring data allows for empirical contextualised analyses and may become a valuable methodological approach in linguistic description.

According to Gledhill (2000), we can nowadays distinguish three clearly distinct corpus linguistics schools: a) corpus-based studies in computational linguistics (Butler, 1985; Oakes, 1998); b) corpus-based research on corpus tagging, parsing and information retrieval (Biber et al., 1998; Danielsson, 2003); and c) corpus analysis to investigate issues that have different applications in modern linguistics, such as language acquisition and language learning (Altenberg & Tapper, 1998; Granger, 1998; Tono, 2009; CASE project[1]), contrastive and translation studies (Baker, 1996; Bernardini & Zanettin, 2000; Xiao & Yue, 2009), discourse analysis (Csmoay, 2005), historical linguistics (Nevalainen & Raumolin-Brunberg, 1996; Rissanen, 2000; Fanego, 2012), and the creation of lexicographical works (Cowie, 1999; Moon, 2007; Hanks, 2009) and grammar reference books (Biber et al. 1999; Huddleston & Pullum, 2002).

McEnery's and Hardie's (2012) account of the most recent methodological and theoretical innovations in the discipline explores how corpus linguistics has developed lately and outlines its role in different types of linguistic analysis, such as language variation, language change, discourse analysis, cognitive approaches to linguistics, metaphor identification in corpora and psycholinguistics, among others. In this sense, we think it can be claimed that computer technologies have contributed significantly to both the description and

[1] http://www.uni-saarland.de/fachrichtung/anglistik/staff/adjunct-faculty/engling2/case.html (learner corpus).

the analysis of language (Sinclair, 2004a; Conrad, 2005; Greaves & Warren, 2007; Granger & Meunier, 2008; Baker 2009; Bennet, 2010). As advocated by McEnery et al. (2006: 7): "the key to using corpus data is to find the balance between the use of corpus data and the use of one's intuition".

In his discussion of the role of corpus studies in linguistics, Meyer (2002) argues that there is no sharp dividing line between descriptive linguists, whose linguistic studies are often based on corpus observation, and theoretical linguists, who are more concerned with building up linguistic theories. Moreover, Meyer states that, despite the differences, both types of linguists can benefit from each other.

Retrieving language data from a corpus by means of corpus-query systems has paved the way for users to access and consider genuine utterances which would have hardly been detected without the assistance of corpus-based methods. As a consequence, the use of computer-assisted tools is now becoming more common in the teaching and learning of second/foreign languages through the analysis of authentic language in use, as can be seen in studies that range from the seminal work by Granger (1994) to more recent ones such as those of Sinclair (2004b), Laso and Giménez (2007), Aijmer (2009), Reppen (2009), Campoy et al. (2010), Granger (2013) and Urzua (2015). As discussed by Carter (1998), learners of English as a Foreign Language (EFL) have some intuition on target language use; however, as Tsui (2004) points out, they still need to be guided in the acquisition of linguistic forms. Nevertheless, we agree with Bernardini (2004) in that the use of corpus linguistics techniques in the language classroom has enabled learners to become language researchers and, as a result, they have gained more autonomy as language learners. As Bernardini (2004: 28) claims, learners' motivation increases and the teacher's role changes and, hence, the environment becomes "supportive" and "non-authoritarian".

Hence, the value of corpora for language pedagogy and, in particular, for the creation of corpus-based materials in the language classroom seems to be unquestionable. Given the success of Huddleston and Pullum's (2002) corpus-based *Cambridge Grammar of the English Language*, the main concern now is how to best integrate corpus linguistics techniques in the classroom and how to overcome possible problems, as in the case of the studies briefly reviewed below.

In the area of second language teaching, the use of learner corpora, as opposed to native corpora, may also facilitate the learning of certain structures in the classroom in a faster and easier way than the use of traditional methods, since learners feel that the data they need to tackle is similar to their own language production. MacDonald et al. (2011), for instance, make use of corpora to carry out error annotation on texts written by Spanish learners of English at two Spanish universities and at different levels of proficiency. Their study is framed in the TREACLE project, which aims at analysing learners' proficiency so as to help in curriculum design. In the same line, the study by Fuster-Márquez and Clavel-Arroitia (2011) proposes a model to integrate corpus linguistics in the English language class in a Spanish university degree with the aim of addressing problems in academic writing that the set textbooks cannot explain. The task consists in working with an academic article from which students have to select and discuss several linguistic features while checking both the British Academic Written English Corpus (BAWE) and the Michigan Corpus of Upper-level Student Papers (MICUSP), two corpora which collect learner essays of academic writing.

A similar methodology has been applied in ESP studies. Cortes (2006), for instance, used corpora to teach certain lexical bundles (e.g., *at the beginning of, by the end of, on the basis of*) to native speakers of English in a History class at university. In the same line, Gilquin et al. (2007) state that the few materials that exist to improve writing skills based on corpora come in many cases from native corpora.

The findings of the studies reviewed above together with our conviction that corpus-based approaches to language teaching provide a highly beneficial context for language learners led us to further investigate the potential of such tools in the linguistics classroom; more specifically, in the teaching and learning of clause patterns in the descriptive grammar classroom at university level. We also have to take into account that some scholars (see Seidlhofer, 2002; Mukherjee, 2004; Breyer, 2009) underline the fact that the influence of applied corpus linguistics on current language teaching practice needs further development. As Fuster-Márquez and Clavel-Arroitia (2011) explain, the two main reasons why a corpus approach in language teaching is still infrequent are that a) corpora may still appear as a difficult tool to be used by learners, and b) few corpora and software are available in the field of language teaching.

Solutions to such issues are likely to appear gradually if teachers and researchers alike collaborate both in the creation and the implementation of new technologies in the linguistics classroom.

3. THE STUDY

3.1 *Context*

The present study took place at the University of Barcelona, more specifically, in the Degree of English Studies and a subject called *Descriptive Grammar of English II* (DGE II). Except for four core subjects in the first year, the degree is taught entirely in English (both by native and non-native English speaking teachers) to (mainly) non-native speakers of English. The groups (usually three) have an average of 60 students each.

DGE II deals with the description of verb complementation in English. This is the second of a series of three modules (DGE I, DGE II and DGE III) that aim at describing the internal workings of English grammar, and are offered over two academic periods. This subject is mostly devoted to the analysis of the English canonical sentence patterns, considering also structural and/or syntactic ambiguity and, thus, the different analyses a given sentence may lend itself to.

3.2 *Participants*

The participants are 3rd-year students of the Degree of English Studies; students from two other degrees -Linguistics, and Modern Languages and Literatures- also take DGE II as an obligatory subject. A total of 148 non-native students participated in the study, grouped in teams of 4-5 students.

3.3 *Procedure*

The group of instructors (N = 9) involved in this study had compiled a corpus of mystery novels, The Whodunnit Corpus, which consists of 5 million tokens. With the aim of making the corpus maximally used, we decided to design a lexicogrammatical database (the *CPDB*) which would not only help store all the data extracted from the corpus, but also create teaching materials for the DGE II course. The *CPDB* has been developed in two stages. In the first stage (Comelles et al., 2010), the examples extracted from the corpus (see below) were analysed and introduced in the database. Secondly, a tree diagram for each of the sentences in the database was produced with the assistance of the *phpSyntax Tree*, a graphical syntax tree generator; then, they were linked to their corresponding database entry.

The resulting tree diagrams were revised thoroughly in order to detect mistakes and ensure coherence. This process triggered discussion among members of the research group about controversial analyses, which later proved useful with students in class.

Third, students were asked to reproduce a similar process to the one we followed when the first stage of the *CPDB* was developed. As will be seen below, students were asked to create a database in Moodle (http://moodle.org/), an open source Virtual Learning Environment (VLE), as part of their tasks for the subject DGE II.

3. 4 *Instruments*

3.4.1 *The Clause Pattern Database*

The *Clause Pattern Database* (*CPDB*) consists of 714 corpus-based sentences, extracted from a 5 million word corpus of bestselling whodunnits by authors such as D. Brown, M. Crichton, J. Grisham and P. Cornwell. This database allows quick access to lexico-grammatical information about each of the 217 lexemes that can be searched for in the database.

The main reason for creating a corpus of contemporary mystery novels was to make students become familiar with a literary genre which is not covered in their curriculum. The starting point to exploit the corpus and create the database was the valency of lexical verbs and clause patterns so as to meet the needs of the learners of the course of DGE II. A selection of 217 prototypical verbs, previously used in class, and illustrative of the 5 canonical patterns established by Huddleston and Pullum (2002), was used to perform several searches by means of *Wordsmith Tools*. These 5 canonical patterns account for: a) SV - intransitive sentences (e.g. *Christmas came one morning*); b) SVCs - complex intransitive sentences (e.g., *She had been secretly proud of her calm, controlled behaviour*); c) SVO - monotransitive sentences (e.g., *His instructions were confusing the participants*); d) SVOCo - complex transitive sentences (e.g., *The afternoon has made the children quiet for a while*); and SVOO – ditransitive sentences (e.g., *Grandmother gave visitors sugar cake and hot coffee*). In the database, verbs performing more than one pattern are also covered (e.g., *make*) and several examples are provided (e.g., SVO – *They won't be making any distance calls*, SVOCo – *They made him their slave*, SVOO – *They made him some soup*). Other syntactic phenomena are also covered such as structural ambiguity and valency alternations, which are an important part of the contents in the DGEII course.

As stated in the procedure section above, in the first stage of the creation of the *CPDB*, the examples from the corpus were analysed and introduced in the database. This analysis implied the identification of the clause pattern, the main lexical verb of the sentence, and the valency. As shown in Figure 1, all this information was introduced by means of several fields which contain: a) the sentence under study (e.g., *He had tasted something very bitter*); b) the lemma of the main verb (*taste*); c) the valency, which includes the grammatical category of the verbal complements (NP-NP); and d) the clause pattern (SVO).

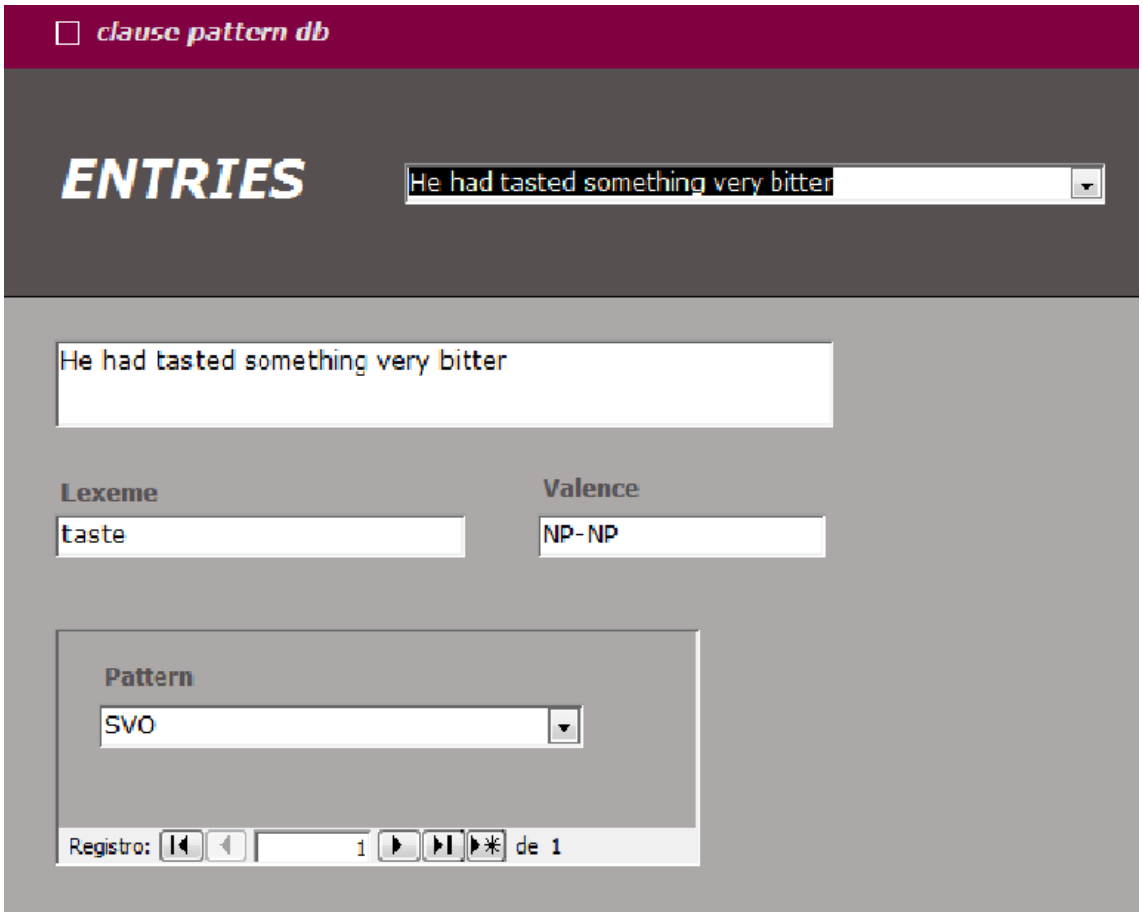

Figure 1. ***Screenshot from the Clause Pattern database***

In the second stage, two tools were used to automatise the production of the tree diagrams and to link them to their corresponding database entry: a) the Charniak parser (Charniak & Johnson, 2005), which uses a regularised MaxEntreranker to select the best parse from the 50 best parses yielded by a generative parsing model; and b) the *phpSyntax Tree* (http://www.ironcreek.net/phpsyntaxtree), which is a free online tree-diagram generator. We opted for these tools because they are publicly available online and they have a user-friendly and quite intuitive interface. The Charniak parser provided us with the bracketed constituent analysis of each sentence; however, some modifications had to be made to the parser output for two reasons. First, being an automatic parser, it may provide wrong analyses. For instance, the sequence *mum walloping* in the sentence *I remember mum walloping him with the broomstick* was analysed as a NP-Object instead of as a sequence of NP-Object (*mum*) + a non-finite VP-X-Complement (*walloping him with the broomstick*). Another example is the analysis obtained for the sentence *Tomorrow Patrick will drive some of them to the airport*, where the sequence *Tomorrow Patrick* was analysed as a proper noun (i.e., name + surname) working as the subject of the sentence instead of an adverb (*Tomorrow*) followed by the proper noun (*Patrick*). Second, the parser output had to be slightly modified and adapted to the type of analysis and labels used in our linguistics classroom. Some of the modifications performed were the adoption of the labels *Verbal* and *Nominal*, which were not provided by the parser. The label *Verbal* was used to allow for a separate node to distinguish complements from adjuncts; likewise, the tag *Nominal* was also introduced to distinguish pre and post modifiers from the head of the NP. In addition, the online version used to analyse sentences only provided part of speech (POS) tags; thus, in order to make the analyses obtained from the parser more similar to the ones shown in class, phrasal tags (NP, ADJP, PP, ADVP and VP) were also added.

Hence, after evaluating all the troublesome areas presented by the sentences in the corpus, a final set of criteria was established for the design of the basic templates for tree diagramming. These criteria aimed to cover the most relevant linguistic issues posed by trees so that all dependencies in the sentences under analysis would be illustrated in a coherent and simplified way. Naturally, the adequacy of these criteria had to be congruent with both the approach and level of analysis of sentences established in the syllabus of our subject, DGE II. In this subject, trees are only considered and used as a tool to graphically represent the dependencies among the various components that sentences might present, including both complements and adjuncts. Dealing with subtle theoretical issues is not the objective of this subject; rather, it focuses on making students aware of the syntactic functions that a specific string may perform at a specific point of an utterance. Thus, tree diagramming helps represent different displays in dependencies triggered, for instance, by structural ambiguity (e.g., *Charlie might have engineered an affair with her*).

Consensus on the type and level of analysis was reached taking into account the relevant issues that our students need to grasp according to the guidelines of DGE II, disregarding the unnecessary complexities posed by idiosyncratic examples. The criteria adopted are the following:

a) To keep structure visually simple, only dependencies among main dependents (arguments and adjuncts) were represented, which implies underspecifying dependencies at lower levels, both phrasal and sub-clausal.
b) To keep a parallelism in the optionality of adjuncts and pre and post modifiers in NPs, for instance, the intermediate nodes *Verbal* and *Nominal* were adopted.
c) To simplify sentence structure, compound sentences were broken down into simple ones; hence, question tags, for instance, were disregarded.
d) To avoid complicating the structure of the diagrams, the negative particle *not* was not accounted for; instead, it was kept as part of the verbal auxiliary in both contracted forms and full forms. Similarly, the negative assertive compound *no one* was taken as a single unit.
e) To avoid dealing with non-basic types of sentences, only declarative sentences were represented.

Once the output was modified, the resulting bracket analyses were transformed into tree diagrams by means of the *phpSyntax Tree* tool (see Figure 2).This process was quite straightforward, since the bracket analyses provided by the Charniak parser were directly copied and pasted into the Phrase box in the phpsyntaxtree interface.

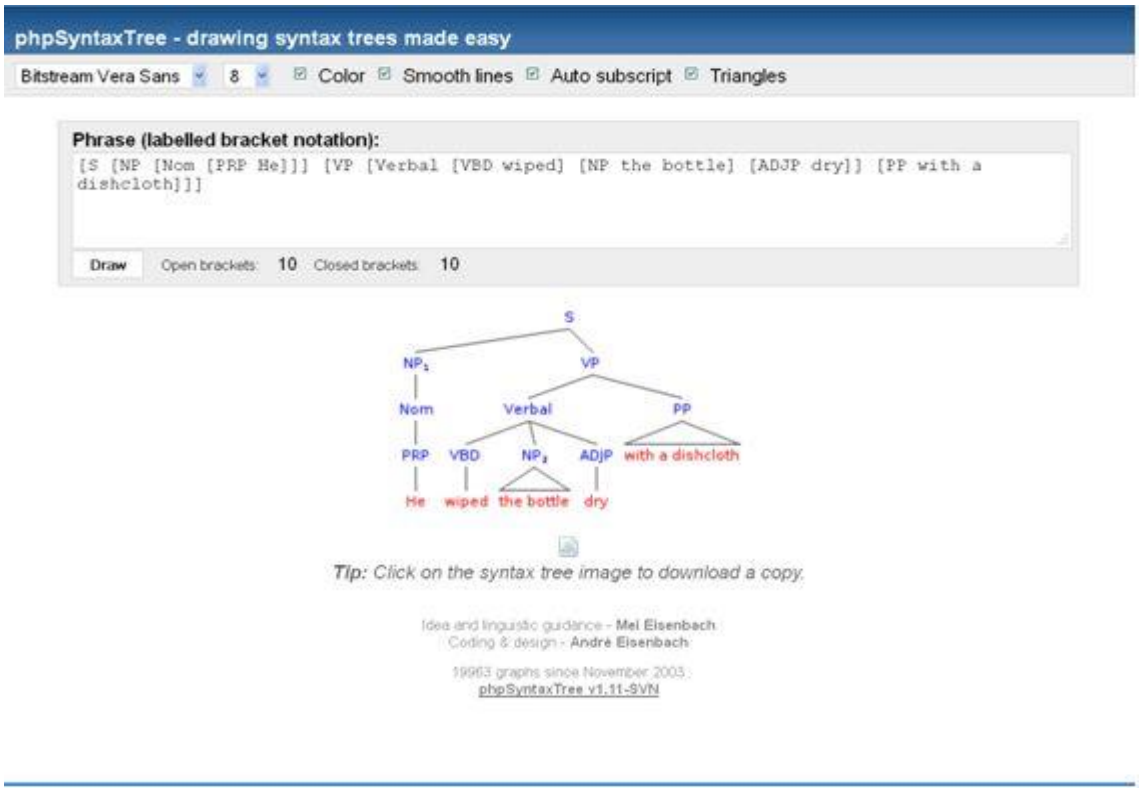

Figure 2. ***Screenshot from the phpsyntaxtree***

After reaching consensus about problematic cases, each tree diagram was linked to its corresponding register in the database. As shown in Figure 3, a complete entry contains the sentence under analysis (e.g., *His quick recognitions made him frantically impatient of deliberate judgements*), the lexeme of the main verb (*make*), the valency of the main verb (NP-NP-AdjP), the clause pattern (SVOCo) and the tree diagram.

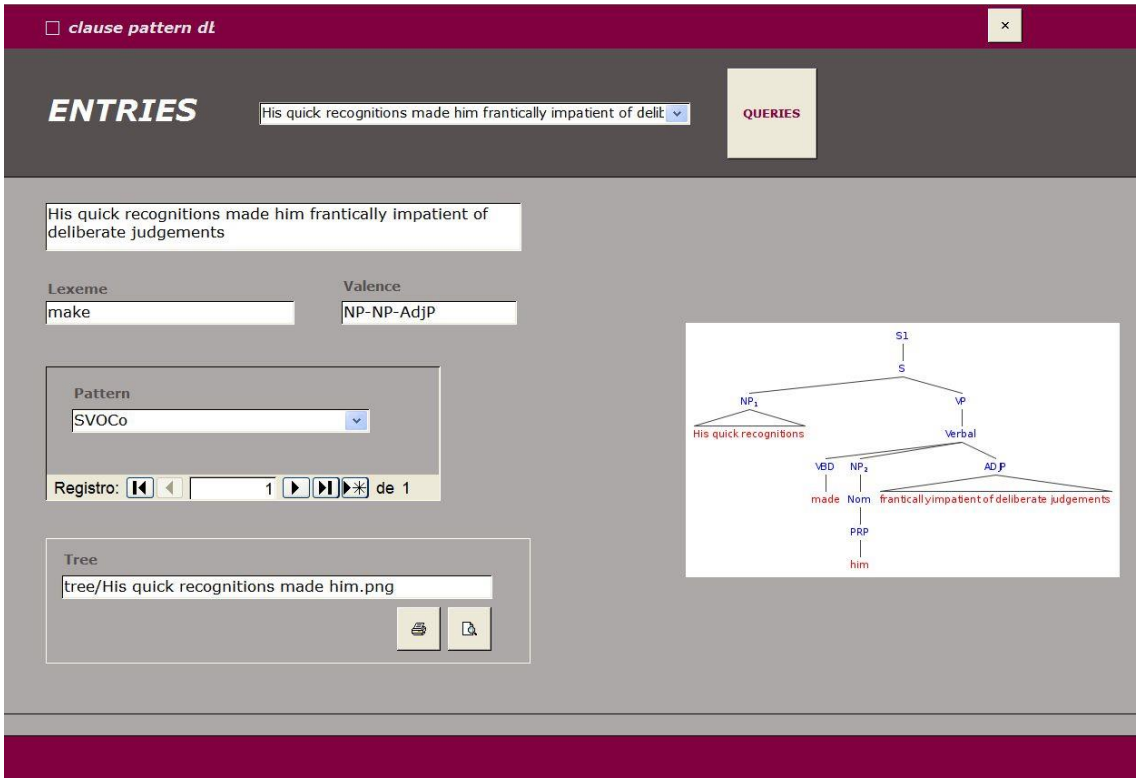

Figure 3. ***Output of a register from the Clause Pattern DB***

3.4.2 Satisfaction questionnaire

Students were invited to answer a questionnaire on their satisfaction with the task, using a five-point Likert scale (i.e., (1) very dissatisfied – (5) very satisfied). The questionnaire contained 10 questions regarding:

- subject knowledge
- analytical and critical abilities
- language analysis techniques
- teacher's instructions
- task as effective link between theory and practice
- interactive an interpersonal skills
- practical task
- user-friendliness of Moodle databases
- teacher as Moodle facilitator
- overall satisfaction

3.5 Task description

As mentioned above (section 3.1), structural ambiguity is an invaluable tool to show very clearly and easily that, when considered in isolation, the same linguistic string (i.e., the same sentence) may yield two different interpretations depending on the different syntactic analyses it may present, which is usually directly translated into different clause patterns. Furthermore, this structural "flexibility" is sometimes correlated with different "part of speech" analyses of the strings (phrases) involved in each of the different readings. This task is therefore thought to enhance students' understanding of verb subcategorisation in English depending on each of the linguistic contexts in which verbs may be found.

Thus, students were asked to reproduce a similar process to the one we followed when the first stage of the *CPDB* was developed; that is, the creation of a database (now in Moodle). Students in teams were required to discuss the issues that needed to be addressed both when completing the task and when having to react to the teacher's feedback to amend any pitfalls their analysis might present.

The task consists of three steps: a) the identification of a variety of sentences that present the verb patterns described in the task guidelines, b) the description of the selected sentences according to the information required by the online database, and c) the drawing of tree diagrams for each of these sentences.

The first step requires students to choose a text from any genre (e.g., a novel, a newspaper article or a children's tale) avoiding those that are addressed to English language learners and therefore do not contain spontaneous data, since they are simplified or tailored to suit specific learning needs. From these texts, they have to extract a total of eight sentences: six of them should illustrate any of the clause patterns studied in class excluding those that contain clausal complements; one sentence should present structural ambiguity (e.g., *I kissed the boy in the bathroom*); finally, one last sentence should allow for dative/benefactive alternation (e.g., *Susan bought her husband a bunch of roses*). Clausal complements are excluded to make the task more challenging since in texts of authentic production it proves quite difficult to find sentences that are free of clausal arguments. In addition, this difficulty makes students aware of the many complexities of real language, where sentences have not been boiled down to their simple structures to facilitate both the teaching and grasping of the basic building blocks of sentence structure.

The second step requires the use of the database created in Moodle. By using this tool, students are given a further opportunity to use new technologies as part of their learning process. In this second step, students have to introduce specific information about their selected sentences in the database, as previously devised by their teachers. This involves some consideration of the different levels of sentence analysis. The database entries that students have to fill in are a replica of those in the *CPDB*, with the addition of two more fields: FURTHER COMMENTS for any comments on the sentence analysed that students might find necessary, and GROUP NAME for the group number previously assigned to each student (see Figure 4).

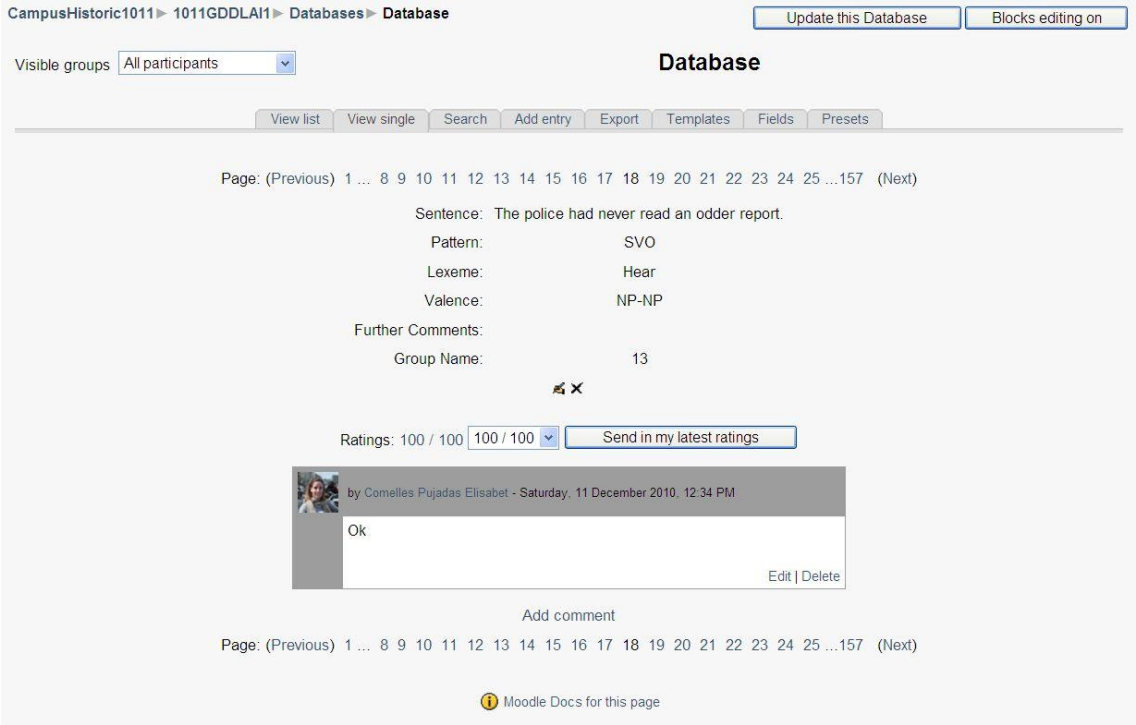

Figure 4. ***Students' Database entry***

Finally, the third step involves the drawing of tree diagrams to illustrate the underlying hierarchy among the constituents in the sentences. When having to graphically represent the relationships among the components in sentence structure, students are made to reflect not only on constituency itself but also on constituent dependency. This may reveal their weak points in handling syntactic complexities, which might involve structural ambiguity as well. For instance, students may find it difficult to distinguish between obligatory and non-obligatory dependents (e.g., *They danced the tango alone* vs. *They danced the night away*), or between (pre or post) modifiers of the same head and embedded (pre or post) modification (e.g., *She bought the book of poems by Penguin* vs. *She bought the book of poems by Kingsley Amis)*, or when telling sentence adjuncts and VP adjuncts apart (e.g., *She kissed him (,) naturally*, where adequate punctuation of sentences is the only clue for adequate analysis), or identifying structural ambiguity in a given sentence (e.g., *He took her flowers*). Thus, tree drawing is used as a tool to develop students' understanding of real data intricacies by graphically illustrating the relationship among dependents.

In the third step, the teachers evaluate a first version of both the information introduced in the database and the hand-drawn tree diagrams. Feedback including comments on the sentences which needed to be revised was provided to students online. Each of the sentences analysed was given a rating (0 for sentences with major errors, 50 for sentences with minor errors, and 100 for correct sentences). A second version of the task with all the required

amendments was to be submitted by a given deadline (see Figure 5). This procedure allowed students to interact and share their knowledge, as well as to consult their teacher when agreement among members could not be reached. The tree diagrams were also to be revised and corrected for a second submission according to the teachers' feedback on the parts that had not been correctly represented in the first version. A grade was finally awarded to the team, taking into account both versions of the task and the progress made throughout the revision process.

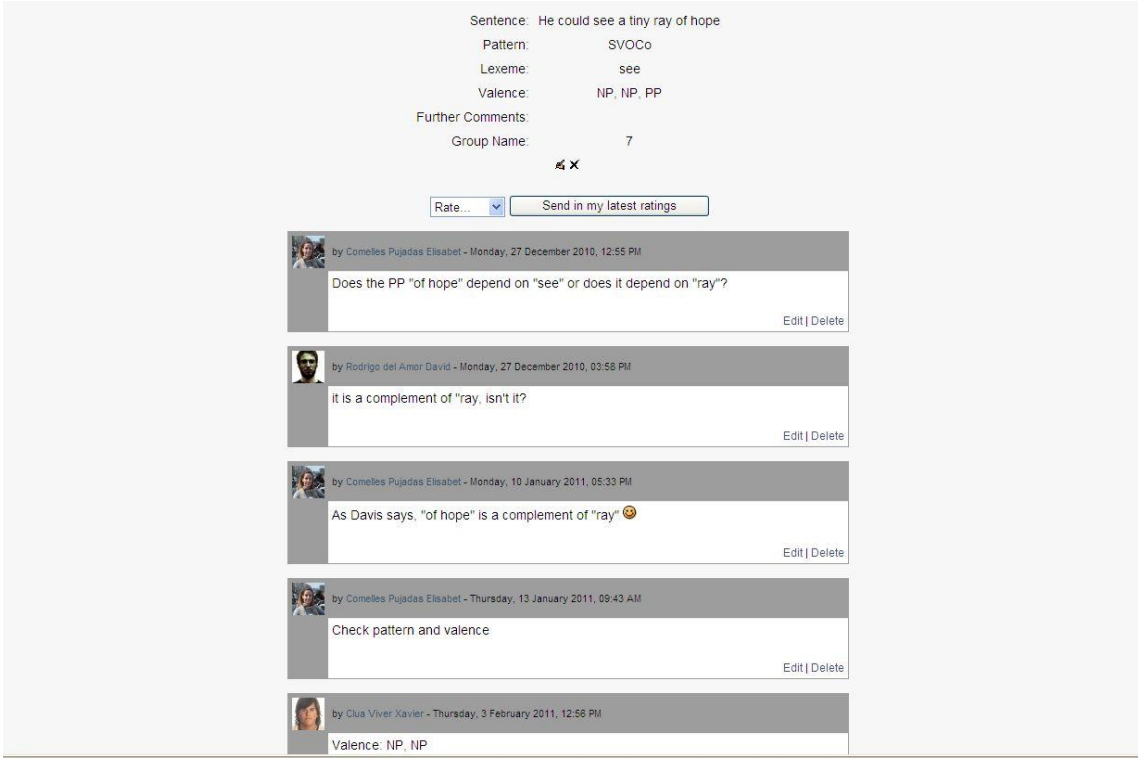

Figure 5. ***Student's Database entry (Teacher's feedback)***

4. RESULTS AND DISCUSSION

The results obtained after completion of the task were highly satisfactory; after the teachers' assessment of the tasks delivered by students, 90% of the student teams arrived at correct versions by the end of the process. In line with Meyer (2002), thus, we can claim that the use of corpus linguistics can be seen as a way of doing linguistics and not as a separate field of study. The task described in the present study has contributed to developing subject contents in two ways; first, we believe that it has helped students engage with authentic data, since they had to apply the concepts dealt with in class for the identification of clause patterns; second, it has made students apply their subject knowledge by using the appropriate terminology so as to fill in the different fields of the database.

A subsequent but no less important aspect of the task lies in its usefulness to help students see in retrospect the importance of being able to identify clause patterns when having to deal with the contents of subjects such as DGE III, which is a higher level course and presupposes knowledge of English sentence structure. The insight acquired in dissecting utterances into smaller units (constituents), which are then analysed at different levels (i.e., syntactic functions, type of phrases or sentences realising those functions), will be applied when studying the structures underlying some of the English sentences described both in DGEII and in other subjects. In DGEII students thoroughly work on the identification of

structural ambiguity, thus identifying constituents and their corresponding syntactic functions is crucial in order to deal with ambiguity. In addition, recognizing the different clause patterns that verbs may enter also help them better understand and be aware of valency alternations (i.e. active-passive, dative-ditransitive and ergative), which directly affects their knowledge of English. The knowledge obtained in DGEII has a direct consequence in other related subjects (i.e., DGEIII). When analysing English marked structures (i.e., those that do not present the SVO basic arrangement), students need to be able to clearly identify the canonical clause patterns in sentences in general in order to tell marked structures apart and identify each of them adequately. Students are ready to understand the structures underlying marked sentences and the grammatical restrictions that shape them only when they can analyse sentences by breaking their linguistic material into different constituents, which are then assigned specific syntactic functions according to the specific slots they occupy in the structure (i.e., presenting different clause patterns); moreover, it is then that they realise that those functions may be performed by any type of linguistic sequence, phrasal and clausal as well (described by the valency field in the database).

There are many other instances to illustrate why asking students to systematically identify the kind of information required by the database will prove helpful in understanding constructions other than the basic canonical ones in English. From our experience as teachers of DGE II, we know that students easily realise that it is necessary to be able to identify the building blocks of the language before they may address more grammatically complex instances. Hence, pedagogically, the usefulness of the task extends the boundaries of the kind of linguistic insight required in DGE II, since it is devoted to enhancing students' ability to identify the basics of English sentence structure, upon which more complex phenomena are based.

The teachers' beliefs about the positive impact of the task were further corroborated by the students' answers to a satisfaction questionnaire. The feedback obtained from 93 questionnaires was remarkably positive, given that the lowest mean score obtained was 3.45 out of 5 (cf. Comelles et al., 2012). We interpret such results to be in line with Bernardini (2004) in that learners become motivated when using NICTs in the classroom.

High scores were specially obtained in questions regarding the teacher as a Moodle facilitator (4.12/5) and teacher's instructions (4.30/5). The effectiveness of the task to link theory and practice (3.78/5), and its usefulness of the task in helping students develop their subject knowledge (3.70/5) also proved to be relevant. Furthermore, the tool was rated as more satisfactory among students who were less familiar with the use of databases and language analysis techniques (i.e., the lower the students' degree of familiarity with the online tool, the higher their degree of satisfaction with the task), which establishes a strong connection between the two instruments: databases and new technological tools[2].

Thus, it must be noted, that the use of a database of clause patterns has promoted students' autonomous learning (see Bernardini, 2004) as well as the use of new technologies in a content-based course. The task described in the present study, therefore, seems to have played a very important role in the type of linguistic analysis required in the university degree where the present study took place. As McEnery and Hardie (2012) point out, corpus linguistics can be considered nowadays an essential tool in linguistic analysis.

[2] For further details about the results of the questionnaire, see Comelles et al. 2012.

5. CONCLUSIONS

The task presented here has proven to be an effective tool for furthering the understanding of the description of the canonical sentence in English. The fact that the *CPDB* shows the various clause patterns that a given lexical verb may appear in allows for a thorough analysis of several lexical and syntactic phenomena dealt with in the linguistics classroom (such as ergative, dative and passive alternation) as well as syntactic and lexical ambiguity. This database is a key tool for the production of new corpus-based teaching materials which provide students with a wide representative sample of clause patterns in context.

The task involves the analysis of authentic English production through the use of an online tool that allows for personalised and specific feedback on the submitted task. Having to identify specific patterns from authentic language sources, students are confronted with challenges posed by uncontrolled language that will not be found in the subject-tailored paradigms addressed in class. Furthermore, online tools such as the database developed for this task promote collaborative learning in the classroom and encourage students to interact and exchange their knowledge of English grammar. Dealing with real data makes students aware of language complexity. Likewise, critical thinking is fostered by promoting students' ability to analyse the intricacies of authentic production, which leads to a better understanding of the inner mechanisms of English syntax in general.

We hope that once publicly available, the *CPDB* will be especially useful in class. From the teacher's point of view, the *CPDB* is a useful and friendly tool to provide students with valuable input, as it contains all relevant information (i.e. sentence under analysis, syntactic analysis and tree diagram) in a single entry. From the students' point of view, the *CPDB* helps them become easily aware of how the linguistic analysis is captured in the tree and helps them better understand the relations established between sentence constituents.

In the future, as a further development of the task, learners will be asked to make use of publicly available syntactic parsers for the syntactic analyses of the examples introduced in the database, as well as to work with the *phpsyntaxtree* generator so that they become more familiar with NICT tools in the linguistics classroom. In addition, it seems worth exploring how to optimise and apply the *CPDB* in other university courses. Likewise, the spread of the *CPDB* may contribute to the design of new materials and the assessment of its effectiveness. Corpus-based teaching tools will not only bring to the forefront the challenges that syntactic and semantic analyses pose for the language observer but also free learners from the limits of student-tailored examples by means of controlled but interactive procedures.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Programa d'Innovació Docent of the University of Barcelona (2013PID-UB/003) and the *Grup de Recerca en Lexicologia i Lingüística de Corpus* group (GReLiC) at the University of Barcelona. We also thank the various reviewers of this article for their thoughtful remarks and suggestions for improvement. Any mistakes remain our own.

REFERENCES

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

Altenberg, B. & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on Computer* (pp. 80-93). London and New York: Addison Wesley.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan Sager* (pp. 175-186). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Baker, P. (Ed.). (2009). *Contemporary Corpus Linguistics*. London and New York: Continuum.

Bennet, G. R. (2010). *Using CORPORA in the Language Learning Classroom: Corpus Linguistics for Teachers*. Michigan: University of Michigan Press.

Bernardini, S. & Zanettin, F. (Eds.). (2000). *Il Corpora della Didattica della Traduzione – Corpus Use and Learning to Translate*. Bologna: CLUEB.

Bernardini, S. (2004). Corpora in the classroom. An overview and some reflections on future development. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp.15-38). Amsterdam: John Benjamins Publishing Company.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22, 153-172.

Butler, C. (1985). *Computers in Linguistics*. Oxford: Basil Blackwell.

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

Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, 52, 43-56.

Charniak, E. & Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. *Proceedings of the 43rd annual meeting of the association for computational linguistics* (pp. 173-180). Michigan: Association for Computational Linguistics.

Comelles, E., Laso, N. J., Verdaguer, I. & Gimenez, E. (2010). Clause pattern DB: A corpus-based tool. In I. Moskowich-Spiegel Fandiño, B. Crespo García, I. Lareo Martín & P.Lojo Sandino (Eds.), *Language Windowing through Corpora. Visualización del lenguaje a través de corpus* (pp. 215-234). Coruña: Universidad da Coruña.

Comelles, E., Laso, N. J., Forcadell, M., Castaño, E., Feijóo, S. & Verdaguer, I. (2012). Using online databases in the linguistics classroom: dealing with clause patterns. *Computer Assisted Language Learning*, 1-13.

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 393-409). Lawrence Erlbaum Associates: New York.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.

Cowie, A. P. (1999). *English Dictionaries for Foreign Learners*. Oxford: Oxford University Press.

Csmoay, E. (2005). Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education,* 15, 243-274.

Danielsson, P. (2003). Automatic extraction of meaningful units from corpora: A corpus-driven approach using the word stroke. *International Journal of Corpus Linguistics,* 8, 109-127.

Fanego, T. (2012). COLMOBAENG: A corpus of late modern British and American English prose. In N. Vázquez (Ed.), *Creation and Use of Historical English Corpora in Spain,* (pp. 101-117). Newcastle upon Tyne: Cambridge Scholars Publishing.

Fuster-Márquez, M. & Clavel-Arroitia, B. (2011). Implementing an academic corpus in the English language classroom in tertiary education. In Mª L. Carrió Pastor & M. A. Candel Mora (Eds.), *Actas del 3 congreso internacional de lingüística de corpus, tecnologías de la información y las comunicaciones* (pp. 695-704). Valencia: Universidad Politécnica de Valencia.

Gilquin, G., Granger, S. & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes,* 6, 319-335.

Gledhill, C. (2000). *Collocations in Science Writing*. Tübingen: Gunter Narr.

Granger, S. (1994). The learner corpus: A revolution in applied linguistics. *English Today,* 10, 25-33.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (145-160). Oxford: Oxford University Press.

Granger, S. (2013). *Learner English on Computer*. New York: Routledge.

Granger, S. & Meunier, F. (Eds.). (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Greaves, C. & Warren, M. (2007). Concgramming: A computer driven approach to learning the phraseology of English. *ReCALL,* 19 (3), 287-306.

Hanks, P. (2009). The impact of corpora on dictionaries. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 214-236). London and New York: Continuum.

Huddleston, R. & Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Laso, N. J. & Giménez, E. (2007). Bridging the gap between corpus research and language teaching. In C. Periñán (Ed.), *Revisiting Language Learning Resources* (pp. 49-64). Newcastle: Cambridge Scholars Publishing.

MacDonald, P., Murcia, S., Boquera, M., Botella, A., Cardona, L., García, R., Mediero, E., O'Donnell, M., Robles, A. & Stuart, K. (2011). Error coding in the TREACLE Project. In Mª. L. Carrió Pastor & M. A. Candel Mora (Eds.), *Actas del 3 congreso internacional de lingüística de corpus, tecnologías de la información y las comunicaciones* (pp. 725-740). Valencia: Universidad Politécnica de Valencia.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based Language Studies. An Advanced Resource Book*. London and New York: Routledge.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Meyer, C. F. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Moon, R. (2007). Sinclair, lexicography, and the Cobuild project. The application of theory. *International Journal of Corpus Linguistics*, 12(2), 159-181.

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). Amsterdam and New York: Rodopi.

Nevalainen, T. & Ramoulin-Brunberg, H. (Eds.). (1996). *Sociolinguistics and Language History. Studies based on the Corpus of Early English Correspondence*. Amsterdam and Atlanta: Rodopi.

Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

phpSyntaxTree. Retrieved from: http://www.ironcreek.net/phpsyntaxtree/ (last visited: 06 April 2016).

Reppen, R. (2009). English language teaching and corpus linguistics: Lessons from the American national corpus. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 204-213). London and New York: Continuum.

Rissanen, M. (2000). The world of English historical corpora: from Caedmon to the computer age. *Journal of English Linguistics*, 28(1), 7-20.

Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 213-234). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Sinclair, J. M. (2004a). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Sinclair, J. M. (Ed.). (2004b). *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing Company.

Tono, Y. (2009). Integrating Learner Corpus Analysis into a Probabilistic Model of Second Language Acquisition. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 184-203). London and New York: Continuum.

Tsui, A. B. M. (2004). What Teachers have always wanted to know - and how corpora can help. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 39-61). Amsterdam: John Benjamins Publishing Company.

Urzua, A. (2015). Corpora, context, and language teachers. In V. Cortes & E. Csomay (Eds.), *Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber* (pp. 99-122). Amsterdam: John Benjamins Publishing Company.

Xiao, R. & Yue, M. (2009). Using corpora in translation studies: The state of the art. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 237-261). London and New York: Continuum.