



Compiling and analyzing a tagged learner corpus: a corpus-based study of adjective uses

Cristina Castillo Rodríguez¹; José María Díaz Lage²; Beatriz Rubio Martínez³

Recibido: 13 de noviembre de 2017 / Aceptado: 18 de febrero de 2020

Abstract. A learner corpus (LC) is widely known as a rich source of information regarding the use of expressions and the errors made by students in their productions. In fact, we, as teachers, can profit from the compilation of their tasks so as to analyze in detail their way of writing. However, the mere compilation of texts does not guarantee a successful exploitation, as more steps than saving texts must be involved in the whole process. Therefore, it seems essential to follow a protocolized methodology of compilation. In this paper we propose five phases for compiling a LC containing texts from the spontaneous written productions from undergraduate and postgraduate students. The outcomes thrown with the LC exploitation will reveal the errors in students' productions regarding the use of plural, comparative and superlative in adjectives and also other fails detected in the tagging phase, most of which are due to students' misuses.

Keywords: English as a foreign language; applied linguistics; learner corpora; error analysis

[es] Compilación y análisis de un corpus de estudiantes etiquetado: un estudio basado en corpus de usos de adjetivos

Resumen. El corpus de estudiantes (CE) se considera una fuente rica de información en lo que respecta al uso de expresiones y errores cometidos por los estudiantes en sus producciones. De hecho, como docentes, podemos beneficiarnos de la compilación de sus tareas para analizar en profundidad su forma de escribir. Sin embargo, la simple compilación de textos no garantiza que la explotación del corpus tenga éxito. Por ello, diseñar una metodología protocolizada se convierte en una tarea esencial en la compilación de un corpus. En este artículo se proponen cinco fases para compilar un CE de producciones escritas espontáneas de estudiantes de grado y posgrado. Entre los resultados de la explotación del CE se encuentran los errores de los estudiantes en cuanto al uso de plural, comparativo y superlativo en adjetivos, así como otros fallos detectados en la fase de etiquetado, que se deben, en su mayoría, a errores de los propios estudiantes.

Palabras clave: inglés como lengua extranjera; lingüística aplicada; corpus de estudiantes; análisis de errores.

How to cite: Castillo Rodríguez, Cristina; Díaz Lage, José María y Beatriz Rubio Martínez (2020). Compiling and analyzing a tagged learner corpus: a corpus-based study of adjective uses. En *Círculo de Lingüística Aplicada a la Comunicación* 81, 115-136, <http://dx.doi.org/10.5209/CLAC.67932>

¹ University of Málaga. E-mail address: cristina.castillo@uma.es

² Universidad Internacional de La Rioja. E-mail address: josem.diaz@unir.net

³ University College Dublin. E-mail address: beatriz.rubio-martinez@ucd.ie

Contents. 1. Introduction. 2. Theoretical framework. 2.1. First attempts in the use of learner corpora. 2.2. Learner corpora in error analysis. 3. Materials and method. 3.1. The learner corpus ENTECOR. 3.2. Participants. 3.3. Tools. 3.4. Procedure. 4. Results. 4.1. Use of plural in adjectives. 4.1.1. Instances of errors (*s_JJ). 4.2.2. Instances of errors (*s_NNS). 4.2. Use of comparatives. 4.3. Use of superlatives. 5. Conclusions. References

1. Introduction

As a branch of applied linguistics, corpus linguistics has conferred upon language researchers around the world—and, indeed, any user interested in languages—a great tool for observing and analyzing words in their context. Even though linguistics, understood in a wider sense, has sometimes been considered too theoretical and as an integrative or complementary part of other disciplines (Corder, 1992), corpus linguistics seems to have a more concrete and applied nature. McEnery, Xiao and Tono (2006) stated that corpus linguistics “should be considered as a methodology with a wide range of applications across many areas and theories of linguistics” (p. 8). In other words, the different uses and applications that corpus linguistics can offer to, among many other disciplines, teaching practice and methodology are certainly numerous. Besides, there is a very wide range of linguistic aspects that can be analyzed and worked on by using corpora in language teaching (Nesselhauf, 2004).

Under the umbrella of the conceptualization of what we know today as learner corpora research (LCR), we present a study in which we analyze the written (and spontaneous) production of students at university (undergraduate and postgraduate levels). As our general objective, we want to offer some easy steps for every foreign language (FL) and second language (L2) teacher in every educative level interested in detecting quantitatively the errors made by students without having to annotate all errors manually. For doing so, in this paper we pay special attention at the use of plural in adjectives in English, as well as the use of comparatives and superlatives. On the other hand, this study has another specific aim: to offer some strategies so as to search for those errors with a raw corpus and with a POS-tagged corpus by means of a corpus management program. Besides, we will analyze the fails detected when automatically tagging with another program.

2. Theoretical framework

2.1. First attempts in the use of learner corpora

Corpus linguistics started to be of special interest to English language teachers around 1987 with the publication of Collins COBUILD English Language Dictionary, which is considered the first corpus-based dictionary specifically created for English students (Gabrielatos, 2005). Since that moment, there have been many studies which focused on how to use corpora to provide more efficient teaching materials and methods (see for instance Biber, Conrad, & Reppen, 1998; Kennedy 1998; McEnery & Wilson, 2001; Meyer, 2002; Tognini-Bonelli, 2001).

However, as Gabrielatos (2005) explains, and despite the general positive criticism about the possibilities that corpora could offer to the improvement of English teaching, some researchers expressed their doubts about how useful corpora could be to describe language use or the possibility of L1 corpora being too intimidating for foreign learners (Gabbrielli, 1998; Widdowson, 1991).

This study is focused on the analysis of a learner corpus (LC), a much newer and less common concept than L1 corpora, which has been used to study child

language since the 1970s (Lozano & Mendikoetxea, 2013). Unlike L1 corpora and studies on second language acquisition (SLA) —which has traditionally nurtured from learners productions (Granger, Gilquin & Meunier, 2015)—, the LC as a field of research dates back to the late 1980s (Granger, 2004), but the compilation of LC started to become popular from the 1990s onwards (Nesselhauf, 2004), corpora of learning writing being the most common type. Despite their relatively recent appearance and use, LCR has already produced a vast amount of studies, above all because of two main advantages (Granger et al, 2015): i) LC are quite large and texts or samples are collected from many students; ii) they can be managed and analyzed with electronic tools, so the cost is rather less than if samples are analyzed manually by humans.

2.2. Learner corpora in error analysis

One of the most accepted definitions for LC (or interlanguage, IL, or L2 corpora, according to Granger, 2003) is “electronic collections of authentic foreign or second language data” (Granger, 2003, p. 465).

The use of LC, which represents the language produced by FL learners (Leech, 1998, cited in Pravec, 2002), has been considered a rich source for studying the authentic language use made by students of FL, which allows us to detect not only errors as far as grammar, spelling, and style aspects are concerned, but also their linguistic behaviour and the tendency of use of expressions.

Although the scientific nature of corpus linguistics seems to have been sufficiently proved, based on the large amount of data with which it deals, the effectiveness of Error Analysis (EA) has been considerably criticized (see Ellis, 2003; and Els, Bongaerts, Extra, Os, & Janssen-van Dieten, 1984). Among many others, some of the most significant criticisms include the fact that learners will try to avoid structures that they think could be erroneous and use only those that they consider correct, or that “analysis is done on a static text, a sort of language photo taken at a certain moment under certain circumstances” (Castillejos López, 2009, p. 676) and therefore, this does not explain the students’ learning progress. EA, however, is an indispensable step towards the efficient exploitation of LC and it should be understood as the first stage in the effective prevention or correction of errors which would make the language learning experience a much more significant process. The different stages in EA as described by Corbet (1974) and summarized by Ellis (1994) are:

- (1) Collection of a sample of learner language.
- (2) Identification of errors.
- (3) Description of errors.
- (4) Explanation of errors.
- (5) Evaluation of errors.

Granger et al (2015) also state that the methodology used for EA using LC is known as computer-aided error analysis. Annotating errors in the texts contained in a LC becomes an essential task in order to detect all the errors made by learners. However, EA is frequently done manually, so it is a very time-consuming task for FL or L2 teachers.

The project in which this study originates consists of detecting errors made by Spanish-mother-tongue students of English as a foreign language (EFL) so as to

determine some innovative methodologies with the aim of reducing or totally eliminating those errors.

Even though the project itself also intends to undertake an error annotation of all the texts compiled for the corpus, it is interesting to take into account other strategies and corpus-based methods so as to look for errors and linguistic behaviour of our students without having to annotate the whole corpus.

Annotation of texts will undoubtedly provide more benefits because all errors will be marked; but, it is true that, we, as teachers, are mindful of the great number of tasks teaching implies and the amount of hours that text annotation requires, as it is a very time-consuming task.

This paper, therefore, is aimed at providing some simple steps for compiling an electronic LC. Besides, it is also aimed at suggesting some strategies so as to search for errors made by students and fails detected when tagging automatically the LC with a tagger system. These simple guidelines will enable FL or L2 teachers to approach the linguistic behaviour of students and fails of the system that one can encounter when using automatic tagging systems in a LC which is supposed to present grammar, spelling, and style errors.

In our case, the analysis of our students' use of the language, and more specifically, the use of plural in adjectives, as well as comparative and superlative adjectives in their written productions, will serve us as a starting point to create, in our future teaching practice, new ways of correcting and preventing students' errors and recurring erroneous grammatical patterns.

3. Materials and method

3.1. The learner corpus ENTECOR

The LC object of our study is called ENTECOR. This macro corpus is intended to comprise several subcorpora compiled from subjects in English belonging to different degrees and master degrees at the International University of La Rioja (UNIR).

Up to now, ENTECOR is composed of two subcorpora, namely, TICOR and SECOR. The first subcorpus contains two components: ICT (component 1), which holds texts collected from the undergraduate subject called 'ICT tools applied to the learning of English language' (Degree in Early Years Education); and TIC (component 2) that gathers texts from the subject 'Tecnologías de la información aplicadas al aprendizaje de lengua inglesa' (Degree in Primary-School Education).

SECOR, on the other hand, contains only one component (TRAINCOR) and is composed by texts collected from the subject called 'Complementos de formación' (Master in Secondary Education).

The three subjects mentioned above are taught entirely in English, and tasks or activities performed by students are also produced in this FL, including the communication via the forum of the subjects.

Our team's next step will be the compilation of more subcorpora, but the current structure of ENTECOR is the following:

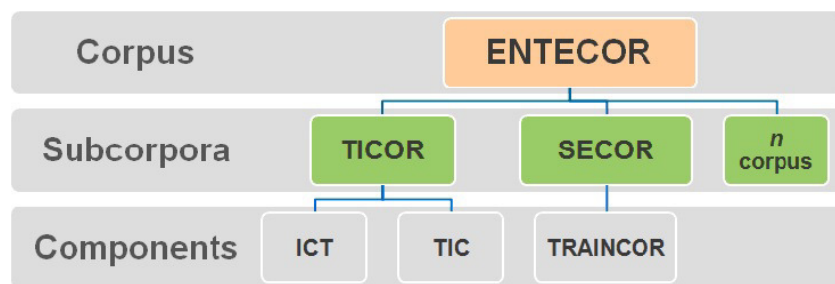


Figure 1. Current structure of ENTECCOR corpus

ENTECCOR contains 527,099 tokens and 13,148 types up to now. These figures correspond to the following:

- TICOR subcorpus contains 107,072 tokens and 4,821 types in the case of the ICT component, and 317,759 tokens and 9,320 types in the TIC component.
- TRAINCOR subcorpus contains 102,268 tokens and 6,380 types.

Before describing participants, tools, and the procedure followed to compile the corpus for its subsequent exploitation with a concordance tool in order to detect errors (students) and fails in tagging (software), it is essential to list some design criteria (Díaz Negrillo & Thompson, 2013; Gilquin, 2015):

- a) medium: our corpus contains written texts extracted from some compulsory forums proposed for subjects in English. The overall project, however, is also aimed at compiling oral productions of students for further research.
- b) annotation: unannotated or raw corpus vs. annotated corpus. The project implies the future error annotation of the texts, but for this study the analyses have been made with the raw corpus and POS annotated corpus.
- c) target language: the corpus is monolingual, that is, it only contains texts (produced by learners) in English.
- d) data conditions: as far as the degree of naturalness is concerned, the forums are proposed with a specific topic in mind. Besides, teachers in charge of the subjects contribute first and provide some guidelines for participating there. However, we should not forget that it is a forum, so some spontaneous written participation is implied. Therefore, we can state that these are semi-natural cases of participation.
- e) time: semi-longitudinal or cross-sectional.

Other variables have been proposed by other authors, but as Granger (2004) claims: “one must admit that... there are so many variables that influence learner output that one cannot realistically expect ready-made LC to contain all variables for which one may want to control” (p. 126).

3.2. Participants

As far as the TICOR subcorpus is concerned, a total of 666 students and their productions have been compiled: i) 155 students have participated with their

contributions for feeding the ICT component since 2013-2014; and ii) 511 students have been gathered for the TIC component since 2011-2012. The reason why the starting date and the number of students differ from one component to another is that the English mention in the subject of the ICT component was officially approved later than in the case of the TIC component. Consequently, the number of students, and also their contributions in academic forums in the TIC component, is higher than the number of students and contributions of the ICT component. With regard to contributions, 231 have been compiled and recorded for the ICT component and 763 for the TIC component, as more than one compulsory academic forum was proposed in both subjects.

For the SECOR subcorpus, on the other hand, a total of 120 files from 120 students have been compiled for the study of the TRAINCOR component. In this component from SECOR subcorpus, only one forum was proposed in the subject selected for collecting data for the corpus. Besides, this subject was proposed in English after the subject of the ICT component.

3.3. Tools

This study requires the use of three tools: EncodeAnt, TagAnt, and AntConc.

EncodeAnt 1.2.0 is a freeware tool, designed and launched by Laurence Anthony, enabling users to detect and convert character encodings. The use of this program is necessary as the tagging program requires a specific format in order to tag successfully all the texts of a given corpus (Anthony, 2016).

TagAnt 1.2.0 is a part-of-speech (POS) tagger, also known as grammatical tagger, that allows the users to tag their corpus (Anthony, 2015b). This program is built on TreeTagger, which means that for commercial purposes, even though TagAnt is freeware, it is necessary to purchase the commercial license of TreeTagger, as indicated in its website (Anthony, 2015b).

AntConc 3.4.4. is a freeware set of tools for analyzing and exploiting corpus of texts, as some of the tools gathered are Concordance, Word List, Collocates, among others (Anthony, 2015a).

3.4. Procedure

Our compilation methodology for successfully exploiting and analyzing ENTECOR consists of five simple steps, as described below.

Phase 1. Download and codification phase

The first phase in every research task in corpus practice is the collection of data to be managed in a suitable software program. Downloading files from forums entails saving them in our computers and labelling them with a suitable and univocal code. Some examples of code are the following: 01TOENICTA, 01TOENICTB, 01TOENTICA, 01TOENTICB, 01TOENTRAIN. 01 means Student1; TO, source text; EN, English; ICT corresponds to the subject from Early Years Education; TIC, to the Primary-school Education degree, and TRAIN, to the Master of Secondary Education. Letter A stands for the first compulsory forum, and letter B for the second forum of every subject containing two compulsory forums during the academic year.

Phase 2. Cleaning and saving phase

Each file contains all the contributions posted by one student in a concrete forum. However, for separating every contribution we have added some indicators of

beginning, such as [Contribution1], [Contribution2], [Contribution3], and so forth. Besides, in the cleaning phase personal information revealing names and surnames of students has been eliminated so as to preserve their privacy. Some examples of these cases are [Studentname], or [Studentnameandsurnames], among other instances (see Castillo & Díaz, 2015, for more information about this phase). After the contributions have been cleaned, all the files are saved in .txt format —which is the format used in most corpus management programs.

Phase 3. Registering code phase

This third phase involves the registration of all the instances in an excel document with the aim of providing all the information about every document. Hence the use of some suitable columns indicating relevant information: code, number of student, real name and surname/s of student, total number of contributions, undergraduate degree/master degree, subject, academic year, and total number of words of the text (Castillo & Díaz, 2015).

Phase 4. Tagging phase

Tagging both subcorpora (TICOR and SECOR) is the fourth phase to compile our LC in order to start managing, observing and analyzing words and expressions in context. This phase requires the use of two different programs, that is, EncodeAnt 1.2.0, and TagAnt 1.2.0. EncodeAnt 1.2.0 has been used for converting files into UTF-8. Once the files have been processed, the program automatically creates a folder with all the .txt files transformed into UTF-8. TagAnt 1.2.0 has been used for tagging the files of our subcorpora: files contained in every UTF-8 folder have been opened into the main screen of TagAnt once it is launched. Then, the start button is pressed and the files appear on the right as tagged files. Tagged .txt documents are saved in the same UTF-8 folder, in such a way that we have at our disposal the original file, encoded in UTF-8, as well as the tagged file.

Phase 5. Exploitation phase.

This phase corresponds to the extraction of results, either errors or mistakes found in the subcorpora or the most frequent expressions used by students in this type of spontaneous written productions. This phase implies the use of the program AntConc 3.4.4 and some concrete word detection is described in detail in the following section.

The following figure illustrates graphically the phases explained before:

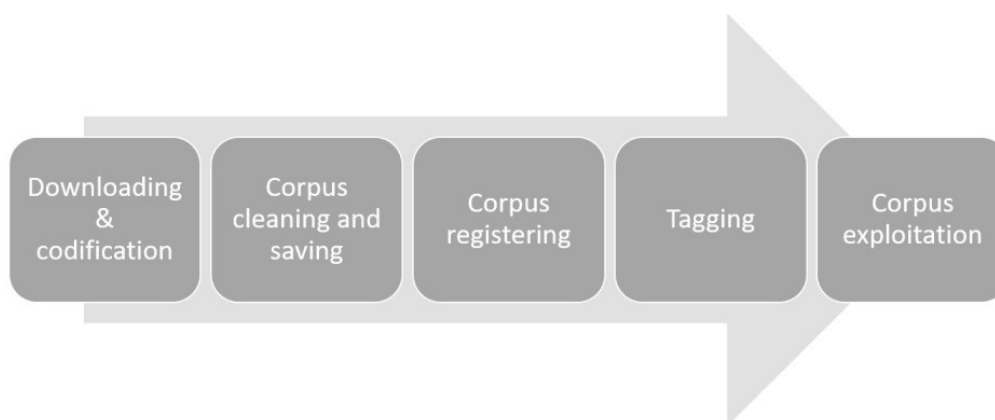


Figure 2. Compilation methodology for ENTECOR learner corpus

4. Results

Exploiting a tagged corpus has some more advantages over using a non-tagged one. But choosing one procedure over the other will always depend on the purposes researchers pursue in each case. In the case with which we are concerned, for detecting some errors made by students and also the fails of the tagger software, we have to search for the corresponding tag in every component from the two subcorpora.

The grammatical tagger TagAnt uses the tagset from TreeTagger, that is, a total of 58 tags; a reduced set of tags if we compare it with the set proposed by CLAWS in its tagsets (132 for C1 tagset or 160 for C6, UCREL, n.d.). However, for the purposes we intend to achieve in this study, the use of this tagger is more than enough for the subsequent exploitation in AntConc.

Once every component has been opened in AntConc we can start observing the use of adjectives in students' contributions.

Concretely, the study with which we are concerned in this paper is divided into three areas:

- a) The use of plural in adjectives.
- b) The use of comparatives.
- c) The use of superlatives.

4.1. Use of plural in adjectives

The detection of errors in the use of adjectives in their plural form has implied searching by means of two tags with the following strings: *s_JJ and *s_NNS.

4.1.1. Instances of errors (*s_JJ)

Firstly, the POS tag JJ —the tag for adjectives according to the tagset indicated before— has been used for the search. Concretely, we have typed in the search box of AntConc the following string: *s_JJ. This means that we search for any word ending in -s and tagged, or recognised by the software, as adjective. The results thrown by AntConc for that string can be observed in Table 1.

Component	Hits	Plural adjectives	Other errors	Fails
ICT	161	0	1	11
TIC	527	0	9	14
TRAINCOR	228	0	6	3

Table 1. Detection of errors and fails with the tag *s_JJ

The number of instances found per word in every component is indicated below in brackets. Besides, errors made by students have been typed in bold, while fails from the system have been highlighted in grey.

As seen in table 1 above, in ICT we have found several adjectives (161 hits) with -s ending, but they are not properly in their plural form. We refer to instances such as: advantageous (1), autonomous (5), cautious (2), conscious (6), continuous (5), countless (1), curious (1), dangerous (2), **ELS** (1), endless (4), enormous (2), famous (2), gorgeous (1), marvelous (1), needless (2), nervous (1), numerous (1),

obvious (5), **ours** (11), previous (12), propitious (1), restless (1), serious (3), spontaneous (1), useless (3), various (3), veracious (1), and 81 cases of saxon genitive and other contractions such as *it's*.

In the TIC component, we have obtained 527 hits. However, as in the case of the ICT component, none of them are errors in the use of plural. The examples thrown by the software have been the following ones: advantageous (4), **andless** (1), **ans** (1), arduous (1), asynchronous (1), autonomous (11), **backwards** (1), cautious (1), [whole/in-]class (1), **conciencious** (2), conscious (9), conscius (1), continuous (22), countless (2), **creativis** (1), curious (6), dangerous (11), disastrous (1), **ELS** (4), endless (20), enormous (11), **express** (2), fabulous (2), famous (5), gorgeous (2), harmonious (2), heterogeneous (1), hilarious (2), homogeneous (2), instantaneous (2), laborious (1), marvellous (2), marvelous (2), monotonous (2), needless (8), nervous (6), **nevertheless** (1), numerous (7), obvious (10), **ours** (8), precious (6), previous (53), priceless (1), religious (1), restless (1), ridiculous (1), rigorous (1), selfless (1), senseless (1), serious (7), simultaneous (1), **synonymous** (1), tedious (1), textless (1), **this** (1), timeless (1), ubiquitous (1), useless (11), various (20), and 236 cases of saxon genitive and other contractions such as *it's*.

In TRAINCOR we have found several adjectives (228 hits) with *-s* ending, but not in their plural form. Here, we refer to examples such as: ambitious (1), **Brits** (1), cautious (1), colourless (1), **conciuous** (1), conscious (2), countless (1), curious (6), dangerous (1), delicious (2), endless (6), **Engliss** (1), enormous (4), erroneous (1), fabulous (2), famous (56), **foquiss** (1), furious (1), generous (1), glorious (1), gross (1), heterogeneous (1), hilarious (6), humorous (2), indigenous (1), **jalous** (1), mass (1), mysterious (1), needless (1), **nowadays** (1), obvious (1), **ours** (2), prestigious (3), previous (21), religious (1), Scots (1), serious (2), tasteless (1), tedious (1), tremendous (1), useless (1), various (4), voluptuous (1), **vous** (1), and 80 cases of saxon genitive and other contractions such as *it's*.

As observed in previous lists, some fails have occurred when tagging words, and also some other errors (different from the use of plural in adjectives) made by students have been detected.

ICT component instances (*s_JJ)

Errors (students)

- Spelling: ELS instead of using ESL, which is the correct initialism.

Fails (software)

- ours: tagged as adjective because all the cases of *ours* found in this component are followed by a noun: *ours children* (1); *ours day* (1); *ours English class* (3); *ours first concern* (1); *ours materials* (1); *ours own materials* (1); *ours students* (3). Therefore, the system might have understood that it was an adjective, even though the correct tag for that word would have been PP\$, which is the POS tag for possessive pronouns. However, even in this case, the system shows some errors as it has only tagged possessive adjectives and only the possessive pronoun *yours* in the three cases found in the component. On the other hand, the rest of possessive pronouns have been tagged with different tags, as: *mine* (2 instances as nouns, _NN), (1 instance as personal pronoun, _PP); *theirs* (1 case as PP) —no cases of *hers* and *his* or *its* (as possessive pronouns) have been found in the component.

TIC component instances (*s_JJ)

Errors (students)

- Spelling: andless (endless), ans (and), conciencious (conscious), creativis (creativity), ELS (ESL).

Fails (software)

- backwards. This is exclusively an adverb, the adjectival form being obsolete; the tag should therefore have been RB. The system has tagged it as an adjective probably because it is used in coordination with *predesigned*; and, indeed, the student seems to intend to use it as an adjective.

- express. This ought to have been tagged as a verb in both instances. However, the system has tagged it as an adjective because both students using it do so incorrectly: in the first instance, as *we can't express ourselves*. Here the incorrect form *ourselves* has been tagged as a plural noun (NNS) rather than as a personal pronoun (PP). As a consequence, *express* has been interpreted as an adjective. In the second instance, we find *they don't know how express their feelings*. Here the problem has been the absence of *to* in the phrase, which has made the system interpret *express* as an adjective rather than as a verb in its base form (VV).

- nevertheless. The system's misrecognition of *nevertheless* as an adjective rather than as an adverb (RB) does not seem to follow any kind of discernible logic. However, we suspect that it might have been, apparently, because the sequence tagged has been *methodologies..Nevertheless* without spacing between those words.

- ours. There are several different phenomena here. In the case of *the resources are not ours, are not made...* (1), the possessive pronoun (PP\$) is correctly used. The system's fail stems probably from the fact that there is no subject in the following phrase. In the case of *in ours classrooms* (1), *ours pupils* (3) and *ours students* (1), as in the ICT component above, the form *ours* is used before a plural noun instead of *our*, which must have been interpreted by the system as an adjective rather than as a possessive pronoun. As for *too many ours in front of...* (1), the system here has misidentified *ours* because it obviously cannot recognize that this is a misspelling of the student: this should of course have been *hours* and have been tagged as a plural noun (NNS). Finally, in *...competence is greater than ours inthis field* (1), the student uses the possessive pronoun correctly, but the fact that *inthis* (an obvious typo) is interpreted by the system as a plural noun (NNS) determines its identification of *ours* as an adjective.

- synonymous. In this case, *this is not a synonymous that all information...* (1), the system has indeed identified the form correctly as an adjective, although the student probably intended to use a noun (*synonym*), even if this would have also been ungrammatical.

- this. As was the case of *nevertheless*, it is difficult to understand why the system would tag it as an adjective rather than as a demonstrative pronoun, as the rest of *this* found in the corpus have been tagged with DT.

TRAINCOR component instances (*s_JJ)

Errors (students)

- Spelling: Brits (British), concious (conscious), Engliss (English), focuss (focus), jalous (jealous).
- Grammar: Not-so-often-nowadays.

Fails (software)

- ours: tagging this word as adjective has been a fail of the system. Besides, the two cases found in the component shows that it was perfectly used as a possessive pronoun, as after those words we have found a comma or an adverb. As in previous cases, it should have been tagged as PP\$, such as in *yours*. However, once again, the system has also missed the use of this tag in possessive pronouns, as we have observed in the different cases like *mine* (7 cases of NN, 2 cases of PP); *theirs* (PP). No cases of *hers* and *his* or *its* (as possessive pronouns) have been found either.
- vous: French word, that is, a foreign word. It should have been tagged as FW, but not as an adjective. Another fail of the system.

4.1.2. Instances of errors (*s_NNS)

Another tag and another string have been used for searching for errors in the use of plural in adjectives. Let us see all the results and analyse them in detail when typing the string **s_NNS*.

As seen before, the system has also shown its own fails when tagging instances. On the other hand, the examples thrown by the system with JJ tag offer no cases of plural adjectives in the three components. Therefore, another string, and another tag, used for detecting this type of errors has been: **s_NNS*. This means all cases of nouns in plural form (with *s* ending) so as to observe if the system has detected any adjectives as nouns. In fact, there have been some interesting outcomes in this respect. Let us see, again per component, all the examples thrown by the system.

In the ICT component, we have found the following cases: *attractives* (1), *availables* (2), *compatibles* (2), *complets* (1), *creatives* (2), *diferents* (2), *differents* (2), *favourites* (1), *foreigners* (1), *negatives* (1), *originals* (1), *spectaculars* (1), *traditionals* (1), *youngers* (1).

Apart from that list of mistakes, other spelling mistakes can be highlighted: *complets*, *diferents*.

In the TIC component, the cases of plural adjectives thrown by the software have been: *attractives* (1), *creatives* (3), *diferents* (4), *differents* (8), *educatives* (2), *falses* (1), *favorites* (5), *favourites* (10), *futures* (1), *importants* (3), *inconvenients* (1), *interestings* (1), *negatives* (3), *originals* (3), *responsables* (1), *similars* (2), *simples* (2), *socials* (1), *specifics* (2), *traditionals* (1), *vacants* (1), *usuals* (1), *yougers* (1).

In the TIC component's list of mistakes related to the use of adjectives in plural, other mistakes have been detected, such as spelling mistakes in the words: *attractives, diferents, responsables, youngers*.

In the TRAINCOR component, only one error has been detected: negatives (1).

The following table (table 2) shows the number of errors in the use of plural of adjectives by using the string *_NNS.

Component	Adjectives in plural
ICT	19
TIC	58
TRAINCOR	1

Table 2. Use of plural in adjectives detected with the string *_NNS

4.2. Use of comparatives

Errors in the use of comparatives have been searched by typing the string and tag *_JJR, which means searching for any word tagged as a comparative adjective form.

In ICT we have found several comparative adjectives (387 hits). We refer to examples such as: **better** (84), bigger (1), cheaper (2), closer (2), easier (41), faster (7), foreigner (1), funnier (3), further (3), greater (4), happier (1), harder (5), higher (3), less (16), lighter (1), longer (1), lower (1), **more** (179), narrower (1), older (15), safer (1), slower (1), smaller (7), weaker (2), wider (1), **worse** (2), younger (2).

In TIC we have found the following comparative adjectives (1133 hits): **better** (227), bigger (6), cheaper (5), clearer (2), closer (7), deeper (5), earlier (2), easier (100), **easter** (1), elder (1), faster (23), fewer (2), friendlier (1), funnier (16), further (6), greater (20), happier (4), **harder** (3), higher (21), larger (4), later (1), less (44), longer (4), lower (3), luckier (1), **moder** (1), **more** (507), newer (2), nicer (1), older (52), quicker (3), richer (5), shorter (1), simpler (1), slower (3), smaller (3), smoother (1), stronger (1), weaker (8), wider (6), worse (5), younger (24).

In TRAINCOR the examples detected as comparative adjectives (237 hits) have been: **better** (26), broader (2), classier (1), clearer (2), closer (4), cooler (2), darker (1), deeper (5), drier (1), earlier (2), easier (5), freer (1), friendlier (2), further (5), greater (1), harder (4), healthier (1), heavier (1), higher (12), larger (1), later (1), less (9), lesser (1), longer (1), lower (7), luckier (1), **more** (110), **northerner** (1), older (10), quicker (1), richer (1), slower (1), smaller (2), stronger (2), wetter (1), wider (3), **worse** (2), younger (4).

Let us analyse the cases highlighted before in the following paragraphs.

ICT component instances

Errors (students)

- more: *more younger*. This is a fairly frequent case in learners of EFL. The comparative adjective (*younger*) is used, but it is incorrectly used in conjunction with the adverb *more*.

Fails (software)

- better. There are a great many instances where this form has been incorrectly tagged as a comparative adjective when it should have been identified as a comparative adverb (RBR), as in: (...) *get to know each other better*; (...) *they will learn better...*; (...) *we can find activities which better fit*; (...) *things that work better in each class*; *Are they better or worse prepared*; (...) *they could learn more and better*; *We have to [...] explain better*; (...) *revise to know it better*; (...) *students [...] can learn better than those that are not motivated at all*; (...) *we only have to know which fits better with our goals*; (...) *they know almost better than she?*; (...) *they use them better than us*; *they know how to use them better than me*. Therefore, the total number of errors in the case of *better* has been 13 out of 84 hits found with this tag.

- more. In many instances where *more* is combined with an adjective (or, in the case of *more clearly*, with an adverb) to form a comparative, *more* should have been tagged as an adverb. This is the case of occurrences such as *more attractive classes*, *more comfortable*, *better materials or more completed one* (where *completed* is a mistake for *complete*), *more complex*, *more confident*, *more creative*, *more different*, *more effective*, *more encouraging*, *more entertaining* (wrongly tagged as a noun) and *attractive*, *more innovative*, *more interesting* (2), *more mature*, *more motivating* (where *motivating* is tagged as a gerund rather than as an adjective), *more persistent*, *more technological*, *more traditional*.

There are eight instances of *more and more*; of these, three are cases of a phrase which is itself adverbial and where in all probability *more* should have been tagged as an adverb: (...) *they wanted to learn more and more*; (...) *the advantages are more and more*; (...) *we rely more and more on ICTs*.

There are seven cases in which *more* is used in combination with *than*: in those instances, *more* should have been tagged as an adverb: (...) *more than a few seconds*; (...) *no more than three months*; (...) *advantages are more than disadvantages*; (...) *more than one Mum/Dad*; (...) *children get to use more than this [sic] technologies*; (...) *english [sic] subject is not more than a tool*; (...) *ICTs have become more than our aid*.

Finally, there are other instances where *more* should have been tagged as an adverb: (...) *children would improve their English more if they'd just watch TV in original version*; (...) *it will help us to enjoy more in our classes*; (...) *I am more in the idea of creating our own material*; (...) *what we play more is...*; (...) *games in which they interact more*; (...) *using more my own imagination*; (...) *we rely more and more on ICTs*; (...) *which source he employs and relies more on*; (...) *teachers needs [sic] to prepare more our classes*; (...) *what do they like more*; (...) *during a lesson you use different activities and exercises and more with little children*; (...) *so it interacted more with students*; (...) *the authoring tool that fits more with your decision*; (...) *I [couldn't/cannot] agree more with...*(3); (...) *I will work more with my designed materials*.

Therefore, the total amount of errors of 178 hits in *more* has been 45.

- worse: amongst the results we find one instance of *worse prepared*, where *worse* should have been tagged as an adverb.

On the other hand, we have typed *more* in the search box and some more errors made by students have been found for the case of *easy*: *more easy* (3).

TIC component instances

Errors (students)

- *more*: one error found in the sequence *more better*.

Fails (software)

- *better*: as was the case with the ICT component, in many instances this form has been incorrectly tagged as a comparative adjective when it should have been identified as an adverb. The following examples may be mentioned: (...) *to learn better a foreign language*; (...) *we can focus better about what the students need*; (...) *it will fit better*; (...) *they work better as they are motivated*; (...) *I understood still better how it can help*; (...) *I feel a little better*; (...) *what activities work better [sic] in his classes*; (...) *it can keep us better in touch*; (...) *the way they learn better*; (...) *are the students reading better*; (...) *students understand them better*; (...) *in order to remember it better*; (...) *work harder and better*; (...) *students will learn more and better*; (...) *students understand it better*; (...) *what activities work better*; (...) *understand better some abstract concepts*; (...) *to know better some countries*; (...) *children can undersatand [sic] better the concepts*; (...) *they can understand better the things*; (...) *in order to understand better the rules*; (...) *they can learn better the contents*; (...) *one that suits better the needs*; (...) *the student to acquire better the new concepts*; (...) *understanding to communicate better with you all*; (...) *they fit better with your class*. The total number of fails in this case has been 26 out of 227 hits.

- *easter*: the single occurrence of this fail is worth pointing out. Perhaps because the student did not capitalise the word, the system has tagged a proper noun, *Easter*, as a comparative adjective.

- *harder*: in two of the three occurrences of the term, its identification is incorrect as it should have been tagged as an adverb: (...) *willing to work harder and better*; (...) *i need to work harder or better than I did*.

- *moder*: there is only one case of this word, a typo for *modern* which the system misidentifies as a comparative adjective.

- *more*: the most numerous cases of misidentification are those in which *more* is followed by an adjective: *more appropriated* (where the system misidentifies the adverb as a comparative adjective because the student has mistakenly used *appropriated* for *appropriate*); *more attractive*; *more better* (where the student's incorrect structure has also resulted in a system misidentification); *more comfortable* (5 occurrences, including one misspelt as *comfotable*); *more common [sic]*; *more complex*; *more comprehensive* (2 occurrences); *more developed* (where the system has failed to identify *developed* as an adjective); *more difficult* (4 occurrences, including one misspelt as *difficult*); *more effective*; *more elaborate*; *more engaging* (where the system has misidentified *engaging* as a gerund); *more entertained*; *more experienced*; *more flexible*;

more genuine; more important (3 occurrences); *more innovative; more interactive* (4 occurrences); *more interesting* (4 occurrences); *more labored; more motivating* (4 occurrences, where the system has misidentified *motivating* as a gerund; here we may also include the single instance of *more motivator*, where *motivator* is the student's incorrect rendering of *motivating*); *more new; more opened* (where *opened* is a student mistake for *open*); *more personalized; more positive* (2 occurrences); *more practical* (2 occurrences); *more related; more skilled* (2 occurrences); *more specific; more standardized; more stimulating* (in this case, again, the system has tagged *stimulating* as a gerund rather than as an adjective); *more traditional* (2 occurrences); *more unfamiliar; more urgent; more visual* (2 occurrences).

As was the case in the ICT component, here we may find seventeen instances of the sequence *more and more*. In five of these, the phrase used is adverbial and each *more* should have been tagged as an adverb too: (...) *everyday more and more*; (...) *I like it more and more*; (...) *assimilate contents more and more*; (...) *the ICT's are more and more in demand*; (...) *the contents are there more and more for every day*.

In the cases where *more* is used in combination with *than*, there are quite a few instances in which *more* should have been tagged as an adverb: (...) *more than one subject*; (...) *they have more than one attempt*; (...) *more than one book*; (...) *more than 1000 words*; (...) *more than 20 years*; (...) *more than once*; (...) *more than 20 students*; (...) *the age of the teacher is more than 50*.

Finally, as pointed out when discussing *better* above, in the sequence *more and better*, *more* should have been tagged as an adverb: *students will learn more and better*.

The total amount of fails in 'more' has been 74 out of 507 instances found.

TRAINCOR component instances

Errors (students)

In the case of the shortest of our subcorpora, there are no student errors to speak of as far as use of comparative adjectives is concerned.

Fails (software)

- better: as observed in previous components, there are a few occurrences of *better* where the RBR tag would have been more appropriate. For instance: (...) *better late than never*; (...) *remember them better*; (...) *understand them better*; (...) *understand better*.

- more: as was to be expected, by far the most numerous cases of misidentification of "more" are those in which it is combined with an adjective; in those cases, as we have already mentioned, *more* should have been tagged as an adverb. Thus *more important; more intelligent; more interesting; more multicultural* (where *multicultural* has been incorrectly tagged as a noun for some reason); *more negative; more punk* (where *punk* should probably have been tagged as an adjective rather than as a noun); *more recent* (2 occurrences).

In a single case, the phrase *more and more* is used as an adverb, which seems to suggest that each *more* should have been tagged thus: *we'll use the new technologies more and more*.

There are thirteen cases in which *more* is followed by *than*. In five of these, *more* should have been tagged as an adverb rather than as an adjective: (...) *surely more than I had*; (...) *more than that, I would like to remark*; (...) *more than 10 times*; (...) *more than 15 years*; (...) *culture involves more than one country*.

There are many more instances where *more* should have in all probability been tagged as an adverb. By far the most frequent cases are those in which *more* is followed by *about*: (...) *you want to read more about it*; (...) *they all get to know more about the culture*; (...) *learning more about the culture*; (...) *learn more about the culture*; (...) *they already know more about England*; (...) *ask their partners and know more about their culture*; (...) *discover more about the culture*; (...) *know more about multiculturalism*; (...) *know more about them*; (...) *learn more about this topic*; (...) *learn more about the subject*; (...) *know more about English speaking countries*.

Cases where *more* is followed by *on* are likewise not infrequent: (...) *they are going to learn more on their own*; (...) *students learn more on their own*; (...) *they should focus more on the culture*.

Other cases include: (...) *more recently*; (...) *more specifically*; (...) *it could motivate more the students*; (...) *not the same any more*; (...) *they have looked more towards the New World*; (...) *I cannot concur more with what [...] here exposes*; (...) *I could not agree more with [...]’s opinion*.

Therefore, 36 fails have been detected in the 110 occurrences found in the component.

- northerner: there is a single case of this, which should have been tagged as an adjective (JJ) rather than as a comparative adjective (JJR): *there is a disproportionate amount of northerner news anchors*.
- worse: in the two instances in which *worse* appears, it should have been tagged as an adverb rather than as an adjective: (...) *we speak worse than in the north*; (...) *we work less and worse*.

Next table (Table 3) shows total number of errors and fails for the string *_JJR described before.

Component	Hits for *_JJR	Errors	Fails
ICT	387	1	60
TIC	1133	1	104
TRAINCOR	237	0	43

Table 3. Detection of errors (students) and fails (software) in comparatives

On the other hand, another search has been performed for detecting other errors made by students. By typing in the search box “more” and ordering instances thrown by AntConc to observe the word in context, we have found the following examples of misuse of comparative.

In the ICT component, the errors found have been *more easy* (3). In the TIC component, *more easy* (5), *more far* (1), *more funny* (3), *more later* (2), and *more little* (1). Finally, in TRAINCOR, the errors detected have been *more deep* (1), *more funny* (1), and *more sad* (1).

These cases have escaped from the search using the tag of comparative adjective because the word *more* has been tagged as an adverb.

4.3. Use of superlatives

The third type of analysis undertaken in this paper is the search for errors in the use of superlatives in the three components of the LC. The tag and string typed in this case is *_JJS. The results thrown by AntConc in this case for the three components are the ones listed and described below.

In the ICT component we have found several words tagged as superlative adjectives (291 hits). The instances thrown have been: **best** (185), biggest (8), cheapest (1), closest (2), fastest (1), fullest (3), funniest (1), greatest (3), highest (2), latest (2), least (17), **most** (55), nearest (1), newest (1), oldest (2), smallest (2), widest (1), **worst** (3), youngest (1).

In the TIC component we have found several superlative adjectives (901 hits). We refer to examples such as: **best** (513), biggest (17), clearest (1), closest (4), deepest (1), earliest (1), easiest (6), fastest (1), greatest (7), hardest (2), highest (4), largest (2), latest (13), least (55), lowest (1), **most** (243), nearest (1), newest (4), oldest (8), quickest (2), rightest (1), simplest (2), **therest** (1), widest (2), worst (5), youngest (4).

In TRAINCOR we have found several superlative adjectives (283 hits). We refer to examples such as: **best** (85), biggest (4), closest (4), craziest (1), earliest (1), easiest (1), funniest (1), greatest (5), best-known (5), largest (1), latest (1), **least** (31), longest (1), loveliest (1), mildest (1), **most** (123), nearest (1), oldest (3), simplest (1), smartest (1), wettest (1), windiest (1), worst (5), youngest (4).

As in the previous analyses, let us describe all the instances per component after observing the words in their context.

ICT component instances

Errors (students)

- worst: two errors have been detected of the three occurrences found in the component: *it can be worst for children; what is worst, I think that children...*

Fails (software)

- best: there are four cases of adverbs wrongly tagged as adjectives: (...) *the ones that suit them best*; (...) *sources of information that best adjust to your didactic goals*; (...) *activities will work best for your students*; (...) *those activities that best meet the individual development*.
- most: the single case of misidentification is *in the most quickly way*. Here, the student has made a double error: *most quickly* for *most quick*, which in turn

should be *quickest*. The system has incorrectly identified *most* as an adjective instead of as an adverb.

TIC component instances

Errors (students)

No instances have been found.

Fails (software)

- best: there are seven instances where *best* has been tagged as a superlative adjective rather than as an adverb: (...) *ready-made materials that best fits* [sic] *for our pupils*; (...) *they can learn best if they are engaged* (2 cases); (...) *the resources that best meet the interests of our students*; (...) *they could choose the one that they like best*; (...) *the student works best*; (...) *the students work best*.
- most: there are four cases where *most* should have been tagged as an adverb. The first one is unambiguous: *most interesting*. The other three are: (...) *those that can help them most in their learning process*; (...) *the one I like most*; (...) *subjects that need multimedia contents the most*.
- therest: the student's typo (*therest* for *the rest*) made the system identify this as a superlative adjective: *each child is different from therest*.

TRAINCOR component instances

Errors (students)

- in two cases superlative adjectives are used where comparative ones should have been: *better or worst* (2).

Fails (software)

- best: three different cases might be noticed here. There are seven instances where *best* is modifying an adjective and should therefore have been tagged as an adverb: (...) *best known writers*; (...) *best known food*; (...) *best known cinema*; (...) *regarding music, the best known*; (...) *best known sport*; (...) *the best known will remain British*; (...) *a presentation of the best ?* [sic] *known symbols*.

None of these forms were written by the students with a hyphen, as they should have been. Remarkably, the system has also tagged *known* in a variety of ways: JJ (3), VVN (3) and NN (1). In the single occurrence of *what I like best about them*, *best* is also incorrectly tagged as an adjective instead of as an adverb.

- least: the only fail occurring here is *when they expect it least*, where the system has tagged *least* as an adjective rather than as an adverb.

- most: there are different types of misidentification concerning this term. The most numerous group includes thirteen cases where *most* is associated to the verb *like* (and, in one instance, *love*): (...) *the sport I like most*; (...) *symbols I like most*; (...) *the three symbols I like most*; (...) *I have liked most*; (...) *the three symbols I like most*; (...) *the symbols I like most*; (...) *the three symbols that I*

like most; (...) the 3 symbols I like most; (...) what I have liked most from the video; (...) what I love most from London; (...) what I liked most in your proposal; (...) which three symbols do you like most?; (...) I liked most the wigs.

The next most numerous group is comprised of cases where *most* modifies an adjective, and is therefore obviously an adverb. In these cases, the system misidentifies it because it similarly misidentifies the accompanying adjective as a noun: *most representative* (5).

Finally, there are two cases where *most* is plainly an adverb but has been identified as a superlative adjective: (...) *one of the things that most shocked me; (...) what impressed me most during my time in London.*

Next table (table 4) summarises the total number of errors and fails described before.

Component	Hits for *_JJS	Errors	Fails
ICT	291	2	5
TIC	901	0	12
TRAINCOR	283	2	29

Table 4. Detection of errors (students) and fails (software) in superlatives

As in the case of comparative adjectives, we have typed the word *most* in search for other errors made by students. However, only the TIC component has shown more errors, apart from the ones analysed before. The three examples of errors found in this component have been *most easy* (1), and *most funny* (2).

Once again, these cases have not been detected because the word *most* has been tagged as a comparative adverb and not as a comparative adjective.

To sum up, all the errors presented in this paper are listed below in Table 5.

Component	Errors in plural adjectives	Errors in comparative adjectives	Errors in superlative adjectives
ICT	19	4	2
TIC	58	13	3
TRAINCOR	1	3	2

Table 5. Total number of errors in the three components

5. Conclusions

Observation and analysis of a LC can offer us, as teachers, more information about the students' style of writing. As stated before, a thorough analysis becomes an essential first stage for preventing errors in the future. Furthermore, this analysis will allow us a better understanding of the particulars of our students' use of grammar, in line with Dellar's contention that "one thing corpora has helped us become more aware of is the fact that grammar is much broader than sentence-based / tense-based grammar would seem to suggest" (Dellar, 2003, para. 21).

In other words, corpora should be considered as a powerful tool to improve and enrich the learning experience, making it more complete and appealing, without silencing the teacher's voice. This is particularly relevant because corpus-based analysis has seldom made the transition from academic research to educational practice. As Meunier (2011) claims: "a clear divide between the exponentially growing number of applied native corpus research and the introduction of corpus data in reference books and teaching materials on the one hand and everyday teaching practices on the other" (p. 461). Bridging this gap is, in our opinion, a priority in the teaching of EFL. In her above-mentioned paper, Meunier (2011) also suggests several ways in which corpus-based research and ESL instruction can be profitably integrated. One of those is the integration of corpus collection and error annotation in normal teaching activities, and the study presented herein represents an initial, and necessarily partial, step towards this objective.

As we have tried to highlight in our case study, there is a wide variety of (correct and incorrect) uses of the same words, constructions and structures in our students' productions. Therefore, compiling them seems crucial so as to analyse in detail their linguistic behaviour. Compilation is, then, seen as one of the main processes for the different stages in Error Analysis as described by Corbet (1974) and summarized by Ellis (1994), as we have exposed before.

On the other hand, the simple compilation of a corpus might not be enough, and, sometimes, tagging all the texts is an essential phase for searching for some concrete POS cases detected in the corpus. It is likely that the "micro-grammar" of words, mentioned by Dellar (2003), can only be fully observed and analysed by using a tagged corpus; at the very least, a tagged corpus will allow for a more systematic scrutiny of the compiled texts.

However, tagging automatically a corpus can also command the use of other strategies for searching for possible errors, as tags might be assigned in a wrong way by the software used to this purpose.

In this paper we have observed and extracted errors made by students, but, also, fails produced by the software. These fails in tagging can also lead us to propose other strings in order to detect more errors in the LC. Besides, fails from the software can also guide us to notice other errors made by students, since, as observed in the context of different instances, some of the tagging fails have been caused because students have wrongly used other words after or before the one wrongly detected, and tagged, by the software.

Consequently, teachers must take into account all these premises when exploiting a LC. The progress from the compilation of the corpus to its automatic tagging, and then to analysis of the compiled and tagged specimens, will probably suggest, at some point in our research, the need to allow for the possible fails made by the tagging software and, hence, the suitability of using longer strings to search for and analyse results or the proper error annotation to exactly record all the instances mistakenly produced by learners.

Acknowledgements

This work has been carried out in the frame of the research project B0036-1617-104-ETEL (Universidad Internacional de la Rioja, 2016-2018).

References

- Anthony, Lawrence (2015a). AntConc (version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://bit.ly/1MeMh0f>
- Anthony, Lawrence (2015b). TagAnt (version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://bit.ly/2jySOc3>
- Anthony, Lawrence (2016). EncodeAnt (version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://bit.ly/2kn7Wx5>
- Castillo Rodríguez, Cristina & Díaz Lage, José María. (2015). Exploitation of a learner corpus: analysing openings and endings in academic forums. *Revista Opción*, 31.6, 192-210. <http://produccioncientificaluz.org/index.php/opcion/article/view/20723/20615>
- Biber, Douglas, Conrad, Susan, & Reppen, Randi (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Castillejos López, Willelmira (2009). Error analysis in a learner corpus: what are the learners' strategies? In Cantos Gómez, Pascual & Sánchez Pérez, Aquilino, eds., *A survey of corpus-based research. Panorama de investigaciones basadas en corpus*, 675-690. Murcia: AELINCO. <http://bit.ly/2kQd4qn>.
- Corder, Stephen Pit (1992). *Introducción a la lingüística aplicada*. México: Limusa.
- Dellar, Hugh (2003). What have corpora ever done for us? *Developing Teachers*. <http://bit.ly/2jQn3yF>.
- Díaz-Negrillo, Ana & Thompson, Paul (2013). Learner corpora: looking towards the future. In Díaz-Negrillo, Ana, Ballier, Nicolas & Thompson, Paul, eds., *Automatic Treatment Analysis of Learner Corpus Data*, 9-29. Amsterdam: John Benjamins.
- Ellis, Rod (1994). *The study of second language acquisition*. Oxford: OUP.
- Els van, Theo, Bongaerts, Theo, Extra, Guus, van Os, Charles & Janssen-van Dieten, Anne-Mieke (1984). *Applied Linguistics and the Learning and Teaching of Foreign Languages*. London: Edward Arnold.
- Gabrielli, Richard (1998). Incorporating a student corpus in your teaching. *IATEFL Newsletter*, 141, 14-15.
- Gabrielatos, Costas (2005). Corpora and Language Teaching: Just a fling or wedding bells? *TESL-EJ*, 8.4. <http://www.tesl-ej.org/ej32/a1.html>.
- Granger, Sylviane (2003). Error-tagged learner corpora and CALL: a promising synergy. *CALICO journal*, 20.3, 465-480. doi: 10.1558/cj.v20i3.465-480
- Granger, Sylviane (2004). Computer learner corpus research. In Connor, Ulla & Upton, Thomas, A. eds., *Applied Corpus Linguistics: A Multidimensional Perspective*, 123-145. Amsterdam: Rodopi.
- Granger, Sylviane, Gilquin, Gaëtanelle, & Meunier, Fanny (2015). Introduction: learner corpus research – past, present and future. In Granger, Sylviane, Gilquin, Gaëtanelle & Meunier, Fanny, eds., *The Cambridge Handbook of learner corpus research*, 1-7. Cambridge: Cambridge University Press.
- Gilquin, Gaëtanelle (2015). From design to collection of learner corpora. In Granger, Sylviane, Gilquin, Gaëtanelle & Meunier, Fanny, eds., *The Cambridge Handbook of learner corpus research*, 9-34. Cambridge: Cambridge University Press.
- Kennedy, Graeme (1998). *An introduction to corpus linguistics*. London: Longman.
- Lozano, Cristóbal & Mendikoetxea, Amaya (2013). Learner corpora and SLA: the design and collection of CEDEL2. In Díaz-Negrillo, Ana, Ballier, Nicolas & Thompson, Paul, eds., *Automatic Treatment and Analysis of Learner Corpus Data*, 65-100. Amsterdam: John Benjamins.
- McEnery, Tony, Xiao, Richard & Tono, Yukio (2006). *Corpus-Based Language Studies*. London and NY: Routledge.

- McEnery, Tony & Wilson, Andrew (2001). (2nd edn.) *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Meunier, Fanny (2011). Corpus linguistics and second/foreign language learning: exploring multiple paths. *Revista Brasileira de Linguística Aplicada*, 11.2, 459-477.
- Meyer, Charles F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Nesselhauf, Nadja (2004). Learner corpora and their potential for language teaching. In Sinclair, John, ed., *How to Use Corpora in Language Teaching*, 125-152. Amsterdam: Benjamins.
- Pravec, Norma. A. (2002). Survey of learner corpora. *ICAME journal*, 26, 81-114.
- Tognini-Bonelli, Elena (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- UCREL, n.d. *CLAWS part-of-the-speech tagger for English*. Retrieved from <http://ucrel.lancs.ac.uk/claws/>
- Widdowson, Henry George (1991). The description and prescription of language. In Alatis, James E., ed., *Georgetown University Round Table on Languages and Linguistics. Linguistics and language pedagogy: The state of the art*, 11-24. Washington, D.C.: Georgetown University Press.