TO BE OR NOT TO BE...?
ANALYSIS OF MISUSES WITH *BE* IN A LEARNER CORPUS
IN AN ONLINE ENVIRONMENT

Cristina Castillo Rodríguez
Universidad Internacional de La Rioja; Universidad de Málaga

Sidoní López Pérez
Universidad Internacional de La Rioja

Abstract

Learner corpora provide teachers with a rich source of real learners' productions in a given language. In fact, teachers from a foreign or second language (FL or L2) can have an immense source of contributions, either written or oral, for analyzing certain linguistic patterns in their students. In this paper, we are aware of this situation and will pay attention to some structures, commonly produced as a misuse: the use, or absence, of the verb *to be* in certain structures such as *born* and *agree*. The learner corpus compiled for this study is formed by written productions of higher-education students whose mother tongue is Spanish and that are enrolled in some non-linguistic subjects using English as a Medium of Instruction (EMI subjects). Therefore, some *literal translation*, very closed to the main term in Spanish, will be found and described in this study.

Keywords: corpus linguistics, learner corpora, error analysis, ESL, applied linguistics.

¿SER O NO SER...? ANÁLISIS DE ERRORES CON *TO BE* EN UN CORPUS DE ESTUDIANTES EN UN ENTORNO ONLINE

Resumen

Los corpus de estudiantes constituyen una fuente rica de datos reales producidos genuinamente por alumnos en una lengua dada. De hecho, los docentes de lenguas extranjeras (LE) o de segundas lenguas (L2) pueden disponer de una gran fuente de producciones, ya sean escritas u orales, para analizar ciertos patrones lingüísticos de los estudiantes. En este artículo, somos conscientes de esta situación y prestaremos atención a estructuras que se suelen utilizar de forma errónea: el uso, o ausencia, del verbo *be* en ciertas estructuras en combinación con *born* y *agree*. El corpus de estudiantes compilado para este estudio está formado por producciones escritas de estudiantes universitarios cuya lengua materna es española y que están matriculados en asignaturas no lingüísticas impartidas en inglés como medio de instrucción. Por tanto, algunos casos de *traducción literal*, muy cercana a los términos en cuestión en español, se describirán en este estudio.

Palabras clave: lingüística de corpus, corpus de estudiantes, análisis de errores, inglés como segunda lengua, lingüística aplicada.

1. INTRODUCTION

The use of learner corpora for second language (L2) or foreign language (FL) research has given rise to multiple studies about the learning of English by students from different mother tongue backgrounds. In fact, a series of learner corpora in English have been released, like, NOSE, SULEC, or WriCLE, to mention a few of them (Centre for English Corpus Linguistics, n.d.) providing an excellent source of real information to analyse the interlanguage of L2 learners. Although those studies provide useful information about the learning of English as a Foreign Language (EFL) or the acquisition of English as a Second Language (ESL), they rely extensively on the collection of samples that include, as an instrument for data collection, specific writing assignments, fill-in-the-gaps tests or other activities. This, however, leaves unattended «students' own spontaneous language production» (Liu and Gleason, 2002: 20), which is also a very rich source to analyse and research about the interlanguage of learners of various language groups.

For the purpose of this study, we have focused on a monolingual comparable corpus, ENTECOR[1], divided into two sub-corpora, TICOR and SECOR, and containing spontaneous written productions of 786 learners. These samples belong to the students' natural contributions to online compulsory forums from three different English-medium subjects. These EMI subjects are included in the curriculum of the Degree in Early Years Education, the Degree in Primary Education and the Master's Degree in Secondary Education, from an online university (UNIR). All of those EMI subjects were utterly taught in an online environment. In this case, we aim at analysing the misuses with *be*, concretely, in the cases of *to be born* and the verb *to agree* in the whole corpus. The verb *to be born* is translated as *nacer* into Spanish and does not take *to be* in that language, whereas the verb *to agree* is translated as *estar de acuerdo* into Spanish and does take the verb *to be* (or *estar* in that language).

As a consequence, and as analysed in detail in this study, our learner corpus reveals that even though a certain level of the English language is required for being enrolled in the three subjects, they keep on omitting the verb *to be* in the structure *to be born* and employing the verb *to be* with the verb *to agree*. This leads to some percentage of literal translation from Spanish into English by our students, which indicates a considerable interference of the learners' mother tongue and a clear language transfer (LT) from Spanish into English that we, as EMI-subject faculty members, should take into account so that we can implement some strategies in further studies and research to avoid those misuses.

2. THE INFLUENCE OF THE NATIVE LANGUAGE
IN THE SECOND OR FOREIGN LANGUAGE LEARNING

When attempting to learn an L2, it has always been assumed that «learners rely extensively on their native language» (Gass and Selinker, 2008: 89). In fact, Lado (1957: 2) made reference to the importance of the native language (NL) in L2 acquisition after establishing LT as the main difficulty of FL learning:

> Individuals tend to transfer the forms and meanings, and the distribution of forms and meanings of their native language and culture to the foreign language and culture–both productively when attempting to speak the language and to act in the culture, and receptively when attempting to grasp and understand the language and the culture as practiced by natives.

However, although there is clear evidence that LT is an important factor in L2 acquisition, there are other factors which are also important and that usually work with LT when attempting to learn a new language (Ellis, 1994). LT has had a controversial history during the course of second language acquisition (SLA) research and its importance in the context of L2 learning has been reassessed on numerous occasions (Lu, 2010).

To start with, many scholars have provided different terms and definitions for LT since the late 19th century. Whitney (1881) made use of the term *transfer* in order to refer to cross-linguistic influence, whereas Corder (1983) proposed the abolishment of the term *transfer* in favour of mother tongue influence. Some years later, Kellerman and Sharwood Smith (1986) suggested the use of the terms cross-linguistic influence or cross-linguistic generalization, while well-known scholar Gass (1988: 385) made reference to LT as «the use of prior linguistic information in a second language context». More recently, Odlin (1989: 27) defined LT as «the influence resulting from the similarities and differences between the target language and any other language that has previously (and perhaps imperfectly) acquired». In addition, scholars have also reached the conclusion that LT can impact the learning of an L2 in both positive and negative ways. Whereas positive transfer takes place when the previous knowledge of the first language (L1) facilitates the subsequent learning of an L2, negative transfer usually occurs when the previous knowledge interferes with new learning. The latter is also referred to as interference (Brown, 2007), which is usually viewed as «the transference of elements of one language into the learning of another at various levels of linguistics which include: phonology, morphology, syntax and the orthography of the languages» (Offiong and Okon, 2013: 899).

At the same time, LT has also been assigned different roles in theories of SLA. In fact, LT is one of the aspects often discussed and approached when dealing with the process by means of which learners acquire an L2 (Yu and Odlin, 2016). Such is the case of the Contrastive Analysis Hypothesis (CAH), a theory elaborated by Lado (1957), which established that the L1 interference is the main barrier for L2 learners (Quesada, 1995). However, Dulay and Burt's Creative Construction Hypothesis (1974) downplayed the role of L1 influence as their proponents defended that learners build

hypotheses of the L2 they are learning according to inherent mechanisms (Gass and Selinker, 2008). Some years later, Dulay, Burt and Krashen (1982) totally denied LT from the native language after clearly believing that language learning capacity relied only on Universal Grammar (UG) (Lu, 2010). Similarly, Zobl (1980) also contributed to minimizing the role of the L1 after establishing the so-called Transfer Hypothesis, a theory that supports the idea that the L1 must fulfil some conditions before being transferred: «it must be productive in the L1 (not some kind of exception), it must be used frequently, and it must not be 'on the way out' historically speaking» (Ellis, 1994: 321). Other scholars such as Eckman (1977), Kellerman (1977), and Corder (1981) focused on analyzing and providing different aspects that could lead to LT, whereas Ellis (1994) listed six factors causing LT that include social and developmental factors, different linguistic levels, markedness, prototypicality and language distance.

As a result, multiple studies about the role and influence of LT in L2 acquisition have been published during the past two decades –especially regarding the use of English as an L2– (e.g Nation, 2003; Lu, 2010; Karim and Nassaji, 2013; Ionescu, 2014; Brogan and Son, 2015; Wang, 2015; Ahmadian, Pouromid and Nickkhah, 2016). Some of these publications focus on how English learners make use of their L1 when writing in the L2, suggesting and implying that L1-based strategies can have beneficial effects on L2 writing (Karim and Nassaji, 2013) and can also contribute to improving L2 composition (Kim and Yoon, 2014). At the same time, when examining the role of L1 in L2 composition, some researchers conclude that the expressions and meanings from L1 used by learners of English are clearly imbedded in the students' written compositions, revealing different cultural expressions of the writers themselves (Ionescu, 2014). However, other studies reveal that negative transfer of L1 is a very important factor to interfere in English writing. Such is the case of Wang (2015), who examines 60 compositions by Chinese college students that prove that the mode of thinking of their native language clearly interferes in the students' improvement in English written production. In the case of the research carried out by Taylor (1975), White (1977), Alonso (1997), Spada and Lightbown (1999), it is also clear that L2 learners of English from different backgrounds rely on the structure of their mother tongue (Spanish and French, respectively) when writing in English. As these studies show, LT becomes the main factor of interference in the students' L2.

On the other hand, the research and analyses carried out all throughout these works demonstrate that LT is one of the multiple aspects «that can affect a person's acquisition, competence, or performance in the TL» (Yu and Odlin, 2016: 29). Among those other factors affecting SLA, several authors coincide that motivation is one of the most important (Khasinah, 2014; López Montero, Quesada Chaves and Salas Alvarado, 2014; Adwani and Shrivastava, 2017). However, these studies also reveal that there are more factors such as attitude, age, intelligence, aptitude, personality, stereotyping, self-efficacy and cultural background, which are also relevant for the process of acquiring an L2.

In addition, previous studies focusing on the error production of non-native speakers of English through the use of different written assignments confirm that a great number of those errors are due to LT from L1 to L2. This leads us to the main purpose of the compilation of a learner corpus (also called interlanguage or

L2 corpus, following the terminology used in Granger, 2003): observing linguistic behaviours and patterns as well as the errors made by students of an L2.

As mentioned before –and as we will see in the analysis of our corpus–, Spanish learners, who are non-native speakers of English, make a wrong use of *to be born* and *to agree*, especially because they are relying extensively on the use of their mother tongue, whilst giving rise to what has been labeled as interlingual errors, that is, those errors «caused by the learner's native language» (Richards and Schmidt, 2002: 26). It should be noted that in Spanish the verb *to be born* does not take the verb *to be*, whereas in the case of the verb *to agree*, the verb *to be* is obligatorily required. As the results will show, LT from Spanish into English is a major cause of error production in the written contributions of our students.

3. METHODOLOGY

3.1. Participants and learner corpus

The learner corpus, called ENTECOR, contains all the contributions produced spontaneously in forums from three subjects taught at UNIR. Two subcorpora have been compiled: TICOR and SECOR. On the one hand, the TICOR subcorpus has two components: ICT –texts from a non-linguistic subject taught in the framework of the Degree in Early Years Education– and TIC –texts from a subject taught in the Degree in Primary School Education. On the other hand, the SECOR subcorpus has only one component, up to now: TRAINCOR, composed of learners' spontanous productions from a subject taught in the Master's Degree in Secondary Education.

All the names and personal information of the learners have been removed with the aim of preserving the learners' privacy. The EMI subjects selected for this study are taught entirely in an online setting, as UNIR is an online university, with no face-to-face sessions. As mentioned before, the subjects use English as a Medium of Instruction (or EMI). That is to say, they are subjects not devoted to learning English, but to learning specific content using English as the language for instruction, which is different from the mother tongue of our students.

ENTECOR contains 527,099 tokens and 13,148 types, from a total of 786 learners of the three subjects mentioned before. These figures correspond to the following:

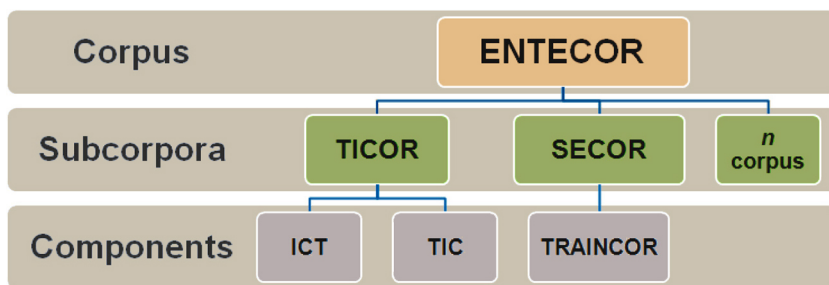| TABLE 1. ENTECOR INFORMATION				
Subcorpus	Component	Tokens	Types	No. Students
TICOR	ICT	107,072	4,821 types	155
	TIC	317,759	9,320	511
SECOR	TRAINCOR	102,268	6,380	120

Figure 1. Current structure of the ENTECOR corpus.

The reason why the TIC component contains more texts, and, therefore, more tokens and types, is because an EMI subject has been proposed first in that degree, while the rest have been incorporated later as EMI subjects. With regard to the academic year when those texts were collected, the TIC component gathers texts compiled from the academic year 2011-2012 to 2015-2016; the ICT component, from 2014-2015 to 2015-2016; and TRAINCOR, from 2015-2016 to 2017.

More contributions (written and oral) produced by learners will be collected for increasing the subcorpora gathered in ENTECOR. The following figure illustrates the current learner corpus and their subcorpora:

3.2. Process of compilation and tools

Every corpus compilation entails the establishment of some design criteria (Díaz Negrillo and Thompson, 2013; Gilquin, 2015). The ones highlighted for this corpus have been:

 i) Medium: written texts.
 ii) Annotation or raw corpus: for this analysis, we have not annotated with errors, that is, we have used the raw corpus.
 iii) Target language: the corpus is monolingual, as it only contains texts (produced by learners) in English.
 iv) Data conditions: as far as the degree of naturalness is concerned, the forums are proposed with a specific goal in the different subjects. The faculty members of those subjects post their contributions first with the aim of starting the debate and specifying the instructions of participation. But, apart from that, the whole environment and the features of the forum activity imply some degree of spontaneity and naturalness in the learners' written participation.
 v) Time: as it is not a study of the same learners across the time, we can state that it is a cross-sectional study.

VI) Academic background: the level of English of the participants is different. The participants whose texts have been compiled for the TICOR subcorpus are supposed to have B1 in English, since the subjects are part of the English mention of the degree they are studying. The participants with contributions compiled for the TRAINCOR subcorpus have B2 in English, as this is the requirement for being enrolled in the postgraduate master.

For the compilation of ENTELEARN, some specific steps have been followed (Castillo and Díaz, 2015), as we summarise below:

1) Downloading and codification. The purpose of this phase is to collect all the data from the forum task from the three subjects in their different academic years and provide a univocal code for every instance. Examples of codes employed in the corpus are: 23TOENICTA, 23TOENICTB, 23TOENTICA, TOENTICB, 01TOENTRAIN. 01 means Student1; TO, source text; EN, English; ICT corresponds to the subject from the Degree in Early Years Education; TIC, to the Degree in Primary School Education, and TRAIN, to the Master's Degree in Secondary Education. The letters indicated in the ICT and TIC components stand for forum 1 (letter A), and forum 2 (letter B) of every subject edition containing two compulsory forums during the academic year. Another code number has been added at the end in cases in which students were enrolled in the subject again.
2) Cleaning and saving. This phase is devoted to eliminating all personal information from students.
3) Registering code. The third phase's goal is to register every item in an excel document for listing all the participants, the degree, the academic year in which contributions were collected, etc.
4) Exploitation. This phase constitutes the analysis and management of data included in the different components and subcorpora of the learner corpus.

The tool used for the analysis in the last phase of our learner corpus' compilation has been AntConc (version 3.4.4.), which is formed by a set of modules of analysis for managing texts of a specific corpus (Anthony, 2015). Concretely, the main module of analysis used for this study has been Concordance, as the main interest was to search for specific strings to observe the expressions used by the students and their potential misuses around the verb *to be*, as analysed in the subsequent section.

4. ANALYSIS OF MISUSES WITH *BE* IN THE LEARNER CORPUS

The potential misuses of the verb *to be* with the structures *born* and *agree* have been detected using the module of Concordance from the AntConc software. The strings typed in the search box and the order of hits thrown by the module are listed and described below:

– born*. All the hits thrown have been ordered in two Levels, that is, Level 1: 0, and Level 2: 1L. This string and order allow us to localize all the instances with the words preceding them, and also the different variants in the endings of the word of search. The asterisk symbol [*] enables us to visualize if some other errors in spelling in the word have been provided by students.
– *agre*. The hits thrown by the system have also been ordered but in three levels, i.e. Level 1: 0; Level 2: 1L; and Level 3: 1R. This enables us to detect the use of the verb *to be* and also the preposition used after the main verb of search. Besides, [*] symbol allows us to observe the possible variants in the use of the word of search in the different tenses used by students.

4.1. Misuses with *born*

The texts gathered in the ICT component (TICOR subcorpus) threw a total of 23 hits for *born** typed in the search box of AntConc. Five misuses have been detected (other misuses in the spelling of other words have been indicated in the examples of the table):

– 3 misuses related to the absence of the verb *to be*.
– 2 grammar errors, as there are two occurrences which included the past and the past participle applying the rule for regular verbs; therefore, instead of using *have been born*, the student wrote *have borned*, and, instead of using *were born*, the student wrote *borned*.
– 1 error not related to the purposes of the study, but detected in the string used in AntConc. It is a spelling error related to the past participle of the verb *to bear* for the sequence *to bear in mind*.

The next table summarises all the aforementioned misuses for the component called ICT from the TICOR subcorpus. Some cells from the right column have not been highlighted because they belong to the same student.

TABLE 2. MISUSES FOUND FOR *BORN** IN ICT COMPONENT	
Examples of misuses	Student ID
Our student [sic] have born In [sic] the century of the technologies...	67TOENICTA
...our students have borned in the digital era...	77TOENICTA
The generation borned in the last 10 years...	99TOENICTA
...they have born surrounded by technology, so they are...	109TOENICTA
...we, as teachers, should borne in mind.	109TOENICTA

In the TIC component, from the TICOR subcorpus, a total of 34 hits have been thrown with the first word of search of the study: *born**. However, in this case, the amount of misuses has been only two. No cases of spelling mistakes have been

found in this component. The following table collects the two examples of sentences where the misuses have been detected.

TABLE 3. MISUSES FOUND FOR *BORN** IN THE TIC COMPONENT	
Examples of misuses	Student ID
...since I born,...	88TOENTICA
They have basically born with a laptop under one arm!	251TOENTICA

Finally, in the TRAINCOR component, from the SECOR subcorpus, the 11 hits thrown with the same search box have not shown any misuse. This means that all of the instances presented correctly the verb *to be* before this word of search, or the past participle was perfectly used.

4.2. Misuses with *agree/disagree*

The second word of search typed in AntConc, as mentioned before, was **agre** with the aim of detecting misuses for the verb *to agree/to disagree*, which, according to our learners' mother tongue, usually presents the tendency of appearing with *be*, as analysed below in the different components of the two subcorpora.

In the ICT component, the total of hits thrown by AntConc has been 351. However, one word has been ruled out, *agreement*, as it is not the purpose of the analysis of this paper. Therefore, the hits analysed (with the use and misuse of verbs containing *agree* or *disagree*) are 350, from which 25 presented misuses. All of the instances of *agree/disagree* were preceding the verb *to be.* The table 4 illustrates these examples of misuses and the code of student who committed the error. As indicated before, some of the cells from the right column have not been highlighted as they belong to the same student coloured in grey in a precedent cell:

TABLE 4. MISUSES FOUND FOR **AGRE** IN THE ICT COMPONENT	
Examples	Student code
I am agree with you, [Studentname].	03TOENICTA
...we all are agreeing with [Studentname]	08TOENICTA
...I'm agree with you that times change...	08TOENICTA
...I'm completely agreed with [Studentname], about...	08TOENICTB
...a lot of us are agree about the importance...	08TOENICTB
I'm agree with you too,	08TOENICTB
...I'm completely agree with us...	14TOENICTA
I couldn't be more agree with you.	40TOENICTB
...I'm completely agree with [Studentnameandsurname]...	49TOENICTA
I am agree with [Studentname] that...	63TOENICTA

...and I am really agree with [Studentname] in her last contribution.	63TOENICTB
...I am totally agree with my partners that...	63TOENICTB
...I am agree with you...	67TOENICTA
...I am totally agree with you, when you say...	67TOENICTB
...why I'm agree with you when you say...	67TOENICTB
...im [sic] not totally agree with her because...	69TOENICTB
...I am totally agree with your idea in which...	69TOENICTB2
...I am completely agree with all what you said.	69TOENICTB2
...I'm totally agree with [Studentname] that ICT tools...	75TOENICTA
Are you agree with this?	101TOENICTB
...I am agree also with [Studentname] about one of...	112TOENICTB
Are you agree with me?	112TOENICTB
I'm agree with [Studentname] about the time...	143TOENICT
...that the majority of us be agrees about...	144TOENICT
...I'm totally agree with you...	149TOENICT

Note that the misuses detected for the ICT component have been committed by 14 students. This figure reveals us that there are students making the same type of mistake (i.e. the use of *be* with the verb *agree*) along their contributions, as the cases of students 08, who fails in the use of this verb in both forums (2 misuses in A and 3 in B), 63 and 67 (both with 1 mistake in A and 2 in B), 112 (2 mistakes in B), and the case of 69, who was even enrolled for the second time in the subject (69TOENICTB2) and presented the same type of error for the same forum.

Besides, as observed in the instances thrown with the Concordance module of analysis of AntConc, in ICT component when the word is used in a question, the number of mistakes is reduced: of 20 questions with *agree*, only 2 showed the structure with the verb *to be*.

In the TIC component, 829 hits have been thrown by AntConc for the second search word object of study. However, the hits valid for study are 822, since sequences in which the search word used is a noun (*agreement* or *disagreement*) are not the purpose of this analysis.

The amount of misuses detected for **agre** typed in the corpus management tool for the TIC component have been 77. All of these instances showed the use of the verb *to be* together with *agree* or *disagree*. The following table shows all the instances of misuses together with the code of student.

TABLE 5. MISUSES FOUND FOR *AGRE* IN THE TIC COMPONENT	
Examples	Student code
I'm agree with you.	50TOENTICB
I am agree with you [Studentname], but...	112TOENTICB

...that I am totally agree and sure...	150TOENTICA
I'm agree with all the opinions...	155TOENTICA
...and I am agree to use ICT in the school.	156TOENTICA
...and they are agree with the same opinion,...	156TOENTICB
I am totally agree with [Studentname],...	179TOENTICA
I'm totally agree too.	190TOENTICA
I'm agree with [Studentname] that...	211TOENTICA
...because they are disagree with...	224TOENTICA
...every of us are agree with the point that...	225TOENTICA
I am completely agree with my colleagues...	231TOENTICB
I am completely agree with you,...	241TOENTICB
I am completely agree with you.	254TOENTICB
I'm agree with you when you say...	254TOENTICB
I'm agree with [Studentname],...	258TOENTICB
I'm agree with [Studentname],...	281TOENTICB
...I'm agree with [Studentname]...	281TOENTICB
...I am agree with [Studentname]...	282TOENTICA
...we are agree in this point...	295TOENTICA
...I am completely agreed with them.	301TOENTICB
I'm agree with you.	310TOENTICB
...I'm not quite agree.	315TOENTICA
Thanks [Studentname] [sic] being agree with the thing that I said.	320TOENTICB
So I am completely agree with all of you...	321TOENTICA
I'm also agree with the idea that...	354TOENTICB
I'm agree with [Studentname] when he said that...	354TOENTICA
...that I'm agree with some other participants.	354TOENTICA
Also I'm agree with her when she said that...	354TOENTICA
...that I'm agree with all of you about...	354TOENTICB
...I'm agree with [Studentname]...	354TOENTICB
...I'm really agree with [Studentname]...	361TOENTICB
I'm agree with you [Studentname]...	365TOENTICB
...I'm totally agree with you,...	389TOENTICB
I'm totally agree with you.	392TOENTICA
I'm totally agree with the content...	392TOENTICB
I'm totally agree with you.	392TOENTICB
...I'm absolutely agree with the use of...	405TOENTICA
I'm totally agree with [Studentname] when she...	405TOENTICA

...I'm absolutely agree with [Studentname] and [Studentname]...	405TOENTICB
...I'm agree with [Studentname]...	405TOENTICB
...I'm 100% agree with some of the statements...	417TOENTICA
...all of us are agree in ICT...	428TOENTICA
...that all of us are agree with the best...	428TOENTICB
We are agree when we talk about...	428TOENTICB
I am agree with you [Studentname].	430TOENTICB
...I'm agree with my classmates about...	438TOENTICA
I'm agree with my partner [Studentname]...	449TOENTICA
I'm agree with you, [Studentname],...	449TOENTICB
...but I'm disagree with you...	449TOENTICB
I'm not totally agree with you, [Studentname].	457TOENTICB
...I can't be more agree with you.	460TOENTICA
...I can't be more agree with you.	460TOENTICB
...and I'm agree with both.	470TOENTICA
I am totally agree with you, because...	472TOENTICA
...I'm agree with you [Studentname],...	472TOENTICB
I'm agree with my partners,..	480TOENTICA
...we are all more or less agree that...	480TOENTICB
...I am also agree with [Studentname] and [Studentname]	481TOENTICB
I am agree with [Studentname] and [Studentname]...	481TOENTICB
Good night, I am agree with [Studentname]...	481TOENTICB
Good afternoon, I am agree with all students but...	483TOENTICA
I'm agree with [Studentname], [Studentname] and [Studentname]	483TOENTICA
I'm agree with you... (×2)	483TOENTICB
...I'm agree with [Studentname] about...	483TOENTICB
I am agreeing with [Studentname].	485TOENTICA
I'm agree with [Studentname]	485TOENTICB
Im [sic] agree with you...	495TOENTICA
Are you agree with me?	495TOENTICB
But I'm agree with [Studentname] that...	495TOENTICB
I'm agree with [Studentname] or [Studentname],...	503TOENTICB
I'm [sic] agree with [Studentname],...	506TOENTICA
I'm agree with [Studentname] and [Studentname]... (×2)	508TOENTICB
I'm agree with you, [Studentname]. (×2)	510TOENTICB

Note once again that some of the misuses have been committed by the same student.

– Code of students with misuses in two forums: 156, 354, 392, 405, 428, 449, 460, 472, 480, 483, 485, and 495;
– Code of students with misuses in one forum: 254, 281, 481.

In this component, only one misuse was found in a question for 12 total questions thrown by Concordance module.

As far as the number of students with misuses, as observed in the table the number for the TIC component is 46.

Finally, in the TRAINCOR component from the SECOR subcorpus, 188 hits have been thrown by the Concordance module of analysis, but once we have ruled out the words which are not verbs (object of study), like *agreement*, the total hits are 186. Two misuses have been detected in this component. Besides, one of those examples shows a spelling mistake, but it might be due to a typo mistake, rather than a misuse in itself (see table 6).

TABLE 6. MISUSES FOUND FOR THE WORD *AGREE** IN THE TRAINCOR COMPONENT	
EXAMPLES	STUDENT CODE
I am totally [sic] desagree with you.	21TOENTRAIN
Are you agree?	50TOENTRAIN

The following table summarises the cases of misuses with *agree* and *disagree* for the three components:

TABLE 7. RESULTS FOR MISUSES OF *AGREE* AND *DISAGREE* IN ENTECOR					
COMPONENT	AGREE/DISAGREE USED IN QUESTIONS	MISUSE IN QUESTIONS	AGREE/DISAGREE USED IN STATEMENTS	MISUSE WITH *AGREE*	MISUSE WITH *DISAGREE*
ICT	20	2 (only *agree*)	331	23	0
TIC	12	1 (only *agree*)	822	75	2 (out of 17)
TRAINCOR	15	1 (only *agree*)	186	0	1 (out of 3)

Finally, we present in the following table the number of misuses found for the two search words in the three components of the corpus ENTECOR:

TABLE 8. TOTAL NUMBER OF MISUSES FOR *BORN** AND *AGREE* IN ENTECOR				
SEARCH WORD	COMPONENT	NO. OF MISUSE	NO. OF STUDENTS	NO. OF HITS
born*	ICT	5	4	23
	TIC	2	2	34
	SECOR	0	0	11
agree	ICT	25	14	350
	TIC	77	46	822
	SECOR	2	2	186

5. FINAL REMARKS AND FUTURE RESEARCH

As observed in the analysis and results shown previously, Spanish students in our learner corpus continue to make a wrong use of the structures *to be born* and *to agree* in English, and, therefore, there is a negative transfer from their mother tongue into the foreign language they use in class. This conclusion of the use of the interlanguage among non-native students is not new in the literature. Other studies such as Bazo Martínez (2001), Peñate Cabrera and Bazo Martínez (2002), Wood and Wood (2002), and Medina Suárez (2013) have researched on the influence of the interlanguage of students using another foreign language.

However, the main novelty of this study is that participants are, first of all, higher education (HE) students, with a B1 level in English (in the case of undergraduate students) and B2 (in the case of postgraduate students); secondly, the study has been carried out in an environment which is utterly online with no face-to-face sessions; third, the activity suggested to check their production in English requires written contributions, even though the forum is a kind of activity in which our students are supposed to provide their posts in a more spontaneous way, similar to oral production; and, fourth, the study has been conducted with HE students enrolled in EMI subjects, so English is not specifically taught in the online instructions.

With regard to the quantitative analysis of the misuses encountered before, the search word *born** appears 68 times in ENTECOR. Of those, 61 are correct and 7 are incorrect, which means that the term is used wrongly in 10.29% of the occurrences. In the case of *agre*, this appears 1172 times in ENTECOR. Of those, 1070 are correct and 102 are incorrect, which means that the term is used wrongly in 8.70% of the hits. Although these percentages are not very high, the wrong use of the structures *to be born* and *to agree* by our Spanish HE students, enrolled in EMI subjects taught entirely in an online setting, clearly indicates that our HE learners keep on showing some misuses with regard to the aforementioned verbs. In our corpus, the learners tend to translate the structures *to be born* and *to agree* directly from Spanish into English, giving rise to multiple errors in the FL.

As explained before, the structure *to be born* does not take the verb *to be* in Spanish. In this language, the verb *to be born* is translated as *nacer* and it cannot take the verb *to be* under any circumstances. However, the verb *to be* is obligatorily required for the structure *to be born* in English; otherwise, it would be grammatically incorrect. This explains why all the incorrect uses of the term *born** in the corpus include the verb *to be*. Our students write a literal translation from Spanish into English because they think that the structure *to be born* follows the same pattern as in Spanish, which is not the case. The same happens with the verb *to agree*. In our students' mother tongue, this verb requires obligatorily the use of *to be* and is translated as *estar de acuerdo*. If such a verb is not used with the structure *to agree*, the sentence will be automatically wrong and grammatically incorrect. It should be noted once again that Spanish does not contemplate the use of the verb *to agree* without the verb *to be*. That is the reason why all the misuses of *agre* in the corpus appear with the verb *to be*. Once again our students are making a literal translation

from their mother tongue to the FL and the results include multiple errors in the conjugation of the structure *to agree*. In this case, Spanish students in the corpus do not conjugate the structure *to agree* with the verb in the base form in the Present Simple. They conjugate this structure with the form of the verb *to be* in the Present Simple as this is the case in Spanish.

Therefore, we can conclude that the errors produced by Spanish students regarding the structures *to be born* and *to agree* in our corpus are mainly due to LT from mother tongue to FL. Whereas LT has been said to affect the learning of a FL in both positive and negative ways, in our corpus LT is clearly negative. At this point, it would be really useful to deal with additional practice and activities into our non-linguistic lessons in the framework of an EMI subject so that we can assist learners with fluency and help them overcome the difficulty in producing natural and correct sentences in a FL.

Finally, one of the future lines we contemplate is the error annotation of the learner corpus and its use in the EMI subjects. Our future purpose is to teach our university students from EMI subjects belonging to degrees and master's degrees how to use the annotated texts from ENTECOR so that they can consult their queries, and check if the misuses caused by LT, like the ones exposed in this work, are totally eliminated so as to shed some light upon the way we could reduce the effect of negative LT into the written and spontaneous production in FL.

REFERENCES

Adwani, Priya and Sanjay Shrivastava (2017): «Analysis of factors affecting second language acquisition», *International Journal of Social Sciences and Management,* 4(3): 158-164.

Ahmadian, Mousa, Sajjad Pouromid and Mehdi Nickkhah (2016): «Improving the quality of second language writing by first language use», *Theory and Practice in Language Studies*, 6(4): 767-775.

Alonso Alonso, María Rosa (1997): «Language transfer: Interlingual errors in Spanish students of English as a foreign language», *Revista Alicantina de Estudios Ingleses*, 10: 7-14.

Anthony, Lawrence (2015): AntConc (version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. URL: http://bit.ly/1MeMh0f.

Bazo Martínez, Plácido (2001): «A qualitative analysis of the interlanguage found in compulsory secondary education students in the Canary Islands». *Revista Canaria de Estudios Ingleses*, (43): 209-218.

Brogan, Franny D. and JyEun Son (2015): «Native language transfer in target language usage: An exploratory case study». *Voices*, 3(1): 47-62.

Brown, Douglas H. (2007 [1980]): *Principles of language learning and teaching.* 5th ed., New York: Pearson Education.

Castillo Rodríguez, Cristina and José María Díaz Lage (2015): «Exploitation of a learner corpus: analysing opening and endings in academic forums», *Revista Opción,* Supl. 31(6): 192-210.

Centre for English Corpus Linguistics. (n.d.): *Learner corpora around the world*. URL: https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html.

Corder, S. Pit (1983): «A role for the mother tongue», in Susan M. Gass and Larry Selinker (eds.), *Language Transfer in Language Learning,* Rowley: MA: Newbury House.

Corder, S. Pit (1981): *Error analysis and interlanguage,* Oxford: Oxford University Press.

Díaz-Negrillo, Ana & Paul Thompson (2013): «Learner corpora: looking towards the future», en Ana Díaz-Negrillo, Nicolas Ballier and Paul Thompson (eds.), *Automatic Treatment Analysis of Learner Corpus Data*, Amsterdam: John Benjamins, 9-29.

Dulay, Heidi C. and Marina K. Burt (1974): «Natural sequences in child second language acquisition», *Language Learning,* 24: 37-53.

Dulay, Heidi C., Marina K. Burt, and Stephen Krashen (1982): *Language two,* New York: Oxford University Press.

Eckman, Fred R. (1977): «Markedness and the contrast analysis hypothesis», *Language Learning,* 27(2): 315-330.

Ellis, Rod (1994): *The study of second language acquisition,* Oxford: Oxford University Press.

Gass, Susan M. (1988): «Second language acquisition and linguistic theory: The role of language transfer», in Suzanne Flynn and Wayne O'Neil (eds.), *Linguistic Theory in Second Language Acquisition,* Dordrecht: Kluwer Academic Publishers, 384-403.

Gass, Susan M. and Larry Selinker (2008): *Second language acquisition: An introductory course.* 3.rd edition. New York: Routledge.

Gilquin, Gaëtanelle (2015): «From design to collection of learner corpora», in Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier (eds.), *The Cambridge Handbook of learner corpus research*, Cambridge: Cambridge University Press, 9-34.

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.

Ionescu, Ioana Daciana (2014): *An analysis of the role of first language in second language acquisition*, (Unpublished doctoral dissertation). San Diego State University. San Diego, CA.

Karim, Khaled and Hossein Nassaji (2013): «First language transfer in second language writing: An examination of current research», *Iranian Journal of Language Teaching Research*, 1(1): 117-134.

Kellerman, Eric (1977): «Towards a characterizations of the strategies of transfer in second language learning», *Interlanguage Studies Bulletin,* 2(1): 58-145.

Kellerman, Eric and Michael Sharwood Smith (1986): *Crosslinguistic influence in second language acquisition*, New York: Pergamon Press.

Khasinah, Siti (2014): «Factors influencing second language acquisition», *Englisia Journal*, 1(2): 256-269.

Kim, Youngran and Hyunsook Yoon (2014): «The Use of L1 as a Writing Strategy in L2 Writing Tasks», *GEMA Online® Journal of Language Studies,* 14(3): 33-50.

Lado, Robert (1957): *Linguistics across cultures*, Ann Arbor: University of Michigan Press.

López Montero, Rosberly, María José Quesada Chaves and Jonnathan Salas Alvarado (2014): «Social factors involved in second language learning: A case study from the Pacific campus, Universidad de Costa Rica», *Revista de Lenguas Modernas,* 20: 435-451.

Liu, Dilin and Johanna L. Gleason (2002): «Acquisition of the article *the* by nonnative speakers of English: An analysis of four nongeneric uses», *SSLA*, 24: 1-26.

Lu, Lanfeng (2010): *Language Transfer: From topic prominence to subject prominence.* (Unpublished master's thesis), University of Leeds, England. URL: https://bit.ly/2uS0DRR.

Medina Suárez, José (2013): *Análisis del tiempo y la concordancia verbales en lengua inglesa en el proceso de interlengua*. [Thesis dissertation]. URL: https://accedacris.ulpgc.es/handle/10553/10712.

Nation, Paul (2003): «The role of the first language in foreign language learning», *Asian EFL Journal,* 5(2): 1-8.

Odlin, Terence (1989): *Language transfer: cross-linguistic influence in language learning*, Cambridge: Cambridge University Press.

Offiong, Offiong Ani and Bassey A. Okon (2013): «Code switching as a countenance of language interference: The case of the EFIK bilingual», *International Journal of Asian Social Science*, 3(4): 899-912.

Peñate Cabrera, Marcos and Plácido Bazo Martínez (2002): «Input and interlanguages in the EFL classroom: a case study with primary school teachers», *Cauce: Revista de filología y su didáctica,* (25): 459-474.

Quesada, Margaret Lubbers (1995): «Theory construction in second language acquisition», *MEXTESOL Journal,* 19(1): 9-17.

Richards, Jack C. and Richard Schmidt (2002 [1985]): *Longman dictionary of teaching and applied linguistics*, 3.rd ed., London: Pearson Education.

Spada, Nina and Patsy M. Lightbown (1999): «Instruction L1 influence and developmental readiness in second language acquisition», *The Modern Language Journal*, 83: 1-22.

Taylor, Barry P. (1975): «The use of overgeneralization and transfer learning strategies by elementary and intermediate students in ESL», *Language Learning,* 25, 73-107.

Wang, Yifei (2015): «Negative transfer of mother language in English compositions by Jiujiang university students», in *SHS Web of Conferences 14*. URL: https://bit.ly/2LqvX4i.

White, Lydia (1977): «Error analysis and error correction in adult learners of English as a second language», *Working Papers in Bilingualism,* 13: 42-58.

Whitney, William Dwight (1881): «On mixture in language», *Transactions of the American Philological Association,* 12: 5-26.

Wood Wood, Manuel (2002): *La transferencia de la L1 a la L2 (español-inglés): el artículo en la lengua escrita de los alumnos en las pruebas de acceso a la universidad.* [Thesis dissertation]. URL: https://accedacris.ulpgc.es/handle/10553/2034.

Yu, Liming and Terence Odlin (eds.) (2016): *New perspectives on transfer in second language learning,* Bristol: Multilingual Matters.

Zobl, Helmut (1980): «The formal and developmental selectivity of L1 influence on L2 acquisition», *Language Learning,* 30: 43-57.