Errors in L1 and L2 University Students' Writing in English: Grammar, Spelling and Punctuation

Errores en la escritura en inglés como L1 y L2 de los estudiantes universitarios: gramática, ortografía y puntuación

CRISTINA LASTRES-LÓPEZ
UNIVERSITY OF SANTIAGO DE COMPOSTELA
GIANPAOLO MANALASTAS
UNIVERSITY COLLEGE LONDON

This paper explores the writing of students at British universities, considering both L1 and L2 speakers of English. Our aim is to identify university students' most common errors and areas of difficulty with respect to 1) grammar, 2) spelling, and 3) punctuation. For our purposes, we analysed samples of writing collected at University College London and we identified and examined frequent problems in native and non-native university student's writing. Our findings report that university students make twice the number of errors in punctuation compared to grammar or spelling; and that their typology of errors is very heterogeneous, showing differences between English L1 and L2 university students in terms of the types of errors they commit.

Keywords: *error analysis*; *writing*; *grammar*; *spelling*; *punctuation*

Este artículo explora la escritura de los estudiantes en universidades británicas, considerando tanto hablantes de inglés como L1 como L2. Nuestro objetivo es identificar los errores y áreas de dificultad más comunes de los estudiantes universitarios con respecto a 1) gramática, 2) ortografía, y 3) puntuación. Con estos fines, analizamos muestras escritas recogidas en University College London e identificamos y examinamos los problemas más frecuentes en la escritura de los estudiantes universitarios nativos y no nativos. Nuestros resultados indican que los estudiantes universitarios cometen el doble de errores en puntuación, si lo comparamos con gramática u ortografía; y que su tipología de errores es muy heterogénea, mostrando diferencias entre los estudiantes universitarios de inglés como L1 y L2 en cuanto a los tipos de errores que cometen.

Palabras clave: *análisis de errores*; *escritura*; *gramática*; *ortografía*; *puntuación*

1. INTRODUCTION

In an increasingly digital age, many students have access to word processing software with built-in correction tools that detect errors at the orthographical, grammatical and punctuation levels. Critics suggesting a decline in written standards often point to an over-reliance on this type of software that leads to a decrease in awareness of prescriptive rules, which can be a somewhat unjust generalisation of the many factors that contribute to these errors being made. In the literature regarding the language of academic discourse there have been many studies focusing on the written abilities of students at all levels of education. These studies have considered the frequency of errors made by college-level students in the US (Connors &

Lunsford, 1988), errors made by L2 speakers of English (Wilcox, Yagelski & Yu, 2014), as well as broader studies focusing on the categorisation of errors made on multiple levels (Wolfe-Quintero, Inagaki & Kim, 1998; McDonald, 2016). This paper gives an overview of a pilot study investigating the errors English L1 and L2 students make in a controlled written exercise on three levels: grammar, spelling and punctuation. The study brings together elements of previous research that have been conducted individually in an effort to identify the most commonly committed written errors. The findings will then go on to contribute to a broader effort to promote strategies and skills to students, which will help to improve their writing skills with regard to the three previously mentioned areas. An additional fourth level of errors, namely lexical, was also considered for analysis; however, as will be discussed in section 3, the data collected from participants did not present errors on this level and thus this analysis was not taken forward.

In Section 2, we give a brief overview of the literature that informed the study and the framework adopted. We also state the aims and hypotheses of our study, based on the literature search conducted. In Section 3, we then outline the methodology for collecting and analysing the data and give a profile of the data collected, before discussing the results in Section 4. The results are discussed in three parts, focusing on grammatical, orthographical and punctuation errors respectively. We then conclude and propose future directions for the research project in Section 5.

2. ARE GRAMMATICAL, ORTHOGRAPHICAL AND PUNCTUATION ERRORS CREATED EQUAL? A VIEW FROM THE LITERATURE

In the introduction to their national study of American college student writing, Lunsford and Lunsford claimed that, amongst other factors, there seemed to be a political agenda that had "bemoaned the state of student literacy and focused attention on what they [reports commissioned by the government] deem significant failures at the college level" (2008: 782, additional parenthesis by the authors). This top-down prescriptive approach to student writing has been the topic of contentious debate for some years: before the turn of the millennium, Truscott (1996) and Ferris (1999) argued opposing sides on the topic of grammar correction in L2 writing, while almost ten years later, Massey, Elliott & Johnson wrote of "the everlasting, if sporadic, debate about changing standards of literacy" (2005: 3). While these debates continue to take place, and while teaching at primary and secondary levels of education continue to heap value on the accuracy of writing, errors in student writing remains "poorly understood, and no consensus has emerged about how to approach error in writing instruction or even whether error constitutes a serious problem, in the development of writing ability" (Wilcox et al., 2014). This 'correctness' in student writing – or as Wilcox et al. put it (perhaps more accurately) – "the avoidance of error" (2014: 1075) – has been given more attention in the US where there seems to be more of a culture of linguistic correctness, if we take the popularity of spelling bee competitions as a standard of measure. Glibness aside, many non-academic publications and opinion-based articles have attempted to make the link between writing ability and performance in other aspects, such as between grammatical knowledge and performance in the workplace (Beason, 2001). Wilcox et al. (2014) make reference to Kyle Wiens' 2012 essay in Harvard Business Review: *I Won't Hire People With Poor Grammar*. Without some form of longitudinal study conducted into writing ability and these performances, it seems somewhat tenuous to make such links.

While there have been previous studies conducted focusing on errors in academic writing, the majority of these studies have either created their own analytical frameworks, or adapted existing approaches. There does not seem to be a universally agreed framework for

codifying the types of errors that can occur in writing. Indeed, a number of studies have adopted a grounded-theory approach to data analysis, given the wide variety of errors that naturally occur in writing. Duskova focused on errors in grammar and lexis, determining errors on a scale of deviation: "the degree of deviation from the normal was such as to leave no doubt of the unacceptability of the form in question" (1969: 12). This coding system relies on a level of subjectivity that is tempered by the knowledge and experience of the coder, a subjectivity that perhaps led Connors and Lunsford to state: "we had no taxonomy of errors we felt we could trust" (1988: 399), when they considered frameworks for their analysis of formal errors in college writing. They went on to acknowledge that the taxonomy they generated themselves was one that was bound by their own cultural perception. Lunsford and Lunsford (2008) then adapted this taxonomy, following the same grounded-theory approach to reorder the hierarchy of errors committed in students' written essays. Over subsequent years other papers investigating written errors in student writing have adapted the Lunsford and Lunsford framework to suit the specific needs of their research: Cook (2010) modified it when studying errors in grammar and usage made by university students, while Wilcox et al. (2014) increased the number of categories when looking at errors made by English L1 and L2 adolescents' writing.

While these methods of analysing errors by hand continued to flourish, alternative methods emerged that began to use electronic corpora as a means of sweeping through data more rapidly. Mediero Duran and Robles Baena (2012: 66) gave broader macro categorisations to the errors they discovered including lexical, grammatical, punctuation, pragmatic and phrasing. These overarching categories were then split into several subcategories, alluding to the wide variety encountered. McDonald (2016) also adopted these categories for her study on errors made by Spanish students' written English.

Taking into account the concerns surrounding the subjectivity of any coding system, the framework adopted for this project takes a combined approach of broader categorisation per Mediero Durán and Robles Baena (2012), with the inductive approach put forward by Connors and Lunsford (1988).

3. DATA AND METHODOLOGY

This paper intends to be a pilot study in the investigation of errors in the writing of university students in the United Kingdom. Partly-stemming from the project 'Improving the English spelling of students' (under funding from UCL ChangeMakers, Centre for Advancing Learning and Teaching, University College London), our aims have broadened from spelling to other aspects which have been found to be more problematic in the writing of university students, such as grammar and punctuation, as discussed in prior research deriving from the aforementioned project (Lastres-López, forthcoming). Therefore, our objective is to extend our analysis in order to identify the most common errors in grammar, spelling, and punctuation in university students' writing. In particular, we aim to give answers to the following questions: 1) what kind of grammatical, orthographical and punctuation errors do university students make in their writing; 2) which of the three metacategories of errors is more frequent; and 3) are there differences between L1 and L2 speakers of English?

For our purposes, we collected samples of writing from 60 university students at University College London (UCL). UCL is a world-leading multidisciplinary university with over 35,000 students from 150 different countries, located in Bloomsbury, in the heart of London. Since we wanted our data to be representative of the reality of the institution, we collected a balanced sample that reflected the internationality and multidisciplinarity of the university. Thus, we examined data from male and female students from different degrees

(Archaeology, Classics, Engineering, English, French, Genetics, History, Linguistics, Literature, Management, Mathematics, Medicine, Pharmacology, Philosophy, Psychology, Physics, and Sustainable Heritage, among others), years[1] (first-, second- and third-year undergraduates, and Master students) and nationalities (American, Australian, Bahraini, British, Canadian, Chilean, Chinese, Dutch, German, Greek, Indonesian, Irish, Italian, Lebanese, and Polish). Since one of our primary aims is to investigate the differences in the writing of native and non-native students, we collected samples of writing from 30 L1 and 30 L2 speakers of English so that the results could be comparable. The L2 speakers are advanced and proficient users of the language (C1-C2 levels of the Common European Framework of Reference for Languages), who have conformed to the strict English language policy of admissions of UCL. Among the non-native speakers we have examined data from students with 10 different L1s (Arabic, Chinese, Dutch, Greek, German, Indonesian, Italian, Japanese, Polish and Spanish), and within the L1 speakers of English we analysed different national language varieties (American, Australian, British, Canadian and Irish).

British English grammar, spelling and punctuation conventions were assumed as the standard, given the location of the university; however, a prescriptivist approach was not considered and variation resulting from different national language varieties (c.f. BrE 'I've already done it'/AmE 'I already did it'; BrE 'colour'/ AmE 'color') was not categorised as a mistake. In line with the text type taxonomy proposed in the British component of the *International Corpus of English* (ICE-GB) (Nelson, Wallis & Aarts 2002) and also carried forward in other national varieties of the ICE corpora, we have considered that timed essays could produce different results from untimed ones, the former being more similar to exams. Thus, we asked students to produce a very short piece of writing in 10 minutes. University experiences were selected as the topic of students' writing, since we considered that familiar topics could encourage students to write more content more freely, and would allow us to examine "real language in use" (Milroy, 1992: 66). The instructions for the task mentioned that students could write about anything related to their university experience. A non-exhaustive list of suggestions for topics of writing included: what they liked or disliked about their university, changes they would like to see implemented at it, or anecdotes related to their experience at university. We briefly alluded to lexical analysis in the introduction, and will expand further before proceeding. As can be gathered from the aforementioned writing topics, participants were asked to write a brief piece that was reflective but not particularly challenging in terms of content. It was therefore expected that, combined with the level of participant language proficiency, there would be little to zero errors committed at the lexical level. We deemed this justification enough for not analysing lexical errors in this project.

A database was then created, analysing manually the errors made by students and classifying them in three metacategories (grammar, spelling and punctuation), with subsequent more fine-grained subcategories for each type. The identification and classification of the errors was carried out in two rounds of analysis. In an initial phase, the team in charge of the analysis of the written samples was composed of three team members and the two authors of this paper, who identified and gave a preliminary classification of the errors. The analysis was later carried forward in more detail by the two authors of this paper, providing a more in-depth classification of the subcategorisation of errors. At both stages of the process, misleading or difficult cases to classify were discussed among the team members, so that the analysis could be sound and reliable. As will be discussed in Section 4.2, grammatical errors were classified depending on whether errors were found at the word level or were associated with the inflection of the word. In the former category, we classified errors according to the part of speech (noun, verb, preposition etc.) involved. Spelling errors,

---

[1] In designing the experiment, we have considered that the level of study is more relevant for our purposes than the age of the participants, since we are looking at a snapshot of university students.

presented in Section 4.3, were categorised in four broad groups, with further subcategories, namely: 1) phonologically-related errors; 2) absence or presence of a grapheme; 3) incorrect choice of grapheme; and 4) splitting of the word. Finally, punctuation errors, which will be examined in Section 4.4, were classified depending on whether there was an incorrect use of a comma, period, apostrophe or capitalisation, or the absence of punctuation marks. The analysis of punctuation was the most difficult to carry out, since the boundaries between correctness and stylistic preference are sometimes fuzzy. An additional level of analysis was therefore implemented to punctuation, to indicate the errors identified which could have multiple options for correction.

4. RESULTS AND DISCUSSION

4.1. *Frequencies*

The manual analysis of the 60 writing samples reveals a total of 175 errors, a mean of around 3 errors per participant in an average text of 109 words. Out of the total number of errors, 55 correspond to grammar, 32 to spelling and 88 to punctuation, as summarised in Table 1 below. This shows that errors in university students' writing are unequally distributed across the three metacategories: almost half of them refer to punctuation exclusively, whereas the other half is, again, unequally distributed between grammar (31.4%) and spelling (18.3%).

*Table 1: **Frequency of grammar, spelling and punctuation errors***

Metacategory of error	N	%
Grammar	55	31.4%
Spelling	32	18.3%
Punctuation	88	50.3%
Total	175	100%

Turning now to the differences between native and non-native speakers, our data reveals that they show almost identical total error frequencies. A closer analysis of the data shows, however, distinct patterns in the three metacategories. While more grammatical errors were found in L2 writing (61.8%), orthography displays a reversed pattern, with more spelling errors (59.4%) in texts written by English native speakers. Punctuation, in which more errors were attested in total, seems to be the category with less variation between native and non-native speakers of English, with very similar figures for the two groups. Absolute and relative frequencies of the errors made in the three metacategories by L1 and L2 speakers are summarised in Table 2 below.

*Table 2. **Frequency of errors made by native and non-native speakers by metacategory***

Metacategory of errors	Native	Non-native	Total
Grammatical errors	21 (38.2%)	34 (61.8%)	55 (100%)
Spelling errors	19 (59.4%)	13 (40.6%)	32 (100%)
Punctuation errors	47 (53.4%)	41 (46.6%)	88 (100%)
Total number of errors	87 (49.7%)	88 (50.3%)	175 (100%)

Our data reveals that more than 90% of the students made at least one mistake in the short text. Table 3 below illustrates the relationship between number of errors and number of participants. It is observed that, per metacategory, most students make either none or a single mistake; whereas in total the number of participants who do not have errors in their writing decreases to only 5, and a single error becomes the most frequent scenario, followed surprisingly, by five or more errors. This shows that much variation occurs at the individual

level. For a more fine-grained account detailing the number of errors made by each participant in each metacategory, see Appendix 1.

Table 3: Relationship between number of errors and number of participants

Number of errors	Number of participants			
	Grammar	Spelling	Punctuation	Total
0	21	41	30	5
1	28	13	9	24
2	7	3	8	9
3	3	1	7	4
4	1	1	1	6
5 or more	0	1	5	12
Total	60			

4.2 *Types of grammatical errors*

Wilcox et al.'s (2014) study on adolescent writing indicate that incorrect verb inflection is one of the most common errors among students. Incorrect verb inflection most frequently derives from errors in grammatical agreement, as in (1) and (2). Subject-verb agreement was also identified as problematic in sentence construction at earlier academic levels, such as students in secondary education (Massey et al., 2005: 56).

(1) *There needs to be improvements [#8N]. Suggested correction: There need to be improvements.[2]

(2) *UCL staff and departmental tutor advices and helps [#7N]. Suggested correction: UCL staff and the departmental tutor advise and help.

Contrary to the expectation, our data shows that this is not the most persistent grammatical error in university students' writing. Out of the 55 grammatical errors identified in the samples analysed, only 7 of these errors correspond to inflections, as shown in Table 4 below, which displays the types of errors encountered – indicating whether such errors occur at the word level or are related to inflectional morphology – and their absolute and relative frequencies.

Table 4: Classification of grammatical errors and frequency of occurrence

Type of grammatical error		Frequency
Missing or wrong word	Verb	13 (23.6%)
	Determiner	11 (20.0%)
	Preposition	8 (14.6%)
	Pronoun	8 (14.6%)
	Adjective	2 (3.6%)
	Noun	1 (1.8%)
	Conjunction	1 (1.8%)
	Relativiser	1 (1.8%)
Missing or wrong inflection		7 (12.7%)
Other		3 (5.5%)
Total		55 (100%)

As shown in Table 4 above, most errors occur at the word level (n=45, 81.8%). Among them, the most frequent concern either verbs or determiners, two categories generally

---

[2] As is conventional, incorrect sentences and shorter fragments are preceded by an asterisk. The coding system in square brackets indicates participant number, followed by N, for native speakers, or NN for non-native speakers.

considered difficult for both native and non-native speakers (Thomas, 1989; Hirsh-Pasek & Golinkoff, 2006; Halford, Wilson, Andrews & Phillips, 2014; Díez-Bedmar, 2015).

Verbs pose a number of problems for university students. The errors encountered in our sample, show that the correlation of verbal tenses is still difficult for university students, as illustrated in example (3) below, in which the student uses the modal verb *would* to hypothesize, but does not establish the correct correlation with the rest of the verbs in the same sentence. Another frequent error is the use of the incorrect verb form, particularly with gerunds, as in (4) and infinitives, as in (5). Errors with verbs are much more frequent among L2 speakers of English (76.9%), as compared with L1 speakers (23.1%).

(3) *Anything I'd like to change is that I want to do more reading if the classes can start again [#39NN]. Suggested correction: Anything I'd like to change is that I'd want to do more reading if the classes could start again.
(4) *[…] is worth to study there [#10NN]. Suggested correction: is worth studying there.
(5) *I had the opportunity to met [#27NN]. Suggested correction: I had the opportunity to meet.

Within the determiner category, most of the errors correspond to article usage, both definite and indefinite, as shown in (6) and (7) below. The determiner category of errors represents 20% of the total number of grammatical errors, with approximately half corresponding to native speakers and the other half to non-native speakers. Although the article system – and by extension the category of determiners – poses problems for L2 speakers of English, it should be borne in mind that the L1 plays a central role here. For example, Díez-Bedmar and Pérez-Paredes (2012: 153) found that the article system is not a major problem in the writing of Spanish learners of English, probably because both systems of articles are similar in the two languages. In contrast, speakers of languages which do not even have an article paradigm, such as Chinese and Japanese, find it much more difficult to learn the correct use of articles (Robertson, 2000; Butler, 2002). Since our data aim to be a balanced sample reflecting the reality of the university of our pilot study, UCL, our data are also highly varied in terms of the different L1s of the students. It is therefore difficult to draw conclusions regarding differences in non-native speakers with presence or absence of article paradigms in their L1s.

(6) *takes up majority of my time [#8N]. Suggested correction: takes up the majority of my time.
(7) *to live in such expensive city [#30NN]. Suggested correction: to live in such an expensive city.

Errors with prepositions and pronouns are also considerably frequent. Regarding prepositions, both the absence of any preposition, as in (8), and the incorrect choice of preposition, as in (9) are frequent. A higher proportion of the errors related to prepositions correspond to non-native speakers (62.5%). Equal in terms of frequency are errors related to the omission of pronouns. Such omission, exemplified in (10), is characteristic of the language of adolescents, deriving from the use of instant messages or textspeak (Crystal, 2006; Tagliamonte, 2016). This tendency to abbreviate words and phrases can also lead to a tendency to suppress certain word classes, typically those considered superfluous in communication to transmit a message successfully. Examples of pronoun omission are presented in (10) and (11). In (10), we illustrate the most common pattern attested for pronoun omission in our sample, that is to say, omission of the pronoun functioning as subject in a main clause; but less frequent patterns, such as pronoun omission in embedding, were

also detected, as illustrated in (11), in which the pronoun is omitted in the subordinate clause. While the pattern in (10), as we said, is frequent in adolescent language, ungrammatical instances such as (11) have only been found in L2 speakers of English.

(8) *I have enjoyed living in London partially because how small and familiar UCL feels [#36N]. Suggested correction: I have enjoyed living in London partially because of how small and familiar UCL feels.

(9) *The overall experience of me and the majority of my colleagues [#33NN]. Suggested correction: The overall experience for me and the majority of my colleagues.

(10) *Like the fact that there are so many different cultures/nationalities/ethnicities here at UCL [#6N]. Suggested correction: I like the fact that there are so many different cultures/nationalities/ethnicities here at UCL.

(11) *I felt that couldn't express myself [#13NN]. Suggested correction: I felt that I couldn't express myself.

4.3 *Types of spelling errors*

Although roughly 90% of the words in English have a predictable orthography (Crystal, 2005), the spelling system of English is still difficult for both native and non-native speakers (Crystal, 2012; Horobin, 2013). One of its main complexities resides in the lack of one-to-one correspondence between graphemes and phonemes (Culpeper, 1997: 18; Van Gelderen, 2006: 15), which, together with a large number of homophonic words, may lead to spelling errors of various types.

Previous research on errors in students' writing considers orthographical errors as a homogeneous category, without paying much attention to different subcategories of spelling mistakes (Lunsford & Lunsford, 2008: 795; Mediero Durán & Robles Baena, 2012: 66; Wilcox et al., 2014: 1082). In line with other studies that have concentrated on spelling only, we consider that spelling errors are a heterogeneous category and we aim to offer a detailed taxonomy of the types of orthographical errors found in our data. Following Elliott and Johnson (2008), our categorisation of spelling mistakes is based on the errors attested, rather than on already existing taxonomies. This allows a more fine-grained analysis of our sample, but does not exclude from the typology the existence of other possible categories which are not included here just because they were not encountered in the sample. The types of spelling mistakes encountered, together with their frequency, are summarised in Table 5 below.

*Table 5: **Classification of spelling errors and frequency of occurrence***

Type of spelling error		Frequency
Phonologically-related errors	Homophones	8 (25.1%)
	Minimal pair	5 (15.6%)
Absence or presence of a grapheme	Double consonant not doubled	3 (9.4%)
	Absence of vowel	3 (9.4%)
	Addition of consonant	1 (3.1%)
	Addition of vowel	1 (3.1%)
	Single consonant doubled	1 (3.1%)
Incorrect choice of grapheme	Incorrect vowel	1 (3.1%)
	Incorrect consonant	1 (3.1%)
Splitting of the word		3 (9.4%)
Other		5 (15.6%)
Total		32 (100%)

Despite the low frequencies encountered, it should be remarked that 19 of the 60 participants (>30 %) showed at least one mistake in their writing (see Table 3 above). These results indicate a decline in the number of spelling errors made by students at university compared with previous educational levels, such as secondary education, as studied by Elliott and Johnson (2008). Out of the total number of spelling errors encountered in our sample, 59.4% are made by native speakers and 40.6% by non-native students. As can be seen, most of the spelling errors are phonologically related: they are either homophones, as exemplified in (12), (13) and (14); or minimal pairs, as in (15). Although mistakes of this type are committed both by L1 and L2 speakers of English, it should be remarked that 61% of these sound-related errors correspond to native speakers.

(12) *Its been demanding to step up from A levels onto a course which really requires wider reading. [#60N]. Suggested correction: It's been demanding […].
(13) *We where obligated to attend classes at a hospital [#33NN]. Suggested correction: We were obligated […].
(14) *I love UCL - it offers great opportunities through societies to volunteering fares. [#3N]. Suggested correction: […] volunteering fairs.
(15) *the amacing opportunity [#54N]. Suggested correction: the amazing opportunity.

Other frequent spelling mistakes include splitting of words, as in (16); or errors deriving from the absence of a required grapheme or the presence of an unrequired one, such as double consonants not doubled, as in (17); and absence of a vowel, as in (18). The other types of errors listed in Table 5 were reported to have a very low frequency in our sample.

(16) *I would define my self as someone who lives in London […] [#53N]. Suggested correction: I would define myself […].
(17) *[…] a structure intended to accomodate need. [#52N]. Suggested correction: accommodate.
(18) *I believe that my lecturers are very knowledgable and supportive [#20NN]. Suggested correction: knowledgeable.

4.4 *Types of punctuation errors*

Analysis of punctuation errors was perhaps the most subjective stage of the project, given the various interpretations that could be placed on a phrase or sentence. It is somewhat impossible – or at the very least, highly presumptuous – to deduce the student's mental process based purely on what they have written down. Beason suggests that: "errors must be defined not just as textual features breaking handbook rules but as mental events taking place outside the immediate text" (2001: 35), which seems a particularly pertinent point in the context of punctuation. The relative dearth of literature focusing on punctuation errors compared to the number of studies on grammar and spelling could be interpreted as a reflection of the difficulties encountered when codifying punctuation error. Given this relative lack of a precedent, we adopted the inductive approach taken to analysing grammar and spelling, with the caveat that there may have been differences of agreement between researchers.

Table 6: Classification punctuation errors and frequency of occurrence

Type of punctuation error	Frequency
Incorrect use of comma: absence	34 (38.7%)
Incorrect capitalisation: proper noun	9 (10.2%)
Incorrect use of comma: unnecessary	8 (9.1%)
Incorrect use of comma: splicing	7 (8%)
Incorrect capitalisation: unnecessary	6 (6.8%)
Incorrect use of period: absence	6 (6.8%)
Absence of apostrophe in contraction	5 (5.7%)
Type of punctuation error	**Frequency**
Absence of apostrophe for genitive	3 (3.4%)
Incorrect use of period: comma sufficient	3 (3.4%)
Incorrect use of period: space missing	3 (3.4%)
Absence of hyphen for modifier	1 (1.1%)
Too broad to categorise	3 (3.4%)
Total	88 (100%)

From the initial pass through the data it was somewhat surprising to encounter more punctuation errors than orthographic or grammatical errors. Of the 60 participants, exactly 30 (50%) made punctuation errors, with an average of 2.9 errors made between them (the highest number of punctuation errors made by a single participant was 9, committed by two participants). Given the previous comment regarding the relevant dearth of publications focusing on punctuation errors, the expectation was to find fewer punctuation errors: from Table 6 we not only found more errors on this level than on the other two, the number of subcategories was also higher, with a count of 12. The total number of punctuation errors committed were also somewhat surprising: 53.4% (n=47) were committed by native English speakers, while 46.6% (n=41) were committed by L2 English speakers. Over half of these errors concerned the usage of commas, with the absence of a comma the most common error identified, as shown in (19) – although caveats to this will be discussed later. Capitalisation of proper nouns also posed an issue (20), while unnecessary capitalisation was also identified (21). Misuse of apostrophes was also noted, in both contractions (22) and in genitive forms (23), while periods were occasionally omitted (24).

(19) *However I have made some great friends and learnt so much in the past 18 months [#7N]. Suggested correction: However, I have made some great friends and learnt so much in the past 18 months.

(20) *The gower street ivory tower should cede to a structure intended to accommodate need. [#52N]. Suggested correction: The Gower Street ivory tower should cede to a structure intended to accommodate need.

(21) *UCL enjoys a high reputation around the World. [#44NN]. Suggested correction: UCL enjoys a high reputation around the world.

(22) *Its not that its not very interesting its just I'd prefer to move towards diverting from retrieving knowledge from the classical stone to actually striving to learn something more. [#52N]. Suggested correction: It's not that it's not very interesting it's just I'd prefer to move towards diverting from retrieving knowledge from the classical stone to actually striving to learn something more.

(23) *It is a one year Masters degree in Project Management [#33NN]. Suggested correction: It is a one year Master's degree in Project Management.

(24) *People are really friendly and you get to meet a diverse range of students [#8N]. Suggested correction: People are really friendly and you get to meet a diverse range of students.

The issue of subjectivity and stylistic preference posed an issue for the research team in the punctuation analysis, particularly in the instances involving commas. The second pass through the dataset resulted in a closer inspection as to whether the analysis could have been the result of differing stylistic approaches. While analysis of spelling errors was relatively straightforward with the use of the OED as guide for standard orthography, there is no equivalent punctuation guideline. Moreover, suggesting corrections for errors in punctuation use is made more difficult by the variety of punctuation marks that can be used in a given situation, which carry different nuances or insights into the cognitive processes informing the participant. The research team therefore aimed to highlight errors that deviated from a continuum of acceptability, while accepting that their suggestions for corrections were susceptible to their own stylistic preference. The table below presents an additional level of analysis applied to punctuation errors, with a column added indicating how many of the errors identified had multiple options for correction, without drastically altering the wording or order of the original sentence.

Table 7: ***Classification of punctuation errors and frequency of errors, showing count of multiple corrections***

Type of punctuation error	Frequency	Multiple correction options
Incorrect use of comma: absence	34 (38.7%)	22
Incorrect capitalisation: proper noun	9 (10.2%)	0
Incorrect use of comma: unnecessary	8 (9.1%)	8
Incorrect use of comma: splicing	7 (8%)	7
Incorrect capitalisation: unnecessary	6 (6.8%)	0
Incorrect use of period: absence	6 (6.8%)	0
Absence of apostrophe in contraction	5 (5.7%)	0
Absence of apostrophe for genitive	3 (3.4%)	0
Incorrect use of period: comma sufficient	3 (3.4%)	2
Incorrect use of period: space missing	3 (3.4%)	0
Absence of hyphen for modifier	1 (1.1%)	0
Too broad to categorise	3 (3.4%)	3
Total	88 (100%)	39

As Table 7 indicates, errors involving commas presented multiple correction options, including corrections regarding its absence ((25a), (25b) and (25c)) and comma splicing ((26a) and (26b)).

(25) *I like spending time with people, my family and friends so I plan in such time to help break away from the course. [#7N]. Suggested corrections:
      a. I like spending time with people, my family and friends, so I plan in such time to help break away from the course.
      b. I like spending time with people – my family and friends – so I plan in such time to help break away from the course.
      c. I like spending time with people (my family and friends) so I plan in such time to help break away from the course.

The options proposed for the error identified in (25) vary in their nuance. Having identified the non-restrictive clause within, it is possible to use either the comma, the em dash or parentheses to place that information in a more formal, more distant or subtle frame. Stylistics therefore plays quite a crucial rule in dictating the punctuation to be used, and we cannot glean this information from the clues left behind.

(26) *The high number of young professors is inspiring as well, it gives you a sense of progress and recognition. [#17NN]. Suggested corrections:

a. The high number of young professors is inspiring as well; it gives you a sense of progress and recognition.
b. The high number of young professors is inspiring as well. It gives you a sense of progress and recognition.

Identifying the comma splice as an error can veer into somewhat contentious territory, given it may be perceived as a stylistic preference in one extreme, to a grammatical error on the other. The preference of a semi-colon or a period are just two of the simpler methods to correct a comma splice, however it may seem overly prescriptive to pull up every instance of a comma splice without knowing the full context of the student's mental process when writing.

5. CONCLUDING REMARKS

This paper has investigated university students' writing in English, examining the differences in errors made by L1 and L2 speakers of English, and analysing for our pilot study a sample of data collected in a university setting in London. Our study reveals differences in the errors made by university students, both in terms of the types of mistakes committed and in relation to the type of speaker (native *versus* non-native). While the total error frequency is similar in both groups, a closer analysis of the three metacategories considered – grammar, spelling, and punctuation – reveals that half of the errors encountered in university students' writing correspond to punctuation, the other half being divided between grammar and spelling. Quantitatively, we have observed not only that native and non-native speakers commit errors at similar error frequencies when they reach high levels of proficiency, as is the case of L2 speakers in an English-speaking university; but, qualitatively, certain aspects are more problematic for certain groups, as is the case with the writing of homophonic words for L1 speakers, for instance.

Regarding grammar, the highest proportion of errors are related to missing or incorrect uses of verbs and determiners, particularly articles. These errors, although present in both groups, have been found to be more prominent among non-native speakers. Conversely, errors in spelling occur most frequently among L1 speakers of English, in particular those related to either homophones or minimal pairs just distinguished by a phoneme. While grammar and spelling favour differences between native and non-native usage, punctuation, the category in which we found more errors, seems to be equally problematic for both L1 and L2 speakers of English. Our results show almost identical frequencies for punctuation errors made by native and non-native speakers, with the absence of commas being the dominant error. This higher frequency of errors in punctuation may probably be due to the fact that it is sometimes more difficult to establish rules with respect to punctuation, or at least that such rules tend to be more flexible, with some usages being related to stylistic preferences. The availability of options when using punctuation must also be considered when analysing these types of errors.

At this relatively early stage of the project, it is too early to make definitive statements about the didactic implications and opportunities for providing students with specific guidance on the types of errors made. Although the results from our study should be confirmed against larger samples of data, we hope to have opened new avenues for future research in university students' writing. We hope to have presented a promising analytical approach that can also be considered for students' formal writing, such as essays and other assessments. Findings from other such projects could combine to provide a more robust understanding of the errors university students make in their writing, and how to avoid committing them.

REFERENCES

Beason, L. (2001). Ethos and error: How business people react to errors. *College Composition and Communication,* 53(1), 33-64.

Butler, Y. G. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, 24, 451-480.

Connors, R. J. & Lunsford, A. A. (1988). Frequency of formal errors in current college writing, or, Ma and Pa Kettle do research. *College Composition and Communication,* 39(4), 395-409.

Cook, D. (2010). The location of error: Reflections on a research project. *Teaching English in the Two-Year College,* 38(1), 22-33.

Crystal, D. (2005). *The Stories of English.* London: Penguin.

Crystal, D. (2006). *Language and the Internet*. Cambridge: Cambridge University Press.

Crystal, D. (2012). *Spell It Out: The Singular Story of English Spelling.* Suffolk: Profile Books.

Culpeper, J. (1997). *History of English*. London: Routledge.

Díez-Bedmar, M. B. (2015). Article use and criterial features in Spanish EFL writing: A pilot study from CEFR A2 to B2 levels. In M. Callis & S. Götz (Eds.), *Learner Corpora in Language Testing and Assessment* (pp. 163-190). Amsterdam: John Benjamins.

Díez-Bedmar, M. B. & Pérez-Paredes, P. (2012). A cross-sectional analysis of the use of the English article system in Spanish learner writing. In Y. Tono, Y. Kawaguchi, & M. Minegishi (Eds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research* (pp. 139-158). Amsterdam: John Benjamins.

Dusková, L. (1969). On sources of errors in foreign language learning. *International Review of Applied Linguistics,* 7(1), 11-36.

Elliott, G. & Johnson, N. (2008). *All the right letters - just not necessarily in the right order. Spelling errors in a sample of GCSE English scripts*. Paper presented at the British Educational Research Association Conference. Edinburgh, September 2008. Retrieved from

http://www.cambridgeassessment.org.uk/Images/109772-all-the-right-letters-just-not-necessarily-in-the-right-order.-spelling-errors-in-a-sample-of-gcse-english-scripts.pdf

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1-11.

Halford, G. S., Wilson, W. H., Andrews, G. & Phillips, S. (2014). *Categorizing Cognition: Toward Conceptual Coherence in the Foundations of Psychology*. Cambridge: MIT Press.

Hirsh-Pasek, K. A. & Golinkoff, R. M. (2006). (Eds.). *Action meets Word: How Children learn Verbs*. Oxford: Oxford University Press.

Horobin, S. (2013). *Does Spelling Matter?* Oxford: Oxford University Press.

Lastres-López, C. (forthcoming). Examining English spelling at British universities. In M. Díaz Ferro, J. Diz Ferreira, A. Pérez Pérez & A. Varela Suárez (Eds.), *Title pending*. Lugo: AXAC.

Lunsford, A. A. & Lunsford, K. J. (2008). "Mistakes are a fact of life": A national comparative study. *College Composition and Communication*, 59(4), 781-806.

Massey, A. J., Elliott, G .L. & Johnson, N. K. (2005). Variations in aspects of writing in 16+ English examinations between 1980 and 2004: Vocabulary, spelling, punctuation, sentence structure, non-standard English. *Research Matters, Special Issue 1*. Retrieved from http://www.cambridgeassessment.org.uk/Images/340982-research-matters-special-issue-4-aspects-of-writing-1980-2014.pdf

McDonald, P. (2016). "We All Make Mistakes!". Analysing an Error-coded Corpus of Spanish University Students' Written English. *Complutense Journal of English Studies,* 24, 103-129.

Mediero Durán, M. E. & Robles Baena, A. (2012). Exploring the errors made by EFL students at the university level. In Álvarez Mosquera, P., Fernández Gil, M. J., Borham Puyal, M., Díez García, M. J., Bautista Martín, S., Ruano García, J. & B. García Riaza (Eds.). *Research challenges for Anglophone studies in the 21st century* (pp. 63-75). Salamanca: Ediciones Universidad de Salamanca.

Milroy, J. (1992). *Linguistic Variation and Change: On the Historical Sociolinguistics of English*. Oxford: Blackwell.

Nelson, G., Wallis, S. & Aarts, B. (2002). *Exploring Natural Language: The British Component of the International Corpus of English*. Amsterdam: John Benjamins.

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research,* 16(2), 135-172.

Tagliamonte, S. (2016). *Teen Talk: The Language of Adolescents*. Cambridge: Cambridge University Press.

Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics,* 10(3), 335-355.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning,* 46(2), 327-369.

Van Gelderen, E. (2006). *A History of the English Language*. Amsterdam: John Benjamins.

Wilcox, K.C., Yagelski, R. & Yu, F. (2014). The nature of error in adolescent student writing. *Reading and Writing*, 27(6), 1073-1094.

Wolfe-Quintero, K., Inagaki, S. & Kim, H.Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii Press.

APPENDIX 1. FREQUENCIES OF ERRORS PER PARTICIPANT

Participant	Type of speaker	Grammar	Spelling	Punctuation	Total
1	L2	1	1	2	4
2	L2	0	0	6	6
3	L1	0	1	0	1
4	L1	1	0	0	1
5	L2	0	0	0	0
6	L1	1	0	0	1
7	L1	2	1	6	9
8	L1	3	1	3	7
9	L1	0	0	1	1
10	L2	1	0	0	1
11	L1	0	0	0	0
12	L1	0	0	0	0
13	L2	2	1	0	3
14	L2	3	0	1	4
15	L1	1	0	0	1
16	L1	0	0	9	9
17	L2	0	0	1	1
18	L2	2	0	0	2
19	L1	1	0	1	2
20	L2	0	1	0	1
21	L1	0	0	2	2
22	L1	0	0	3	3
23	L1	0	1	3	4
24	L1	0	0	0	0
25	L1	0	0	0	0
26	L1	0	0	2	2
27	L2	4	2	2	8
28	L1	1	0	0	1
29	L1	2	1	3	6
30	L2	1	2	2	5
31	L2	1	0	1	2
32	L2	1	0	3	4
33	L2	3	3	8	14
34	L2	0	0	4	4
35	L1	1	0	0	1
36	L1	1	0	0	1
37	L2	0	1	2	3
38	L2	1	0	0	1
39	L2	1	0	0	1
40	L2	1	0	0	1
41	L2	1	0	1	2
42	L2	2	0	0	2
43	L2	1	1	1	3
44	L2	2	0	2	4
45	L2	1	0	0	1
46	L2	1	0	0	1
47	L2	1	0	0	1
48	L2	1	0	0	1
49	L2	1	0	1	2
50	L2	0	0	1	1
51	L1	1	0	0	1
52	L1	0	6	9	15
53	L1	0	4	2	6
54	L1	1	1	0	2

Participant	Type of speaker	Grammar	Spelling	Punctuation	Total
55	L2	1	1	3	5
56	L1	1	0	0	1
57	L1	1	0	0	1
58	L1	0	1	0	1
59	L1	1	0	0	1
60	L1	2	2	3	7