



Patterns of errors in texts written by Costa Rican university English learners: A corpus-aided study

Patrones de errores en textos escritos por aprendices universitarios de inglés en Costa Rica: Un estudio asistido por corpus

Volumen 23, Número 1

Enero - Abril

pp. 1-31

Marisela Bonilla López

Citar este documento según modelo APA

Bonilla López, Marisela. (2023). Patterns of errors in texts written by Costa Rican university English learners: A corpus-aided study. *Revista Actualidades Investigativas en Educación*, 23(1), 1-31. Doi. <https://doi.org/10.15517/aie.v23i1.51485>

Patterns of errors in texts written by Costa Rican university English learners: A corpus-aided study

Patrones de errores en textos escritos por aprendices universitarios de inglés en Costa Rica:
Un estudio asistido por corpus

Marisela Bonilla López¹

Abstract: The present corpus-aided study sought to identify the grammatical and non-grammatical second language (L2) error patterns of Costa Rican university English learners at all academic levels of a public university. Specifically, a total of 360 English as a foreign language learners, who were enrolled in the B.A in English or B.A. in English Teaching during the second semester of 2019, took the Quick Oxford Placement Test to ascertain their English Proficiency level and composed an argumentative text to elicit their written errors. Results from the placement test showed that the participants' proficiency level ranged between B1 (low intermediate) and C1 (low advanced). In addition, the quantitative nature of the study required not only converting the handwritten compositions with a speech recognition software but also identifying and tagging all L2 errors with a tagging system. Analyses of a statistical software for data management revealed that the learner corpus contained a total of 33 L2 error patterns, which were classified as follows: 17 grammatical, 10 stylistics, and 6 lexical. Main descriptive findings indicated that although some error frequencies lowered to the point of having none as learners advanced in the major (e.g., capitalization and superlatives), other linguistic problem areas persisted all throughout (e.g., word form errors, fragments, and word order). Concluding remarks highlight that because the error frequencies of some L2 error categories still ranked high over time, learners' L2 knowledge of lexical, syntactic, morphological, and stylistic domains could need more expert input (in the form of explicit instruction and/or feedback) depending on the complexity of the target structure.

Keywords: foreign languages, university students, Linguistic research, Writing

Resumen: El presente estudio asistido por corpus buscó identificar los patrones de error gramaticales y no gramaticales de la segunda lengua (L2) de los aprendices de inglés en todos los niveles académicos de una universidad pública. Específicamente, para determinar el nivel de competencia en inglés y para obtener los errores escritos, un total de 360 estudiantes de inglés como lengua extranjera, matriculados en el Bachillerato en inglés o el Bachillerato en la Enseñanza de Inglés durante el segundo semestre de 2019, completaron el Quick Oxford Placement Test y escribieron un texto argumentativo, respectivamente. Los resultados de la prueba de ubicación mostraron que el nivel de competencia de las personas participantes osciló entre B1 (intermedio bajo) y C1 (avanzado bajo). La naturaleza cuantitativa del estudio requirió no solo convertir las composiciones escritas a mano con un software de reconocimiento de voz, sino también identificar y etiquetar todos los errores L2 con un sistema de etiquetado. Los análisis de un software estadístico para la gestión de datos revelaron que el corpus contenía un total de 33 patrones de error de L2, los cuales se clasificaron de la siguiente manera: 17 gramaticales, 10 estilísticos, y 6 léxicos. Los hallazgos descriptivos principales indicaron que, aunque algunas frecuencias de error se redujeron hasta el punto de no tener ninguno a medida que el estudiantado participante avanzaba en la carrera (por ejemplo, mayúsculas, superlativos, modales y cuantificadores), otras áreas de problemas lingüísticos persistieron independientemente del nivel académico (por ejemplo, errores de forma de palabras, fragmentos, y orden de las palabras). Las observaciones finales destacan que debido a que las frecuencias de error de algunas categorías no disminuyeron con el tiempo, el conocimiento de los dominios léxicos, sintácticos, morfológicos y estilísticos de L2 de los estudiantes podría necesitar más aportes de expertos (en forma de instrucción explícita y/o retroalimentación) dependiendo de la complejidad de la estructura de meta.

Palabras clave: Lengua extranjera, Estudiante universitario, Investigación lingüística, Expresión escrita

¹ Docente propietaria e investigadora de la Escuela de Lenguas de la Universidad de Costa Rica, San José, Costa Rica. Doctorado en Lingüística de la KU Leuven en Bélgica. Orcid: <https://orcid.org/0000-0002-1194-7721>

Dirección electrónica: marisela.bonilla@ucr.ac.cr

Artículo recibido: 22 de junio, 2022

Enviado a corrección: 31 de agosto, 2022

Aprobado: 24 de octubre, 2022

1. Introduction

The latest edition of the world's largest ranking of countries by English skills, carried out by *EF EPI* (Education First English Proficiency Index, 2021), indicates that out of 112 countries, Costa Rica ranks 44 and has moderate English proficiency with a score of 553 (vis-à-vis the Netherlands, which has very high proficiency and ranks 1st on the list with a score of 663). Some may argue that a survey of this type cannot by any stretch paint a completely accurate picture due to sampling procedures², yet it certainly shows a preview of a larger reality: that the English proficiency level of Costa Rican pupils and youngsters seems to be stagnant. In fact, recent news reports pose a problem that authorities of the Ministry of Public Education (MEP in Spanish) have yet to grapple with. In 2021, the Foreign Language Assessment and Training Program (*PELEX* in Spanish) from the School of Modern Languages of the University of Costa Rica administered a language competence test in all public high schools nationwide. The results were not encouraging: 64% of the students were placed on the A1 or A2 band based on the Common European Framework of Reference for Languages. Such results did not yield the B1 minimum that MEP was hoping for, and they certainly do not look promising to reach bilingualism by 2040 (Ruiz, 2022).

The foregoing implies that teaching English could represent a daily challenge in the life of Costa Rican second language (L2) practitioners generally and L2 writing teachers particularly, especially considering that any L2 issue that highschoolers may drag could show at the university level. Hence, this scenario calls for one action that could be useful in furthering knowledge of English teaching in a context with a clear educational need: learner corpus research. This line of inquiry “has primarily relied on collecting and analyzing second language ... learner writings” (Granger, 2008 cited in Alexopoulou et al., 2017, p. 1) with the purpose of, among other things, identifying frequency of use of given L2 structures (Neff et al., 2004) and ascertaining areas of L2 struggle (Arjan et al., 2013). Indeed, the fact that learners' language collections can be computerized has made it possible to have large learner corpora of over 40 million words (e.g., the Cambridge Learner Corpus) as well smaller ones collected with a specific research purpose in mind (Díaz-Negrillo, 2009).

However, there is narrowed down corpus data about the linguistic problems of EFL learners from various first language (L1) Spanish backgrounds: available knowledge emerges

² Test takers consist of people who are interested in taking a proficiency test to measure their English language skills. This means that the sample is self-selected, which brings its representativeness into question.

mainly from EFL learners in Spain—be it from large (Díaz-Negrillo and Valera, 2010) or small learner corpora (Díez-Bedmar, 2005). What is more, to the best of the researcher’s knowledge, no major college wide study has been conducted in the context of this investigation. Hence, in an attempt to assist in the understanding of EFL in Costa Rica generally and from an undergraduate standpoint specifically, the present corpus-aided investigation seeks to identify the L2 error patterns of Costa Rican university writers at all academic levels of an English major. Specifically, the research question that guided this study was the following: What are the grammatical and non-grammatical L2 error patterns of university writers across academic levels of an English major of a public university?

2. Theoretical background

With the advent of Contrastive Analysis (CA) in the late 1950s and Error Analysis (EA) in the 1960s, researchers sought to analyze learners’ L2 errors by looking for differences between learners’ L2 and first language (L1) (i.e., CA) and to classify L2 learners’ errors to explain what caused them (i.e., EA) (for a review, see Bitchener and Ferris, 2012). From these studies (Bhela, 1999), it was possible to determine that learners’ L1 may have an influence on L2 written inaccuracies. Specifically related to Spanish L1, different researchers have shed light on the nature of errors of speakers learning English as a FL. One such example is Alonso (1997), who conducted a study with twenty-eight first year EFL high school students in Spain. According to the author, errors from compositions about the last film the participants had seen were mostly interlingual errors, that is, those “that reflect the learner’s first language structures” (Dulay et al., 1982, p. 23). Similar to Alonso (1997), Calsín (2011, cited in Vargaya, 2019) analyzed the participants’ texts—in this case, 4th and 5th year Linguistics and English students—and found Spanish L1 influence on written errors related to the absence of the –s for the third person conjugation in simple present tense (omission error), the unnecessary addition of –s in adjectives (addition error), and the lack of accuracy in placing the adverbs of frequency in the correct order (lack of sentence order).

Nevertheless, criticism to EA and CA theories because they were too limited in their focus (Bitchener, 2016; Ellis, 1994), on the one hand, and the incorporation of computers in data collection, on the other hand, shifted empirical efforts to a line of inquiry with a methodology that studies language use beyond the causes of L2 errors and L1 comparisons to understand them: that is, corpus linguistics. Lindquist (2009) defines *corpus* as “a collection of texts which

is stored on some kind of digital medium and used by linguists to retrieve linguistic items for research or by lexicographers for dictionary-making” (p. 3). As a result, there are large native corpora that contain all sort of samples of English, which is the most studied language thus far (Granger, 1998a). Some of these are the Brown/Frown Corpus, the London-Lund Corpus of Spoken English (LLC), the Bank of English (BoE), the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the International Corpus of English (ICE). Interestingly, the emergence of native English corpora made it clear that there was also a need for corpora that studied English as used by L2 learners, hence the term *learner corpora* (Díaz-Negrillo, 2009; Nesselhauf, 2004). Among the most prominent learner corpora are the International Corpus of Learner English (ICLE), the Longman Learners’ Corpus (LLC), and the Hong Kong University of Science and Technology (HKUST) learner Corpus (for a comprehensive list, see Lindquist, 2009; Pravec, 2002). Currently, learner corpora “give us access not only to errors but also to learners’ total interlanguage” (Granger, 1998b, p. 6)³. One lack, however, is that much of the understanding of English errors at a university level comes from seminal work on native-speaker corpora (Connors and Lunsford, 1988; Hodges, 1941; Johnson, 1917; Lunsford and Lunsford, 2008; Witty and Green, 1930), and when studies with EFL university learners have been conducted, the context is situated mainly in Europe (e.g., Dagneaux et al., 1998) and Asia to a lesser extent (Narita, 2013).

Thus, few of the investigations on the overall written production of Spanish L1 EFL university writers are Díaz-Negrillo and Valera (2010), Neff et al. (2004), and Díez-Bedmar (2005), out of which just two explore learners’ errors. To illustrate, Neff et al. investigated fourth-year university learners’ lexico-grammatical patterns of writer stance (e.g., it is + (adverb) adjective + that; it is + (adverb) said/thought + that) and compared them with those of professional writers and English L1 university students. The participants were EFL writers whose first languages were Dutch, Belgian-French, Italian, and peninsular Spanish and the language samples were extracted from ICLE. Main results showed an overuse of *it is* + (adverb) adjective + *that* and the agentless passive by the EFL learners, whereas the *it is* + adjective pattern showed no significant differences. Different from Neff et al., Díez-Bedmar (2005) analyzed first-year students’ essays to identify L2 learners’ written errors at a morphological, syntactic, semantic, and pragmatic level. Overall, the findings revealed that some of the most

³ Interlanguage is “a separate linguistic system based on the observable output which results from a learner’s attempted production of a TL [target language] form” (Selinker, 1972, p. 214).

problematic areas were punctuation, spelling conventions, verb tenses, and articles. Then, as an error frequency study, Díaz-Negrillo and Valera (2010) examined a sample of the Non-native Corpus of English (NOCE, Díaz-Negrillo, 2009) and found a complex picture where comma usage, for example, seemed highly problematic along with lexical issues such as wrong word choice.

Clearly, albeit their significant findings, previous investigations are not enough to gain sound insight into Spanish L1 EFL learners' interlanguage and to inform in turn L2 educators and researchers alike. Consequently, the need to further broaden current knowledge of Spanish L1 EFL university writers at different academic levels in Costa Rica inspired this study.

3. Methodology

3.1 Approach

Different researchers agree that the word *corpus* speaks of a methodology being used rather than a topic in linguistics being studied (e.g., Díaz-Negrillo, 2009; Lindquist, 2009; Nesselhauf, 2004). For instance, currently "*corpus* is almost always synonymous of *electronic corpus*, i.e., a collection of texts which is stored on some sort of digital medium and used by linguists to retrieve linguistic items for research or by lexicographers for dictionary-making" (Lindquist, 2009, p. 3). Against this background, the present quantitative study used corpus methods both to create the learner corpus from the participants' written samples of the second semester of 2019 (i.e., IIC2019) and to display the ensuing descriptive findings (see 3.4 for a detailed description). Indeed, in terms of current distinctions in corpus linguistics (i.e., *corpus-based*, *corpus-driven*, and *corpus-aided*), this study is corpus-aided (also known as *corpus-supported*) because corpora are used to find illustrative examples of, in this case, L2 error patterns (see Lindquist, 2009, p. 26 for a description).

3.2 Participants and context

This study took place at the School of Modern Languages from the University of Costa Rica (UCR), a public university located in San José at Rodrigo Facio Branch in IIC2019. Specifically, to create the written learner corpus, only courses with a writing component were visited across all academic levels of the English major: first year (Integrated English I and Integrated English II), second year (English Composition I and English Composition II), third year (English Rhetoric I and English Rhetoric II), and fourth year (English Rhetoric III and

English Rhetoric IV). Hence, the selection criterion was purposive. In its initial stage (see 3.3.2), consent forms from 383 individuals were gathered, but after discarding those whose data were not complete due to absenteeism ($n = 20$) and those whose L1 was not Spanish ($n = 3$), the total number of participants was 360 (male = 61.9%, female = 38.1%) and distributed as follows: first year ($n = 78$), second year ($n = 123$), third year ($n = 95$), and fourth year ($n = 64$). The large majority of the EFL participants ($M_{\text{age}} = 23$, $SD = 5.52$) were Costa Rican ($n = 355$). The rest came from countries such as El Salvador ($n = 1$), Venezuela ($n = 1$), Nicaragua ($n = 2$), and Colombia ($n = 1$). Thus, in all cases, the participants' L1 was Spanish. As for their English proficiency, it differed by academic level: first year (low intermediate; $SD = .807$), second year (low intermediate; $SD = .741$), third year (high intermediate; $SD = .805$), and fourth year (low advanced; $SD = .889$).

3.3 Design

3.3.1 Instruments

3.3.1.1 Learner Profile Sheet

The participants completed a learner profile sheet to provide not only their general personal information but also their specific background information related to their L1 and L2 history (See Appendix A).

3.3.1.2 Placement Test

To ascertain learners' proficiency level, the Oxford's Quick Placement Test (OQPT) was administered (see results in 3.2). The exam could be completed in two versions: online if—based on the course schedule—a language laboratory was available at the time of administering the instrument or print if such availability was not present.

3.3.1.3 Argumentative texts

To create the learner corpus, the participants were provided with a list of six prompts (See Appendix B). Opinion writing (i.e., argumentation) was chosen because it was the only rhetorical pattern that all learners had had some exposure to across all academic levels. Any other rhetorical pattern (e.g., comparison/contrast or cause/effect) would not have given learners equal writing conditions. With this in mind, a specific number of words was also not required. They were, however, encouraged (irrespective of the prompt of their choice) to explain

their reasons clearly and use examples from their own experience to support their ideas. This was done to maximize the chances of a similar text length across levels. All compositions were written on paper since there was no availability of language labs at the time of writing the texts. After conversion of the texts to an editable format (see 3.4), the total number of words in the learner corpus was 57 054 ($M = 158.4$, $SD = 43.6$). As for average length per year, it was as follows: first (Sum = 8871, $M = 113.7$, $SD = 32.8$), second (Sum = 19831, $M = 161.2$, $SD = 38.5$), third (Sum = 17094, $M = 179.9$, $SD = 35.7$), and fourth (Sum = 11258, $M = 175.9$, $SD = 35.7$).

3.3.2 Procedures

Conversations with course instructors preceded the two-week data collection process. Those meetings were necessary to discuss logistics, namely, the chronogram, class time availability, and number of students in the course. Then, Week 1 was spent asking for the participants' consent as well as administering the learner profile sheet and the placement test. On the one hand, the consent form part (i.e., the explanation of the research objective, the summary of both the benefits and implications of participating, and the wait for the signatures in class) took 10 minutes approximately. On the other hand, the allotted time for completing the learner profile sheet and the placement test was 30 minutes.

A week later (Week 2), learners had the chance to choose one writing prompt and develop the answers in the sheets provided. They had 30 minutes to complete the task. Because no language lab was available at the time of writing, all argumentative compositions were pen-and-paper texts. However, if one was available during the schedule of test taking, learners were able to take the online version of the proficiency test.

3.4 Data coding and analysis

After the two week-long data collection period, all handwritten compositions ($N = 360$) were converted into a digital document. To transcribe all texts, the speech recognition software Dragon Naturally Speaking was used. Whenever the software was not able to transcribe an error, it was inserted manually. Then, drawing on Bonilla et al., (2017), each converted text was assigned a code that contained the following information: the setting, the year of data collection, the native language, the target language, the proficiency level, and the participant number (e.g.,

UCR-20-SP-EN-B1-92). The purpose of coding each text was to keep the data coding anonymous.

Specifically, as in previous reports on college writing errors (Lunsford and Lunsford, 2008), all errors present in the text were marked, meaning that error types “emerge[d] out of the data rather than being imposed on them prior to data collection and analysis” (Patton, 1990, p. 306). Thus, after having traced all existing error types and confirmed acceptable interrater and intrarater reliability (see Cronbach’s alpha values in Table 1)⁴, thirty-three error types were identified—all of which belonged to either of the grammatical ($n = 17$) and non-grammatical error categories ($n = 16$). The latter was then further subdivided as (i.e., stylistics) ($n = 10$) and lexical ($n = 6$) for a more fine-grained analysis. All throughout the reference manual was A Comprehensive Grammar of the English Language (Quirk et al., 1985).

Table 1
Reliability (Cronbach’s alpha) for interrater and intrarater consistency per error type

Index	Grammar	Stylistics	Lexis
Interrater	.88	.91	.85
Intrarater	.92	.96	.94

Source: **Elaborated by author (2022)**

4. Results and Discussion

The research question that guided this study sought to identify the grammatical and non-grammatical patterns of university writers at all academic levels of an English major of a public university. Table 2 displays the descriptive statistics of ranked error patterns in first-year university writers. Table 3 presents the descriptive statistics of ranked error patterns in second-year university writers. Table 4 summarizes the descriptive statistics of ranked error patterns in third-year university writers. Table 5 shows the descriptive statistics of ranked error patterns in fourth-year university writers.

⁴ The researcher recoded 10% of the data six months after the first analysis round to establish intrarater reliability. Similarly, an independent experienced rater coded 10% of the data to establish interrater reliability. When discrepancies were found, a number of meetings took place until agreement was reached.

Table 2
Descriptive statistics of frequency of error patterns in first year EFL learners at UCR in IIC2019

Ranking	Error type	Frequency	M	SD
1	lexis.derivation	71	.91	1.153
2	punctuation.comma splice	70	.90	1.401
3	grammar.verb.person.misselection	68	.87	1.155
4	lexis.misselection	67	.86	1.224
5	grammar.sentence fragment	64	.82	.950
6	grammar.article.definiteness	62	.79	1.085
7	punctuation.comma.conjunction.omission	53	.68	.875
8	grammar.ordering	52	.67	.989
9	grammar.article.definiteness.indefinite	49	.63	.968
10	grammar.subject.omission	49	.63	.941
11	grammar.pronoun	48	.62	.929
12	grammar.verb.form.misselection	48	.62	.841
13	spelling.grapheme	46	.59	1.086
14	grammar.parallelism.omission	45	.58	.961
15	punctuation.comma.conjunction.overinclusion	43	.55	.878
16	grammar.sentence structure.multiple error	43	.55	.962
17	lexis.omission	34	.44	.695
18	punctuation.fused sentence	31	.40	.827
19	lexis.overinclusion	31	.40	.779
20	punctuation.comma.introductory phrase.omission	27	.35	.770
21	punctuation.comma.verb.object.overinclusion	26	.33	.474
22	grammar.quantifier.misselection	23	.29	.537
23	grammar.verb.tense.misselection	21	.27	.596
24	spelling.orthographical case	19	.24	.461
25	grammar.adjective.degree.comparative	19	.24	.514
26	lexis.collocation	16	.21	.493
27	lexis.foreign	16	.21	.406
28	grammar.noun.case.genitive	13	.17	.375
29	grammar.auxiliary.modality	12	.15	.363
30	punctuation.comma.non-restrictive elements.omission	10	.13	.336
31	grammar.noun.number	9	.12	.394
32	grammar.adjective.degree.superlative	5	.06	.247
33	punctuation.comma.appositive.omission	0	.00	.000

Source: Elaborated by author (2022)

Table 3
Descriptive statistics of frequency of error patterns in second year EFL learners at UCR in IIC2019

Ranking	Error type	Frequency	M	SD
1	grammar.sentence fragment	67	.54	.781
2	punctuation.comma.conjunction.overinclusion	62	.50	.803
3	grammar.article.definiteness	60	.49	.881
4	lexis.derivation	56	.46	.812
5	punctuation.comma splice	55	.45	.832
6	punctuation.comma.conjunction.omission	55	.45	.760
7	grammar.parallelism.omission	54	.44	.780
8	spelling.grapheme	53	.43	.758
9	lexis.misselection	51	.41	.789
10	grammar.pronoun	50	.41	.745
11	grammar.verb.person.misselection	49	.40	.807
12	lexis.omission	42	.34	.722
13	grammar.ordering	41	.33	.721
14	grammar.subject.omission	37	.30	.572
15	grammar.verb.form.misselection	37	.30	.639
16	grammar.sentence structure.multiple error	33	.27	.628
17	punctuation.comma.introduutory phrase.omission	32	.26	.663
18	grammar.article.definiteness.indefinite	27	.22	.536
19	punctuation.fused sentence	26	.21	.547
20	grammar.verb.tense.misselection	20	.16	.468
21	lexis.overinclusion	20	.16	.468
22	lexis.foreign	20	.16	.371
23	lexis.collocation	19	.15	.406
24	punctuation.comma.appositive.omission	18	.15	.355
25	grammar.quantifier.misselection	14	.11	.367
26	punctuation.comma.verb.object.overinclusion	11	.09	.287
27	punctuation.comma.non-restrictive elements.omission	11	.09	.287
28	grammar.noun.number	11	.09	.287
29	spelling.orthographical case	10	.08	.274
30	grammar.adjective.degree.comparative	8	.07	.279
31	grammar.noun.case.genitive	7	.06	.233
32	grammar.adjective.degree.superlative	1	.01	.090
33	grammar.auxiliary.modality	0	.00	.000

Source: Elaborated by author (2022)

Table 4
Descriptive statistics of frequency of error patterns in third year EFL learners at UCR in IIC2019

Ranking	Error type	Frequency	M	SD
1	punctuation.comma.conjunction.overinclusion	70	.74	1.013
2	punctuation.comma.introductory phrase.omission	50	.53	.932
3	punctuation.comma.conjunction.omission	40	.42	.752
4	lexis.derivation	40	.42	.766
5	punctuation.comma splice	39	.41	.692
6	lexis.misselection	39	.41	.692
7	grammar.parallelism.omission	37	.39	.624
8	grammar.ordering	37	.39	.689
9	grammar.sentence fragment	29	.31	.566
10	lexis.omission	28	.29	.563
11	grammar.article.definiteness	27	.28	.595
12	spelling.grapheme	27	.28	.595
13	grammar.article.definiteness.indefinite	25	.26	.622
14	grammar.subject.omission	22	.23	.555
15	grammar.sentence structure.multiple error	21	.22	.587
16	grammar.verb.person.misselection	21	.22	.549
17	grammar.verb.form.misselection	19	.20	.557
18	lexis.overinclusion	18	.19	.490
19	grammar.pronoun	18	.19	.445
20	punctuation.fused sentence	13	.14	.346
21	punctuation.comma.appositive.omission	10	.11	.309
22	lexis.collocation	10	.11	.341
23	grammar.verb.tense.misselection	10	.11	.309
24	grammar.quantifier.misselection	9	.09	.329
25	punctuation.comma.non-restrictive elements.omission	5	.05	.224
26	lexis.foreign	5	.05	.224
27	spelling.orthographical case	3	.03	.177
28	grammar.noun.number	3	.03	.176
29	grammar.adjective.degree.superlative	2	.02	.144
30	grammar.auxiliary.modality	2	.02	.144
31	grammar.noun.case.genitive	1	.01	.103
32	punctuation.comma.verb.object.overinclusion	0	.00	.000
33	grammar.adjective.degree.comparative	0	.00	.000

Source: Elaborated by author (2022)



Table 5
Descriptive statistics of frequency of error patterns in fourth year EFL learners at UCR in IIC2019

Ranking	Error type	Frequency	M	SD
1	punctuation.comma.conjunction.overinclusion	48	.75	.873
2	lexis.derivation	38	.59	.868
3	punctuation.comma.conjunction.omission	31	.48	.617
4	grammar.parallelism.omission	31	.48	.756
5	punctuation.comma.splice	26	.41	.660
6	punctuation.comma.introduutory phrase.omission	24	.38	.787
7	spelling.grapheme	21	.33	.536
8	grammar.sentence fragment	16	.25	.535
9	lexis.misselection	14	.22	.417
10	lexis.omission	13	.20	.406
11	grammar.ordering	13	.20	.406
12	grammar.verb.person.misselection	13	.20	.443
13	punctuation.fused sentence	10	.16	.366
14	grammar.subject.omission	10	.16	.444
15	lexis.overinclusion	9	.14	.393
16	grammar.article.definiteness	9	.14	.350
17	grammar.verb.form.misselection	8	.13	.333
18	grammar.sentence structure.multiple error	8	.13	.378
19	grammar.pronoun	7	.11	.362
20	lexis.collocation	6	.09	.294
21	punctuation.comma.appositive.omission	3	.05	.213
22	grammar.article.definiteness.indefinite	3	.05	.213
23	lexis.foreign	2	.03	.175
24	grammar.verb.tense.misselection	1	.02	.125
25	grammar.noun.case.genitive	1	.02	.125
26	grammar.adjective.degree.comparative	1	.02	.125
27	punctuation.comma.verb.object.overinclusion	0	.00	.000
28	punctuation.comma.non-restrictive elements.omission	0	.00	.000
29	spelling.orthographical case	0	.00	.000
30	grammar.adjective.degree.superlative	0	.00	.000
31	grammar.auxiliary.modality	0	.00	.000
32	grammar.noun.number	0	.00	.000
33	grammar.quantifier.misselection	0	.00	.000

Source: Elaborated by author (2022)



As can be seen, the areas of linguistic issues differed across academic levels. Even though first-year learners' number one error category was lexical derivation ($n = 71$), overall, the error categories with higher occurrences were grammar oriented, ranging from subject-verb agreement issues ($n = 68$) and fragment ($n = 64$) to word order ($n = 52$) and (in)definite article confusion along with subject deletion ($n = 49$). A few non-grammatical issues also appeared at the top, namely comma splice ($n = 70$) followed by comma omission before coordinating/correlative conjunction joining clauses ($n = 53$).

It can also be observed that when compared to their first-year counterparts (cf. Table 2), some grammatical error categories remained in the top 10 of second-year writers. Table 3 reveals that such is the case of sentence fragment ($n = 67$) and missing or overinclusion of definite article ($n = 60$) errors. A similar situation occurred with a non-grammatical error type such as spelling, which ranked high both in first ($n = 46$) and second year ($n = 54$). In addition, punctuation-related issues that ranked lower in first year (e.g., unnecessary comma before coordinating/correlative conjunction joining words or phrases) had a higher ranking in second year ($n = 62$). Others remained equally problematic, for example, comma splices ($n = 62$) and comma omission before a coordinating/correlative conjunction joining clauses ($n = 55$). As far as lexical errors are concerned, word formation ($n = 56$) and word choice ($n = 51$) issues had lower counts unlike missing lexical items, which increased ($n = 42$).

From Table 4 it can be seen that third-year writers' predominant error categories consist of non-grammatical issues, being comma-related errors in the top three. It is also evident that out of the six types of lexical errors, *lexis.derivation* ($n = 40$) *lexis.misselection* ($n = 39$)—albeit the lower frequency of occurrence when compared with those of first- and second-year learners—were still troublesome. Similarly, despite the lower sum, grammatical categories such as parallelism ($n = 37$) word order ($n = 37$), sentence fragment ($n = 29$), and article-related issues were ranked high. It is also worth highlighting that error categories involving an unnecessary comma between verb and object as well as comparative adjective issues had no error counts, which was not the case in first (cf. Table 2) and second (cf. Table 3) year.

Table 5 shows a similar pattern to Table 4: (a) error types with a higher frequency of occurrence were non-grammatical rather than grammatical, (b) *lexis.derivation* remained in the top five, and (c) the grammatical issues in both academic levels (i.e., 3rd and 4th year) were the same except that they had lower counts (i.e., parallelism, sentence fragment, and word order issues). One difference, however, is that there was no error trace in seven error categories, out of which three were non-grammatical and three were grammatical.

Thus far, Table 1 to 5 clearly render an intricate linguistic scenario. That is why the theoretical and practical implications emerging from the results will be explained in light of key methodological variables from previous research (4.1) as well as relevant factors in the EFL classroom (4.2).

4.1 Past empirical efforts

Just as previous corpus-oriented work on college writing errors (Connors and Lunsford, 1988; Lunsford and Lunsford, 2008), this study sheds more light on error patterns of university writers. Nonetheless, by (1) including learners' academic year as a variable, (2) having a sample that consists of Spanish L1 English (Teaching) majors only, and (3) employing a computer-tagging system, the present exploratory study renders a fine-grained analysis not available thus far. More specifically, if the L2 error patterns of this study were to be displayed as a whole, the ranked categories—as previously shown from Table 2 to Table 5—would paint a completely different picture. While not exhaustive, Table 6 summarizes a historical top ten error list. This list seeks to compare English errors as found in native corpora (Connors and Lunsford, 1988; Hodges, 1941; Johnson, 1917; Lunsford and Lunsford, 2008; Witty and Green, 1930) and in the present study. As can be observed, participants across studies share similar problem areas. To illustrate, two non-grammatical error types that are recurrent in Table 6 are related to spelling and the use of comma—all present in 4 out of 5 lists. However, differences in findings could be explained by taking a close look at key methodological variables. The list below briefly describes each of them.

4.1.1 Analysis across levels

Notwithstanding their significant contribution, a bird's eye view of university writers' L2 error patterns whether from a large learner corpus (Connors and Lunsford, 1988) or a few samples (Sajid, 2016) may not be accurate enough if it does not provide a nuanced outlook of the specific linguistic problem areas per academic level. For instance, from available literature (Ali Al-Khairi, 2013; Al-Jamal, 2017; Connors and Lunsford, 1988; Lunsford and Lunsford, 2008), errors related to verbs, articles, pronouns, punctuation, word choice, spelling, agreement, and singular/plural noun endings seem to be the most common irrespective of differences in L1 backgrounds. Interestingly, a more fine-grained analysis suggests that error frequencies may as well vary per level. Table 6 illustrates this point. For instance, while the global ranking of this study does not include pronoun errors in the top ten, it was indeed an important language issue but mainly in first- and second-year learners and not so much on more advanced learners such as their third- and especially fourth-year counterparts.

Table 6
Historical top ten error lists

Johnson (1917) 198 papers	Witty and Green (1930) 170 timed papers	Hodges (1941) 16 000 papers	Lunsford and Lunsford (2008) 877 papers	The present study (2022) 360 papers
Spelling	Faulty connectives	Comma	Wrong word	punctuation.comma. conjunction. overinclusion lexis.derivation
Capitalization	Vague pronoun reference	Spelling	Missing comma after an introductory element	punctuation.comma splice
Punctuation (mostly comma errors) Careless omission or repetition	Use of "would" for simple past tense forms Confusion of forms from similarity of sound or meaning	Exactness Agreement	Incomplete or missing documentation Vague pronoun reference	punctuation.comma. conjunction. omission grammar.sentence fragment lexis.misselection
Apostrophe errors Pronoun agreement	Misplaced modifiers Pronoun agreement	Superfluous commas Reference of pronouns	Spelling error Mechanical error with a quotation Unnecessary comma	grammar.parallelism. omission grammar.article. definiteness
Verb tense errors and agreement Ungrammatical sentence structure (fragments and run- ons)	Fragments Unclassified errors	Apostrophe Omission of words	Unnecessary or missing capitalization	grammar.verb.person. misselection
Mistakes in the use of adjectives and adverbs Mistakes in the use of prepositions and conjunctions	Dangling modifiers Wrong tense	Wordiness Good use	Missing word Faulty sentence structure	spelling.grapheme

Source: Adapted by author (2022) with information from Lunsford and Lunsford (2008)

Similar patterns of change depending on the academic level can be observed in other categories of grammatical, non-grammatical, and lexical errors. Four more examples (among others) exemplify the aforementioned: (a) when compared with first-year writers, fragment errors in fourth year were rare; (b) most punctuation issues that involved coordinating and correlative conjunctions took place in advanced levels (cf. Table 5 and Table 6); (c) spelling errors were mostly problematic in the first two years of the major (cf. Table 1 and Table 2); and (d) lexical problems due to L1 interference were more frequent in first year. Attributing factors to these results could be the more advanced language proficiency of higher academic levels, which comes from more years of syntactical and lexical input.

4.1.2 *Handwritten texts*

When participants are allowed to use basic word processing (e.g., Lunsford & Lunsford, 2008), the spell check tool will aid learners—unless they are deactivated. Indeed, Lunsford and Lunsford (2008) hypothesized that the spell check function may explain why their sample had lower frequencies of spelling errors (when compared with Connors and Lunsford, 1988) and a large number of wrong word errors. However, such an explanation does not apply to the present study because all texts were handwritten, and no dictionaries were used. This implies that all participants were indeed writing to the best of their ability, meaning in turn that their output may have been a truer reflection of their interlanguage. Such a possibility has noteworthy practical implications, especially when considering that previous work comparing the effects of word processor on the quality of essays written by EFL students has—not surprisingly—found an advantage of word-processed texts vis-à-vis handwritten ones (Darus et al., 2008).

4.1.3 *Task type*

Previous research attempts on L2 error identification have analyzed all sorts of text types ranging from term papers (Amiri and Puteh, 2017) and letter writing (Ali Al-Khairi, 2013) to essay writing (Al-Jamal, 2017) and cover letters (Lunsford and Lunsford, 2008). The relevance of this methodological difference lies in the ensuing practical implications. For example, based on the results obtained in this study, an overgeneralization would be to conclude that EFL university writers across levels seem not to struggle with mechanical errors in a quotation—at least not in a way that other learner types would (Lunsford and Lunsford, 2008 in Table 6). Nevertheless, the reality is that participants in the present study showed no

problems with sources and attributions because no sources were required after all. What is more, all texts were opinion compositions with instructions that explicitly stated that learners needed to use examples from their own experience, not sources (Appendix B). Arguably, had other task types been included in the analyses (e.g., a research report), other error types may have emerged. One such example could be punctuation errors in bibliographical entries when attempting to use a referencing style (Amiri and Puteh, 2017).

4.2 L2 errors in EFL writing

Second language acquisition (SLA) is not linear; L2 learners may seem to master a given structure only to regress in time as they may still be in the path towards L2 development (Bitchener, 2016). On the other hand, it cannot be denied that a number of pedagogically oriented questions may be gleaned from the findings of this study. To illustrate, results that reveal recurrent L2 errors across academic levels could understandably prompt L2 (writing) instructors to ask themselves why that is. For instance, although some error frequencies lowered to the point of having none as learners advanced in the major (e.g., capitalization, superlatives, modals, and quantifiers), other error frequencies suggest that a given problem area persisted irrespective of the academic level (e.g., word form errors, fragments, comma splices, run-ons, and word order).

To address a potentially attributing factor for this scenario, defining L2 input and how it is processed is called for. Input is defined as “language that is available to the learner through any medium” (Gass and Mackey, 2006, p. 5). This means that L2 learners are exposed to all sorts of input types—be it authentic or modified: songs, newspaper articles, billboards, video games, documentaries, books, chats, posters, movies, peer talk, teacher talk, peer feedback, teacher feedback, etc. However, as explained in Leow (2015), not all the input that learners are exposed to is taken in. That is, due to attentional and cognitive constraints, some of the input may be lost and not further processed into the internal system. The input that is indeed taken in makes it to learners’ L2 internal grammar, which will reflect learners’ interlanguage. Such L2 knowledge will be seen in learners’ output (oral or written), which will evince in turn to what extent L2 knowledge of a given TL structure needs more opportunities for consolidation. On the other hand, if the input is not taken in, no L2 development can even commence and more input will be necessary (for the fine-grained description of theoretical framework of L2 learning process in SLA, see Leow, 2015, p. 17).

Consequently, the aforementioned description raises a couple of questions. What should EFL university learners be capable of writing if their background TL history is reduced? What would then be a reasonable expectation for EFL university writers if their TL entry profile was already questionable to begin with? Conclusive answers to these questions cannot be provided in the absence of more learner corpora in the context of this investigation—hence, the relevance of this springboard study. However, two facts can be irrefutably stated: (1) due to budget constraints and lack of qualified L2 teachers, the English coverage in public kindergartens reaches only 17.7%—far less than the 100% coverage that MEP authorities wanted by 2022 (Cerdas, 2022), and (2) Costa Rican youngsters do not meet the English exit profile when they finish high school—a fact that has repeatedly made the news over the years (Cascante, 2013; Cordero, 2019; Garza, 2015, 2020; González, 2021; Ruiz, 2022).

Understandably, against this background, at a university level more L2 knowledge gaps will need to be filled, more L2 problems will be dragged to higher academic years, and a bigger effort on the part of L2 instructors and learners will need to be made. After all, from a linguistic perspective, grammatical, non-grammatical, and lexical errors require understanding of different domains of knowledge (Truscott, 2001) and treating lingering L2 issues will imply dealing with the different degrees of complexity of those domains. As a matter of fact, there is growing evidence from written corrective feedback (CF) research that error complexity plays a major role in the extent to which diverse error categories are responsive to correction (Bonilla-López et al., 2021; Diab, 2015; Ferris and Roberts, 2001; Shintani and Ellis, 2013). To illustrate, main findings in Bonilla-López et al. (2021) showed that even after feedback plus revision, EFL learners were not able to show short-term gains in error categories related to pronouns, subject deletion, subject-verb agreement, and spelling. More interesting results yielded evidence of errors such as fragments, subject repetition, and verbs having no response at all to feedback provided under certain conditions (See Table 7, p. 61 for the authors' description of potential sources of error complexity in Spanish L1 EFL learners).

Thus, the results in this study, in which there is a seemingly recurrent nature of some errors despite learners' advancement in the major (e.g., word form errors, fragments, comma splices, run-ons, and word order), may bring into question not only learners' exposure to a pivotal input type (i.e., written CF) but also their instructors' classroom actions to have them notice that input. Simply put, if over the years the experience that these participants have had with written CF has been deficient, the findings in this study are not surprising. To put the word

'deficient' into perspective, the stages of cognitive processing of input in Gass (1997) could help: besides being given written CF, L2 learners must have opportunities to (1) attend to it. Noticing this input (2) enables a cognitive comparison that will allow learners to (3) match that input with existing stored knowledge. They will then (4) process the information and (5) modify their output, which will reflect whether there is repair or not. If there is repair (i.e., successful error correction), there is evidence that learners are in the process of L2 development (see also Bitchener and Ferris, 2012). If there is no evidence of repair (e.g., repetition of error), learners will need more input and opportunities to consolidate their L2 knowledge. Consequently, considering that L2 learning cannot take place if there is no attention (i.e., noticing) (Leow, 2015; Schmidt, 1990), a deficient feedback practice is one that provides no feedback or that provides feedback but does not ask learners to do something with it. Therefore, even though the present study did not elicit data that could elucidate potential sources of error frequency, the jury is out when it comes to the quality and quantity of input (in the form of written CF) that these participants may have received over the years.

Furthermore, the participants' potential exposure to detrimental feedback practices at some point of their TL acquisition history plus the fact that the complexity of errors makes some more amenable to correction than others (Bonilla et al., 2021; Diab, 2015) might have confounded with a key contextual variable in this investigation, which involves the learning-to-write and the writing-to-learn-language dimensions (Manchón, 2011). This means that the EFL university writers in this study were learning how to write texts and at the same time using writing as a vehicle to learn the TL, making them in dire need of L2 input and posing in turn a stark difference between the participants in the present study and those of native corpora. Such a need for vast input gains even more importance by taking a closer look at the participants reported history of TL exposure. For example, the metadata revealed that the language spoken at home as they grew up was Spanish (94%), that Spanish was the medium of instruction in primary (91.8%) and high school (90.4%), and that majority had never been in an English-speaking country (73%). Against the aforementioned, it would seem reasonable to speculate that had learners been exposed to English and efficient feedback practices from the start of their academic years, not only could education authorities be closer to reach the L2 learning goals they set for the country, but also learners' areas of grammatical and non-grammatical struggle before entering the university and across levels may differ. However, due to the novelty of a

corpus-aided study in the context of this investigation, more studies (with both quantitative and qualitative data) are in order to substantiate the interpretation of the findings.

5. Conclusion

The present corpus-aided study widens current knowledge of the error patterns of L2 university writers generally and Costa Rican EFL learners at UCR in IIC2019 particularly. In a nutshell, main findings rendered a complex linguistic scenario worth highlighting: (1) even if first-year learners' highest error frequency was lexically related, the predominant L2 issues were more grammar oriented, (2) second- and first-year learners had similar grammatical issues on top, yet punctuation issues in comma usage started to rank higher as learners progressed in the major, and (3) while some error frequencies decreased in fourth year to the point of not appearing at all, some lexical and syntactic matters—at a phrase and clause level—remained problematic across academic years. These findings lend support to the belief that if there is something that “corpus research has helped clarify is error” (Wilder and Yagelski, 2018, p. 384).

In fact, even though the present results emerge from a particular L2 learning environment, this study could still be illuminating for stakeholders in similar circumstances. First, it might be common belief that once L2 learners pass a course and advance in their study plan, they should show L2 improvement (even mastery) of the L2 linguistic content they were exposed to. Nevertheless, as a contribution to L2 education, the results refute this common misconception and show that this may not always be the case and that, as far as writing is concerned, L2 error frequencies could vary across and within academic levels. Such findings have relevant pedagogical and theoretical implications because they add support to SLA research, which has stated that L2 acquisition is complex, dynamic, and non-linear (Larsen-Freeman, 1997, 2003). Therefore, if L2 exposure does not immediately equate with L2 development—let alone L2 acquisition (for a distinction, see Bitchener and Storch, 2016, p. 2), L2 practitioners may want to reflect on their teaching practices. For instance, keeping in mind the Theoretical Framework for the L2 Learning Process in SLA (see Leow, 2015, p. 15) and the Stages of Cognitive Processing of Input (see Bitchener and Storch, 2016, p. 18), a good starting point would be asking oneself: How many classroom activities am I implementing to maximize learners' chances to consolidate L2 knowledge? Am I providing written CF? Am I making sure learners process such feedback? Am I exposing learners to sufficient TL input?

Interestingly, in the context of this investigation, the latter may also be worth asking to interested parties at the highest levels of government (e.g., MEP authorities) since at primary and secondary levels, teaching English in Spanish has been customary (Ugarte, 2015), which clearly is a detrimental practice for learners' L2 development and the country's goal to reach bilingualism by 2040. Hence, in non-predominant English countries, supervision and training from decision-making authorities are also needed for any real L2 change to be seen nationwide. Clearly, seeking to raise awareness and understanding of how a L2 is learned is necessary not only for those at the front line of the L2 classrooms but also for those on top at the ministry level. In this respect, the present study offers a springboard for such a discussion, contributing in turn to Costa Rican L2 education.

Second, the results of this study could also be of use for corpus linguistics researchers to inform their research design and account for all variables. To list one example, the present results differed from those of other studies on error identification (Amiri and Puteh, 2017) because of differences in key variables (e.g., task type). This was somewhat expected because as Granger (1998a) states, "learner output has been shown to vary according to the task type" (p. 8). As a matter of fact, the author further adds that "the topic is also a relevant factor because it affects lexical choice, while the degree of technicality affects both the lexis and the grammar" (p. 8). Nevertheless, while not surprising, differences in findings do bring to the fore the need to rigorously report all variables to determine to what extent research findings may (or not) be applicable to other learning environments. For this same reason, caution should be exercised in the interpretation of the results in this study.

Third, it is hoped that L2 practitioners and L2 learners alike can benefit from the bird's eye view of the L2 error patterns of the EFL learners in this investigation. That is, the fact that the error frequencies of some L2 error categories still ranked high over time seems to suggest that learners' L2 knowledge of lexical, syntactic, morphological, and stylistic domains could need more expert input (in the form of explicit instruction and/or feedback) depending on the complexity of the target structure—an aspect already touched upon in the bulk of written CF studies (Bonilla-López et al., 2021; Diab, 2015). Taking this into consideration, the present findings may be useful to increase L2 practitioners' awareness of potential areas of struggle of FL college learners and to inform, as a result, the creation of classroom materials that will cater to their students' linguistic needs.

Finally, future studies might want to consider the following caveats. It is hard to characterize an entire learner type on the basis of a small learner corpus based on one text type and collected at a given point in time (i.e., synchronic). Therefore, studies that aim for a larger sample and that create a learner corpus consisting of varied rhetorical patterns, emerging from one prompt only (per pattern), and having similar text length remain in order. Doing so would improve control of variables such as task, topic, and text length and allow a fairer comparison among learners (see Caines and Buttery, 2018 for an explanation of *opportunity of use*). As a matter of fact, in the context of this investigation, there is a clear need to conduct a nationwide corpus investigation that gathers a variety of texts both at a high school and university level and for larger stretches of time (i.e., diachronic). In other words, besides administering much-needed L2 competence tests (e.g., PELEX efforts), analyzing learners' actual L2 output through corpus data may be the only way of painting a complete picture of the country's L2 English situation as far as proficiency is concerned. In this respect, while targeted at college level, the present corpus-aided investigation constitutes a start in that direction.

6. References

- Alexopoulou, Theodora., Michel, Marije., Murakami, Akira., and Meurers, Detmar. (2017). Task Effects on Linguistic Complexity and Accuracy: A Large-Scale Learner Corpus Analysis Employing Natural Language Processing Techniques: Task Effects in a Large-Scale Learner Corpus. *Language Learning*, 67(S1), 180–208. <https://doi.org/10.1111/lang.12232>
- Ali Al-Khairy, Mohamed. (2013). Saudi English-Major Undergraduates' Academic Writing Problems: A Taif University Perspective. *English Language Teaching*, 6(6), p1. <https://doi.org/10.5539/elt.v6n6p1>
- Al-Jamal, Dina AbdulHameed. (2017). Students' Fossilized Writing Errors: EFL Postgraduates at Jordanian Universities as a Model. *Journal of Al-Quds Open University for Educational & Psychological Research & Studies*, 6(19). https://digitalcommons.aaru.edu.jo/cgi/viewcontent.cgi?article=1228&context=jaqou_edpsych
- 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. <https://doi.org/10.14198/raei.1997.10.01>
- Amiri, Fatemeh., and Puteh, Marlia. (2017). Error Analysis in Academic Writing: A Case of International Postgraduate Students in Malaysia. *Advances in Language and Literary Studies*, 8(4), 141. <https://doi.org/10.7575/aiac.all.v.8n.4p.141>

- Arjan, Asmeza., Hayati Abdullah, Noor., and Roslim, Norwati. (2013). A Corpus-Based Study on English Prepositions of Place, in and on. *English Language Teaching*, 6(12), p167. <https://doi.org/10.5539/elt.v6n12p167>
- Bhela, Baljit. (1999). Native language interference in learning a second language: Exploratory case studies of native language interference with target language usage. *International Education Journal*, 1(1), 22–31.
- Bitchener, John., and Ferris, Dana. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Bitchener, John. (2016). To what extent has the published written CF research aided our understanding of its potential for L2 development? *ITL - International Journal of Applied Linguistics*, 167(2), 111-131. <https://doi.org/10.1075/itl.167.2.01bit>
- Bitchener, John., and Storch, Naomy. (2016). *Written Corrective Feedback for L2 Development* (Vol. 96). Multilingual Matters.
- Bonilla-López, Marisela., Van Steendam, Elke., and Buyse, Kris. (2017). Comprehensive corrective feedback on low and high proficiency writers: Examining attitudes and preferences. *ITL - International Journal of Applied Linguistics*, 168(1), 91-128. <https://doi.org/10.1075/itl.168.1.04bon>
- Bonilla-López, Marisela., Van Steendam, Elke., Speelman, Dirk., and Buyse, Kris. (2021). Comprehensive corrective feedback in second language writing: The response of individual error categories. *Journal of Writing Research*, 13(1), 31–70. <https://doi.org/10.17239/jowr-2021.13.01.02>
- Caines, Andrew., and Buttery, Paula. (2018). The effect of task and topic on opportunity of use in learner corpora. In Vaclav Brezina and Lynne Flowerdew (Eds.), *Learner Corpus Research: New Perspectives and Applications* (pp. 6-27). Bloomsbury.
- Cascante, Luis Fernando. (2013, November 11). Costa Rica: Reprobada en inglés. *La República*, 4.
- Cerdas, Daniela. (2022, March 10). Inglés tardó 24 años en alcanzar a 18% de niños de preescolar. *La Nación*, 4.
- Connors, Robert., and Lunsford, Andrea. (1988). Frequency of Formal Errors in Current College Writing, or Ma and Pa Kettle Do Research. *College Composition and Communication*, 39(4), 395. <https://doi.org/10.2307/357695>
- Cordero, Monserrat. (2019). *MEP: Colegios Públicos tienen nivel básico en dominio del inglés*. CONICIT. http://www.conicit.go.cr/prensa/historico/historico_noticias/MEP_ingles_basico.aspx
- Dagneaux, Estelle., Denness, Sharon., and Granger, Sylviane. (1998). Computer-aided error analysis. *System*, 26(2), 163–174. [https://doi.org/10.1016/S0346-251X\(98\)00001-3](https://doi.org/10.1016/S0346-251X(98)00001-3)

- Darus, Saadiyah., Ismail, Kemboja., and Ismail, Mohamed Bashir. (2008). *Effects of Word Processing on Arab Postgraduate Students' Essays in EFL*. 7(2), 16.
- Diab, Nuwar. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing*, 24, 16–34. <https://doi.org/10.1016/j.asw.2015.02.001>
- Díaz-Negrillo, Ana. (2009). *EARS: A user's manual*. LINCOM EUROPA.
- Díaz-Negrillo, Ana., and Valera, Salvador. (2010). A learner corpus-based study on error associations1. *Procedia - Social and Behavioral Sciences*, 3, 72–82. <https://doi.org/10.1016/j.sbspro.2010.07.014>
- Díez-Bedmar, María Belén. (2005). Struggling with English at university level: Error patterns and problematic areas of first-year students' interlanguage. In P. Danielsson and M. Wagenmakers (Eds.), *The Corpus Linguistics Conference Series*. <https://ininet.org/struggling-with-english-at-university-level-error-patterns-and.html>
- Dulay, Heidi., Burt, Marina., and Krashen, Stephen. (1982). *Language two*. Oxford University Press.
- Education First. (2021). *EF EPI EF English Proficiency Index*. (2021). www.ef.com/eipi
- Ellis, Rod. (1994). *The Study of Second Language Acquisition*. Oxford University Press.
- Ferris, Dana., and Roberts, Barrie. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10(3), 161–184. Doi [https://doi.org/10.1016/S1060-3743\(01\)00039-X](https://doi.org/10.1016/S1060-3743(01)00039-X)
- Garza, Jeffrey. (2015, November 9). Costa Rica se estanca en nivel de inglés. *La República*. https://www.larepublica.net/noticia/costa_rica_se_estanca_en_nivel_de_ingles
- Garza, Jeffrey. (2020, November 23). Costa Rica retrocede seis lugares en dominio del inglés. *La República.Net*. <https://www.larepublica.net/noticia/costa-rica-retorcede-seis-lugares-en-dominio-del-ingles>
- Gass, Susan. (1997). *Input, interaction, and the second language learner*. Erlbaum.
- Gass, Susan., and Mackey, Alison. (2006). Input, interaction and output: An overview. *AILA Review*, 19(1), 3–17.
- González, Andrea. (2021, February 15). 70% de los egresados de educación pública tiene un inglés muy básico. *SINART Costa Rica Medios*. <https://costoricamedios.cr/2021/02/15/70-de-los-egresados-de-educacion-publica-tiene-un-ingles-muy-basico/>
- Granger, Sylviane. (1998a). *Learner English on Computer*. Longman.

- Granger, Sylviane. (1998b). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger, *Learner English on Computer* (1st ed., pp. 3–18). Routledge. <https://doi.org/10.4324/9781315841342-1>
- Hodges, John. (1941). *Harbrace handbook of English*. Harcourt.
- Johnson, Roy Ivan. (1917). The Persistency of Error in English Composition. *School Review* 25(8), 555–580.
- Larsen-Freeman, Diane. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141–165. <https://doi.org/10.1093/applin/18.2.141>
- Larsen-Freeman, Diane. (2003). *Teaching language from grammar to grammaring*. Heinle/Thomson.
- Leow, Ronald. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.
- Lindquist, Hans. (2009). *Corpus linguistics and the description of English*. Edinburg University Press.
- Lunsford, Andrea., and Lunsford, Karen. (2008). “Mistakes are a fact of life”: A national comparative study. *College Composition and Communication*, 59(4), 781–806.
- Manchón, Rosa. (2011). Writing to learn language: Issues in theory and Research. In *Learning-to-write and writing-to-learn in an additional language* (Vol. 31; pp. 61–82). John Benjamins Publishing Company.
- Narita, Masumi. (2013). The use of articles in Japanese EFL learners’ essays. In Sylviane Granger, Gaëtanelle Gilquin, and Fanny Meunier (Eds.), *Twenty Years of Learner Corpus Research: Looking back, moving ahead* (pp. 357–366). Presses Universitaires de Louvain.
- Neff, JoAnne., Ballesteros, Francisco., Dafouz, Emma., Martínez, Francisco., Rica, Juan-Pedro., Díez, Mercedes., and Prieto, Rosa. (2004). Formulating writer stance: A contrastive study of EFL learner corpora. *Applied Corpus Linguistics*, 77–89. https://doi.org/10.1163/9789004333772_006
- Nesselhauf, Nadja. (2004). Learner corpora and their potential for language teaching. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 125-152). John Benjamins.
- Patton, Michael Quinn. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage.
- Pravec, Norma. (2002). Survey of learner corpora. *ICAME Journal*, 26(1), 8-14.
- Quirk, Randolph., Greebaum, Sidney., Leech, Geoffrey., and Svartvik, Jan. (1985). *A comprehensive grammar of the English Language*. Longman.

- Ruiz, Francisco. (2022, January 27). Costa Rica reprueba en inglés: La mayoría de colegiales no supera nivel mínimo definido por el MEP. *El Financiero*. <https://bit.ly/3A0wYYS>
- Sajid, Muhammad. (2016). Diction and Expression in Error Analysis Can Enhance Academic Writing of L2 University Students. *Advances in Language and Literary Studies*, 7(3). <http://dx.doi.org/10.7575/aiac.all.v.7n.3p.71>
- Schmidt, Richard. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://nflrc.hawaii.edu/PDFs/SCHMIDT%20The%20role%20of%20consciousness%20in%20second%20language%20learning.pdf>
- Shintani, Natsuko., and Ellis, Rod. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286–306. <https://doi.org/10.1016/j.jslw.2013.03.011>
- Truscott, John. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics*, 27(2), 93-108.
- Ugarte, Joselyne. (2015, November 15). Las razones por las que la enseñanza del inglés es mala en el país. *Crhoy.Com*. <https://archivo.crhoy.com/las-razones-por-las-que-la-ensenanza-del-ingles-es-mala-en-el-pais/nacionales/>
- Vargaya, Abimael. (2019). *Interferencia morfosintáctica del español en la producción de textos descriptivos en inglés que presentan los estudiantes del Instituto de Idionas "Concord" - Juliaca 2018* (Tesis para obtener el título de Licenciado en Educación, Especialidad Lingüística e Inglés). Universidad Peruana Unión, Juliaca, Perú. http://200.121.226.32:8080/bitstream/handle/20.500.12840/3437/Abimael_Tesis_Licenciatura_2019.pdf?sequence=4&isAllowed=y
- Wilder, Laura., and Yagelski, Robert. (2018). Describing cross-disciplinary analytic moves in first-year student writing. *Research in the Teaching of English*, 52(4), 382-403. <https://www.jstor.org/stable/26802704>
- Witty, Paul., and Green, Roberta La Brant. (1930). Composition Errors of College Students. *The English Journal*, 19(5), 388-393. <https://doi.org/10.2307/803405>

Appendix A. Learner profile sheet

Name:	Last names:
Email Address:	
Gender: Male <input type="checkbox"/> Female <input type="checkbox"/>	Date of birth:
Place of birth:	Current place of residence:
Major currently enrolled in: English <input type="checkbox"/> English Teaching <input type="checkbox"/>	
Current year of study: First <input type="checkbox"/> Second <input type="checkbox"/> Third <input type="checkbox"/> Fourth <input type="checkbox"/>	

Your native language: Spanish Other(s): _____

Your father's native language: Spanish Other(s): _____

Your mother's native language: Spanish Other(s): _____

Language(s) spoken at home: Spanish Other(s): _____

Medium of instruction in primary school: Spanish Other(s): _____

Medium of instruction in high school: Spanish Other(s): _____

Age at which you started to formally study English: _____

Have you stayed in an English-speaking country before? Yes No

Where? _____

How long? _____

Please estimate your **English** proficiency by putting a check (✓) in the boxes below. Consider the following numbers to answer:

1: low beginner 2: high beginner 3: low intermediate 4: high intermediate
 5: low advanced 6: high advanced

Skill	1	2	3	4	5	6
Listening						
Speaking						
Reading						
Writing						

If you speak other languages, please indicate it and circle your overall language proficiency.

Language: _____ Proficiency: 1 2 3 4 5 6

Language: _____ Proficiency: 1 2 3 4 5 6

Appendix B. Writing prompts

Dear student: From the options below, please circle the writing prompt you wish to write about. After that, use the remaining space in this handout and the extra sheet attached to develop your answer. In doing so, please, explain your reasons clearly. You may also use examples from your own experience to support your ideas. Thank you for your cooperation!

PROMPT A. Do you think people should diet more and eat less to live a healthy life? Why or why not? Explain.

PROMPT B. Do you agree that success equals big money? Why or why not? Explain.

PROMPT C. Do you agree that people should exaggerate the truth or outright lie in their resume if that will help them to get a job? Why or why not? Explain.

PROMPT D. Do you agree that *Ampliación* exams should be eliminated? Explain.

PROMPT E. Do you agree that lack of participation in presidential elections (i.e., abstentionism) should be penalized with a fine? Why or why not? Explain.

PROMPT F. Do you agree that technology is limiting creativity? Why or why not? Explain.

PROMPT G. Do you agree that spanking constitutes child abuse? Why or why not?

PROMPT H. Do you agree that criticism against millennials is justifiable? Why or why not?

Appendix C. Tagging system (adapted from Díaz-Negrillo, 2009)

Error tag (meaning)	Description (example)
GM.PO (grammar.pronoun)	faulty pronoun agreement/misselection (economic independence has her pros and cons)
GM.QF.MS (grammar.quantifier.misselection)	quantifier misselection (I haven't slept many time)
GM.NN.NB (grammar.noun.number)	noun shift or error in number (They were little childs)
GM.SF (grammar.sentence fragment)	sentence fragment (Although when we were children, we used to watch cartoons. / For example, when you have to work the whole day.)
GM.SF (grammar.sentence structure.multiple error)	faulty sentence structure (because you administer your money how do you want / why we want to lie in a resume? / They are persons capable to do so much)
GM.VR.TN.MS (grammar.verb.tense.misselection)	verb tense (In my first day I look at everyone / a few friends who are here since last week / you can enjoy more if you share with friends who had been with you all your life)
GM.VR.FR.MS (grammar.verb.form.misselection)	verb form (when you lie, you must think that you getting a job unfairly / but that does not always happens / if you didn't wanted to lie, you would have thought about it / selling them have became)
GM.PA.OM (grammar.parallelism.omission)	lack of parallel structures when using coordinating/correlative conjunctions (A person with money can not only buy anything but also he can travel anywhere / A healthy lifestyle can be eating lots of fruits and go to the gym)
GM.AX.MD (grammar.auxiliary.modality)	omission or misselection of modal auxiliaries (It's very possible that you get a job / I can't found)
GM.AR.DF (grammar.article.definitness)	missing or overinclusion of definite article (when I go out the friends / when I check on Internet / the creativity is important to express our ideas.)
GM.AR.DF.IN (grammar.article.definitness.indefinite)	'a/an' and/or 'the' confusion (an university student may sometimes not have a healthy diet because he is busy / this person can go to work by a car)
GM.VR.RS.MS (grammar.verb.person.misselection)	subject-verb agreement (people makes use of technology in a terrible way.)
GM.AD.DG.CV (grammar.adjective.degree.comparative)	comparative adjective error (this makes our life more easier / it has made him a little bit good)
GM.AD.DG.CV (grammar.adjective.degree.superlative)	superlative adjective error (I think the most young people / here the problem is worst)
GM. OR (grammar.ordering)	word order (People while on diet may experience weight loss.)
GM.SB.OM (grammar.subject.omission)	subject deletion (Creativity is useful to find new ways of dealing with problems but is not all a person needs / First, is an opportunity to climb the ladder is we do a good job.)
GM.NN.CA.GT (grammar.noun.case.genitive)	overinclusion or missing genitive case (the literature's class / American's traditions)
LX.MS (lexis.misselection)	wrong word choice (exaggerating or adding information that is not true might be a wonderful idea <i>in</i> the moment / she went

	to an <i>interchange</i> program / to watch TV among 4-5 hours per day you have to <i>dominate</i> 2 or 3 languages)
LX.FR (lexis.foreign)	literal translation from L1 Spanish (this happens because the different nutrients are related directly with the physiological changes in the human body / fortunately, this does not always happen)
LX.OM (lexis.omission)	missing lexical item at a phrase level (they are states that hold an army because reasons of defense / on one hand, I know that it will have consequences)
LX.OV (lexis.overinclusion)	redundancy; repeated/unnecessary lexical item (I have a friend who she is American / he would be against of this use of technology)
LX.CC (lexis.collocation)	wrong collocation (just like if the world was at his feet)
LX.DR (lexis.derivation)	word formation in either compounding, prefixation, and suffixation (the people who eat vegetables, fruits and nuts in your diet have less risk to development these diseases / that produces a lost in the relationship / all that lying was not even worth / that was completely unnecessary / a person with money has different opinions)
PN.CS (punctuation.comma splice)	comma splice (Nowadays, technology has a solution to almost anything, we have a lot of devices for doing different tasks.)
PN.CM.IP.OM (punctuation.comma.introductory phrase.omission)	missing comma after introductory word or phrase (In the past if you needed to look for information you went to the library)
SP.OS (spelling.orthographical case)	missing or unnecessary capitalization (Second, It is not good to eat a lot of carbohydrates.)
PN.CM.EL.OM (punctuation.comma.non-restrictive elements.omission)	missing comma for non-restrictive clauses (For example, carbohydrates which can be important food nutrients come in all sorts of food.)
SP.GR (spelling.grapheme)	Spelling: single (e.g., methode) or double grapheme (i.e., beautifull) either with consonant or vowels (Lieing in a resume should not be done /Second, It is not good to eat a lot of carbohydrates.)
PN.CM.CJ.OV (punctuation.comma.conjunction.ov erinclusion)	unnecessary comma before coordinating/correlative conjunctions joining words or phrases (However, I prefer to eat a lot, but in a healthy way, and do exercise.)
PN.CM.CJ.OM (punctuation.comma.conjunction.omission)	missing comma before coordinating/correlative conjunctions joining clauses (Nowadays, people are eating a lot of junk food and they are not aware of their eating habits.)
PN.FS (punctuation.fused sentence)	fused (run-on) sentence (In the past if you needed to look for information you went to the library now we have web pages to do so.)
PN.CM.AP.OM (punctuation.comma.appositive.omission)	missing comma with appositive (The principal of our school Don Jorge always told us that lying has consequences.)
PN.CM.VR.OB.OV (punctuation.comma.verb.object.ov erinclusion)	unnecessary comma between verb and object (They went to the supermarket and bought, lettuce, tomatoes, and all sorts of healthy food.)

Revista indizada en



Distribuida en las bases de datos:

