

Syntactic Complexity in Journal Research Article Abstracts Written in English¹

Rodrigo Tovar Viera², Technical University of Cotopaxi, Ecuador

Abstract

The reader's ability to connect new information to existing knowledge is crucial when reading a text. Nonetheless, text complexity, in many ways, is more linguistic than cognitive. It encompasses the degree of sophistication, and how challenging a reading section is. Depending on the section, such difficulty may appear on the vocabulary level, in the organizational structure, or with coherence, and cohesion. In the globalized world of scientific communication, research articles published in non-Anglophone academic journals require English abstracts to access an international database and citation possibilities. This paper describes the syntactic complexity in journal research article abstracts. A corpus of abstracts written in English and published in Anglophone and non-Anglophone contexts were sampled. The English sub-corpora underwent software-based text analysis using fourteen syntactic complexity measures with the second language (L2) Syntactic Complexity Analyzer (Lu, 2010). Significant differences appeared in only four of the fourteen syntactic indices between texts in Anglophone and non-Anglophone journals, and out of these fourteen measures, non-native groups reported thirteen lower mean values. The study affords insights for L2 writing research to produce accurate texts in content and structure. Ideally, findings will uncover pedagogical implications and applications for academic writing instructions.

Resumen

La capacidad de un lector de conectar información nueva con conocimiento previo es clave al momento de leer un texto. Sin embargo, la complejidad del mismo, en muchos sentidos, es más lingüística que cognitiva. Abarca el grado de sofisticación y el desafío que implica cada sección de lectura. Dependiendo de la sección, la dificultad puede aparecer en el nivel de vocabulario, la estructura organizativa, coherencia y cohesión. En el mundo globalizado de la comunicación científica, los artículos de investigación publicados en revistas académicas no anglófonas requieren resúmenes en inglés como sección indispensable para acceder a una base de datos internacional y posibilidades de citación. Este artículo describe la complejidad sintáctica en resúmenes de artículos de investigación. Para la conformación del corpus, se tomaron muestras de resúmenes escritos en inglés y publicados en contextos anglófonos y no anglófonos. El subcorpus anglófono se sometió a un análisis de texto basado en software utilizando 14 medidas de complejidad sintáctica con L2 *Syntactic Complexity Analyzer* (Lu, 2010). Se observaron diferencias significativas en solamente cuatro de los 14 índices sintácticos entre textos de revistas anglófonas y no anglófonas y, de las 14 medidas, los grupos no nativos demostraron 13 valores medios más bajos. El estudio aporta información para la investigación de la escritura en una segunda lengua para producir textos precisos en contenido y estructura. Idealmente, los resultados revelarán implicancias pedagógicas y aplicaciones para la enseñanza de la escritura académica.

Introduction

The importance of syntactic complexity in writing research in second and foreign languages has been recognized as an indicator of text difficulty (Lu, 2011), readability (Graesser et al., 2014), and the relationship between L2 writing and L2 proficiency (Ortega, 2003). Syntactic complexity as an explicit and varying component in text difficulty includes syntax. Frantz et al. (2015) thereby stated that grammar contributes to the meaning of the text and such grammatical meaning indirectly affects text comprehension. This analysis has been extended to scholarly works, such as research articles (RA) and research article abstracts (RAAs). The latter increased the analysis exponentially in both national and international academic production. After titles, abstracts in which authors summarize their works are the most widely read type of research literature freely available online. In many non-English-speaking countries, many journals require authors to include abstracts in English to accompany articles (Martín, 2003). The significant contribution of abstracts in disseminating the knowledge production have made them the subject of research studies across disciplines and languages (Ai & Lu, 2013; Hyland, 2004; Huang, 2018; Lee, 2018; Lorés, 2014; Lu, 2017; Yang et al., 2015). These studies have collectively brought useful insights into how texts display peculiarity patterns in rhetoric and written composition structures.

Accordingly, numerous studies have often been focusing on investigating RAAs from different dimensions, for example, writing quality (e.g., Lu, 2017; Yang et al., 2015), syntactic complexity (e.g., Ai & Lu, 2013; Lee, 2018), vocabulary richness (e.g., Crossley et al., 2016; Laufer & Nation, 1995; Meara, 2005), rhetoric patterns (Can et al., 2016; Lorés, 2004), and factors that influence their syntactic complexity (e.g., Lu & Ai, 2015; Ortega, 2003). Most of these contrastive studies have shown that, in many ways, the first language (L1) and prior L2 learning come to influence the ways of organizing ideas and structuring information when writing in English. The latter, 'structuring information,' attempts to describe possible differences between

¹ This is a refereed article. Received: 25 February, 2022. Accepted: 27 September, 2022. Published: 6 April, 2022.

² rodrigo.tovar@utc.edu.ec, 0000-0002-5960-3020

the writing compositions of the two groups of scientific discourse communities. The current work could account for the linguistic or cultural discrepancies or both, in the sense that RAAs have been found to be much more diverse and lexically dense than other types of writing. Although it is far from conclusive, predicting communicative-situated language use based on the assumptions of linguistic and cultural traits, “analytical research suggests that authors’ L1 and L2 schema differ in their preferred ways of organizing ideas which can influence academic writing.” (Hyland, 2014, pp. 401).

In many ways, syntactic complexity and text complexity are indicators of text comprehension and interpretation; the former rates writing accuracy through orthography, lexicon, and syntax, and the latter encompasses the challenges a reading section presents. The realization of these two linguistic elements can uncover specific features of the language embodied in a text. For instance, vocabulary level difficulty, organizational structure, and features of the language are complex from a linguistic rather than a cognitive setting, as they force the readers to make inferences to connect sentences. It is also assumed that the lexicon and grammar choices impact reading comprehension. Lu and Ai (2015), for instance, discovered that the length of the production unit decreases, and the degree of phrasal sophistication increases with writers’ language proficiency. Similarly, Ai and Lu (2013) found that the length of the production unit of the non-native English groups was shorter than those of the native ones. As Ortega (2003) pointed out, the variety of lexico-grammatical structures and the degree of sophistication of these structures determine the text complexity and, therefore, its processing difficulty.

Although a well-written text makes effective use of vocabulary and better organizational structure, writers’ language knowledge and their linguistic and cultural backgrounds affect text complexity and the production of intelligible texts. That is to say, the preferred lexical and grammatical choices determine “how easy or difficult a particular text is to read” (Common Core State Standards Initiative, 2010, p. 4). The levels of difficulty of the texts discussed above allow researchers to explore if RAAs published in (non-)Anglophone contexts share similar or different composing patterns, thereby equating syntactic complexity with text difficulty.

The research on abstracts written in English is further studied in Bondi and Lorés Sanz’s (2014) studies and in several other published investigations devoted to linguistic and pedagogical classroom applications (e.g., Swales & Feak, 2004; 2009). Therefore, understanding how RAAs published in different English-speaking contexts persuade readers to read the whole article has been the goal of many academics and researchers (e.g., Amnuai, 2019; Tseng, 2011). Although abstracts report variation in their rhetoric, syntactic complexity, and lexical richness (Tankó, 2017), research on this genre suggests writing accurate information is most important so the complete article gains indexing and citation possibilities. Despite the fact journal RAAs written in English have been extensively investigated, evidence of research literature in this type of text is scarce, particularly in Ecuador, except for the pioneering work of Tovar-Viera (2019a), who reported variability in the surface representation of the lexico-grammatical choices in agronomy abstracts. The scarcity of research on abstracts published in (non-)Anglophone academic contexts and their perceived linguistic status justify the research interest.

Despite the fact that many disciplines and fields are highly divergent and use to different discursive practices, it is crucial to investigate how their abstracts present the information content to hook readers’ attention. In particular for university-level researchers, for whom local or international publications are prerequisites for scholarly promotions, the abstracts written in English (or the translated version from Spanish) seem not to be of great interest. Indeed, writers perceive it as a mere requirement set by the journal editors (Lorés, 2014). The present study, then, examines RAAs written in English to describe their syntactic complexity. Therefore, the study aims at answering the following research question:

Are there differences in the syntactic complexity of English abstracts published in American and Ecuadorian journals (AEJ), if so, what are these differences?

Methodology

The study involves both descriptive and interpretative approaches; it is qualitative and quantitative research in nature. It focuses on measuring the writing quality in journal RAAs. The rationale of using these approaches was to display simple or complex structures and the sophistication of such composing patterns while interpreting text complexity. Concerning the research design, this is a comparative-descriptive study, in which the techniques’ selection includes both quantitative and qualitative procedures. For the quantitative (result-oriented) approach, the technique used was a content analysis of the output text of the computerized program using the SPSS statistical analyzer tool. Then, the *exegesis technique*; “a running commentary on

the product that reveals something of its dynamic unfolding as a process" (Halliday and Hasan, 1989 p. 11) so-called *thematic and discourse analysis overview*; and literature review was used when undertaking qualitative (process-oriented) data to interpret the research outcomes and describe linguistic, pedagogical, and possible cultural implications.

Data collection

The study analyzed two English sub-corpora of RAAs written in the disciplines of education and electronics and published between 2010 and 2018 in the fields of science and humanities. The rationale for choosing education and electronics disciplines is because according to Biber and Gray (2016) "the language of science research writing is quite different from the language of humanities prose." (p. 7) It claims the assumption that academic texts in different knowledge fields present different types of complexity. In this way, these types of texts can "employ complex and elaborated grammar; exhibit an explicit maximum meaning; and be conservative and resistant to linguistic change" (p.7). Another reason is because since the 2008 constitutional reforms of Ecuador there have been significant increases in the number of academic publications in science, social science, and humanities, to nearly 13.227 in April, 2021, according to a search in *Scimago* (<https://www.scimagojr.com>). Thus, research related to education (894), on the one hand, and those of electronics (3779), on the other, have been the foci of researchers, practitioners, and school administrators to disseminate the need for academic processes to link higher educational institutions with the society in general.

The research explored the education and electronics journal RAAs because, according to Ecuador's Constitution (Cancillería de Ecuador, 2008), these disciplines are the bases for innovation, promotion, development, and dissemination of knowledge and culture internationally. These areas are also the professions with ample job opportunities (Instituto Nacional de Estadística y Censos, 2018) and interest in the higher education system; they are linked to the global and technological production to generate scientific and technological research. Academics must share their research findings to gain knowledge in their studies and improve academic writing skills, mainly when writing abstracts of scientific papers in English.

Table 1 illustrates in detail the characteristics of the two datasets of article abstracts. Although there are considerable differences, for instance, in the mean length of abstracts between the two sets of groups, these differences do not affect this type of comparison and contrastive analysis. This is, in great part, because the fourteen syntactic complexity dimensions are calculated according to the degree and sophistication of composing patterns used independently from one text to another (Lu & Ai, 2015). Following the criteria collection employed in Tovar (2019b), the two English sub-corpora were compiled for the examination. First, research articles published in the periods 2010 – 2018 were collected (288 RAA: 144 in Anglophone and 144 in non-Anglophone journals). Those texts reporting a peer-review and editorial scrutiny were then considered; this to ensure that the genre-based analysis represents the context and text of abstracts written in education and electronics in Ecuador and American journals. American journals were selected because they endorse the conventional American English model to which Ecuador adheres to its teaching practices and academic and professional development methods. Therefore, each corpus in AEJ contains "abstracts selected from journals that meet the criteria of a) representativeness—*appropriate sample testing group*, b) reputation—*indexation and double-blind peer-reviewed*, and c) accessibility—*in print or online database*". (Tovar, 2019b, p. 77)

Context	Discipline	RAA	Abstracts Length		# sentences per abstract		Sentence length		Total words
			Mean	SD	Mean	SD	Mean	SD	
American	Education	30	164.11	23.97	6.80	2.04	19.82	12.28	4162
	Electronics	30	192.03	53.61	8.23	2.79	22.73	10.16	5761
Ecuadorian	Education	30	172.97	52.16	6.33	2.53	37.68	7.68	5189
	Electronics	30	163.56	44.16	5.88	2.18	35.20	14.88	4097
TOTAL	2	120							19209

Table 1. Summary of the dataset.

Corpus selection

Journal research article abstracts that represent the "national or international indexing, double-blind peer-reviewed, unstructured text, and single paragraph condensed summary" (Tovar, 2019b, p. 77) were part of

the corpus-based design criteria. Additionally, the journal RAA should have been published between 2010 and 2018 in (non-)Anglophone speaking contexts. Then, the publication context is the focus of the abstracts' selection rather than the nativeness of academic writers.

The data set of American journals includes 60 English RAAs published by the *American Journal of Education* (AJE), *Journal of Teacher Education* (JTE), *Journal of Electronic Materials* (JEM), *Journal of Electronic Packaging* (JEP). The rationale for choosing Ecuadorian journals was their accessibility and the verified record of the indexing process in Scopus. This process includes highest percentage of the degree of English language proficiency, rhetorical and writing style, knowledge of the disciplinary field, and the overall level of comprehensibility (88%, 72%, 72%, 71%, respectively). Thus, the content quality and quantitative measures of the articles published in these journals are expected to be based on standard scientific English accepted inside research and discourse communities. Additionally, these journals reported a respectable level of publication ranking and citation as domestic and international journals; they are indexed in some of the international databases to which non-anglophone journals belong. Such characteristics are positioning them as journals to guarantee controlled comparison analysis. It was, therefore, logical to expect that the hierarchy of these journals be the same as the content quality deployed in the sample RA abstracts. The corpus of each discipline contains thirty RA abstracts. These journals are indexed in *ERIC*, *EBSCOhost*, *Elsevier*, the American statistical association, *SciSearch*, *SCOPUS*, *Applied Science and Technology*, and *GoogleScholar*.

Similarly, the Ecuadorian corpus consists of sixty abstracts written in English and published in Ecuador in the following journals: *Alteridad* (journal of education), *UTCiencia* (science and humanities), *Revista Tecnológica ESPOL* (science & humanities), *Sophia* (philosophy & education), *ACI Avance* (science & engineering), *Enfoque* (scientific engineering journal), *Ingenius* (science and technology), *Maskay* (electric & electronics). It encompasses thirty abstracts in each discipline Ecuadorian journals report indexing in *Latindex* (regional cooperative online information system for scholarly journals from Latin America, the Caribbean, Spain, and Portugal) and an international database, like *Elsevier*, *DOAJ*, *Dialnet*, *REDIB*, *DRJI*, *SIS*, *MIAR*, *SciELO*, and *EBSCOhost*. Thus, sample articles from Ecuadorian journals were expected to employ the informational sections of their abstracts precisely and skillfully per se compared to those from American journals. The higher education system hosts Ecuadorian journals. Unlike American journals, Ecuadorian ones are mixed, which devote special sections and space for the disciplines mentioned earlier.

Corpus sampling procedure

After selecting journals from each discipline, a corpus of 288 abstracts was listed: 144 texts from Anglophone journals and 144 from non-Anglophone ones. Both Anglophone and non-Anglophone journals included 72 abstracts per discipline, like education and electronics. Using the random sampling selection excel *fx=randbetween* program, out of 288 texts, we obtained the final 120 RAAs: sixty abstracts in Anglophone journals and sixty in non-Anglophone journals. To give the *fx=randbetween* program functionality, abstracts, in a spreadsheet, were first numbered from 1 to 144 with their respective discipline and publication context. Then, the *fx=randbetween* function was activated, and the selection of thirty abstracts per discipline started.

Once the choice of the texts ends, abstracts were rearranged through a normalization process using the function *fx=dataremoveduplicates* to eliminate double or repeated texts. After completing this process, the sampling corpus registered thirty abstracts in the required disciplines. Finally, the functions *VLOOKUP* (lookup_value, table_array, col_index_num, [range_lookup]) and *SEARCH* (find_text, within_text, [start_num]) identified the discipline and its publication context.

Data analysis and instruments

The two English subcorpora underwent software-based text analysis. With the top-down and bottom-up approach, sentences were the unit analysis of the study. The top-down approach focuses on the information content while the bottom-up approach looks for linguistic signals. L2 Syntactic Complexity Analyzer—L2SCA (Lu, 2010) gauges the syntactic complexity of journal RAAs. It measures fourteen syntactic dimensions covering (1) length of production units, (2) amounts of coordination, (3) amounts of subordination, and (4) degree of phrasal sophistication. L2SCA provides a comprehensive assessment of the sophistication and complexity of writing samples by calculating and returning the frequency counts of the fourteen indices of syntactic complexity use in the individual or set of texts, as can be seen in Table 2.

Measure	Dimensions	Code	Definition
Length of production unit			
	Mean length of sentence	MLS	# of words / # of clauses
	Mean length of T-unit	MLT	# of words / # of sentences
	Mean length of clause	MLC	# of words / # of T-units
Amount of subordination			
	Clauses per T-unit	C/T	# of clauses / # of T-unit
	Complex T-units per T-unit	CT/T	# of complex T-units / # of T-units
	Dependent clauses per clause	DC/C	# of dependent clauses / # of clauses
	Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Amount of coordination			
	Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
	Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
	T-units per sentences	T/S	# of T-units / # of sentences
Degree of phrasal sophistication			
	Complex nominals per clause	CN/C	# of complex nominals / # of clauses
	Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
	Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units
	*Clauses per sentences	C/S	# of clauses / # of sentences

Note: * Overall sentence complexity

Table 2. Syntactic complexity features

The output texts of the L2SCA software analyzer were mined using SPSS Statistics to determine text complexity. Accordingly, after obtaining the syntactic complexity indices for each English-sub corpus, a set of independent sample t-tests were run to compare and contrast RAA published in AEJ. Additionally, one-way ANOVAs were run to compare the two non-Anglophone written texts against the Anglophone groups. L2SCA, a web-based program, was chosen because of its accessibility, high reliability, and extensive use as an automatic text analyzer in a large number of contrastive studies. This research employed qualitative analysis supported by quantitative data from computerized programs. After all, the purpose of the study was to examine and describe the written complexity of abstracts in actual use of English language. Thus, the study consisted of the segmentation of the data collected from the quantitative approach, such as coding, selection, and categorization of relevant and meaningful information or clearance data irrelevant related to the research goal. That is to say, the continuous decision-making process was essential to set up data interpretation with the use of the *exegesis technique* (Halliday and Hasan, 1989).

For the quantitative design, the two English corpora, abstracts published in Anglophone speaking contexts (NG) and those of non-Anglophone speaking contexts (NNG), were submitted to the Web-based L2SCA: Batch Mode to analyze the fourteen syntactic complexity measures and dimensions of abstracts published in both academic journals. The output texts of L2SCA were subsequently imported into spreadsheets for a further statistical analysis using SPSS, presented in tables to visualize the results. Finally, up to four English sub-corpora, 2 in NNG and 2 in NG, were created to examine the syntactic complexity of abstracts in education and electronics disciplines; this analysis followed the same Web-based L2SCA design and statistical process mentioned above.

When investigating the syntactic complexity of texts, Hunt (1966) indicated that the length of T-unit is a better index of *maturity* —the average number of T-units per sentence. I will call it *L2 knowledge* wherein learners can rewrite sentences by the process of reduction and consolidation. According to Hunt, a T-unit is “the shortest units into which a piece of discourse can be cut without leaving any sentence fragments as residue; they are thus minimal terminable units” (p. 737). The author also claimed that nearly half the sentences could encompass a single T-unit in everyday prose, and the other half may contain two or more T-units, frequently connected with and’s conjunctions; such sentence types are called compound or compound-complex. For instance, a sentence may have two or more T-units when independent clauses (main clauses) are linked, as in sentence 1, and a single T-unit appears when one or more clauses (subordinate clauses) are embedded in the main clause like in sentence 2.

Sentence 1. There are many students in the Language Center, and they (students) study English as a foreign language. (S + S) = two T-units (two main clauses)

Sentence 2. There are many students in the Language Center who study English as a foreign language. [S (-s)] = one T-unit

Results

The cross-disciplinary and linguistic analysis looks for differences in the dimensions of syntactic complexity. Research outcomes (see Table 3) shows that out of the fourteen syntactic complexity measures in the study, the NNG had thirteen lower mean values than the NG. That is, just the T-units per sentence (T/S) had higher values ($M= 0.788, SD= 0.019$) than the others. Independent sample t-tests were run to determine if the NNG and NG groups' differences were statistically significant. Results suggest that the two groups (NG and NNG) significantly varied in the length of production unit, particularly in MLS indices ($t= -2.155, p = .05$); amount of coordination as of CP/C ($t= -3.622, p = .005$), CP/T ($t= - 3.783, p = .004$); and the degree of phrasal sophistication, namely CN/C ($t= -5.245, p= .001$), CN/T ($t= -5.953, p = .001$), and VP/T ($t= - 3.819, p = .003$).

The sample t-test between NG and NNG groups ($t= 2.755, p = .010 < .05$), and between the two disciplines (education: $t=1.939, p = .03 < .05$) and (electronics: $t= -1.791, p = .04 < .05$), revealed significant differences and variations among the syntactic complexity dimensions. In the same way, outcomes of running a one-way t-test inform that the two groups considerably differed in two amounts of coordination gauges. Thus, CP/T and CP/C reported significant variability with p-values of .004 and .005, at both Anglophone and non-Anglophone written texts.

Measure	Code	NNG Mean	NG Mean	t	p
Length of production unit					
Mean length of sentence	MLS	46.039	50.609	-2.155	.050
Mean length of T-unit	MLT	58.456	65.942	-2.294	.060
Mean length of clause	MLC	46.709	51.071	-2.943	.103
Amount of subordination					
Clauses per T-unit	C/T	1.251	1.291	-1.567	.148
Complex T-units per T-unit					
Dependent clauses per clause	DC/C	0.126	0.155	-1.444	.179
Dependent clauses per T-unit	DC/T	0.190	0.211	-1.462	.174
Amount of coordination					
Coordinate phrases per clause	CP/C	0.238	0.273	-1.693	.121
Coordinate phrases per T-unit					
T-units per sentences	CP/T	0.130	0.197	-3.622	.005
	CP/T	0.163	0.255	-3.783	.004*
	T/S	0.788	0.770	.932	.373
Degree of phrasal sophistication					
Complex nominals per clause	CN/C	1.445	1.632	-5.245	.001*
Complex nominals per T-unit	CN/T	1.806	2.107	-5.953	.001*
Verb phrases per T-unit					
	VP/T	1.292	1.361	-3.819	.003*
Sentences complexity					
Clauses per sentences	C/S	0.961	0.994	-1.285	.228

Note: *statistical significance at the p level of $p < .05$

Table 3. Mean values and fourteen syntactic complexity indices in the NNG and NG

Lu and Ai (2015) recognize the synergy of complexity, accuracy, and fluency (CAF) as the main feature to measure language proficiency and writing quality. In practice, complexity or more precise linguistic complexity, looks for "the extent to which the language produced when performing a task is elaborated and varied" (Ellis, 2003, p. 340), characterized by the diversity of syntactic structures and the writing sophistication of those composing patterns. In this way, data in Table 4 reported differences in the syntactic complexity indices between the two disciplines, at both NG and NNG groups. In particular, the length of the production unit shows higher degrees of variability in MLS, MLT, and MLC (see notation in Table 2) compared to the gauges of the other syntactic indices that have lower syntactic variability (see Table 3, for a detailed analysis between the mean values). Abstracts published in an Anglophone speaking context had the higher indices of production unit than those of non-Anglophone speaking contexts. However, electronics texts in NG indicated the highest variability (SD) and highest figures (mean) in the three syntactic dimensions, as MLS, MLT, and MLC (see Table 4 below).

Sub-Corpora	Discipline	Length of production unit					
		MLS		MLT		MLC	
		mean	(SD)	mean	(SD)	mean	(SD)
NNG	Edu	44.210	(1.083)	56.227	(1.744)	46.050	*(2.703)
	Elec	47.868	*(3.992)	60.686	*(4.781)	47.368	(1.386)
NG	Edu	50.665	(1.739)	63.559	(1.795)	49.738	*(2.291)
	Elec	50.555	*(6.110)	68.325	*(11.231)	52.403	*(8.214)

Note: * higher degree of variability

Table 4. Mean of length of production unit

Significant variability was observed in the amount of subordination and coordination (see Table 5, below). Written texts in Anglophone journals used more subordination than non-anglophone ones; for instance, the electronics abstracts of texts in Anglophone journals produced significantly more C/T ($M=1.303, SD= .036$) and DC/T ($M= 0.273, SD= .028$) than the less production of electronics in non-Anglophone written texts ($M=1.228, SD=.187; M=0.236, SD=.040$). Regarding the amount of coordination, Anglophone groups significantly differed from those of non-Anglophone ones. Nonetheless, education and electronics texts, in non-Anglophone groups, made minor uses of coordinating sentences ($M=0.360, SD=0.302$) compared to Anglophone ones ($M=0.407, SD=0.258$). The average lower occurrence of coordination in non-Anglophone written text, according to Norris and Ortega (2009), may be an indicator of writers' low levels of L2 knowledge, as well as the lack of academic writing instructions.

Sub-Corpora	Discipline	Amount of subordination				Amount of coordination		
		C/T	CT/T	DC/C	DC/T	CP/C	CP/T	T/S
		<i>mean</i> (<i>SD</i>)	<i>mean</i> (<i>SD</i>)	<i>mean</i> (<i>SD</i>)	<i>mean</i> (<i>SD</i>)	<i>mean</i> (<i>SD</i>)	<i>mean</i> (<i>SD</i>)	<i>mean</i> (<i>SD</i>)
NNG	Edu	1.222 (0.051)	0.146 (0.019)	0.197 (0.036)	0.241 (0.007)	*0.118 (0.008)	*0.144 (0.0260)	*0.787 (0.024)
	Elec	1.280 (0.055)	0.106 (0.187)	0.183 (0.065)	0.236 (0.040)	*0.143 (0.060)	*0.182 (0.017)	0.789 (0.037)
NG	Edu	1.279 (0.028)	0.177 (0.037)	0.214 (0.026)	0.273 (0.022)	0.213 (0.032)	0.273 (0.024)	0.797 (0.018)
	Elec	*1.303 (0.036)	0.132 (0.048)	0.209 (0.022)	*0.273 (0.028)	0.181 (0.040)	0.236 (0.038)	0.743 (0.029)

Note: Anglophone group = NG, non-Anglophone groups = NNG
Degree of variability *

Table 5. Mean of variability in coordinating and subordinating indices

Consistent with the analysis for the other two syntactic complexity measures, data in Table 6 presents variation in the dimensions of phrasal sophistication and overall sentence complexity between Anglophone and non-Anglophone groups. However, education and electronics disciplines in Anglophone groups compared to those of non-Anglophone ones, reported significant differences in the average use of phrasal sophistication, namely complex nominals as of CN/C ($M=1.633$), CN/T ($M=2.107$) and verb phrases as VP/T ($M=1.361$). Non-Anglophone written texts, particularly education, showed an average lower incidence of clauses ($M=0.553$), as well as a lower occurrence of complex sentences ($M=0.962$) than their Anglophone counterparts.

Sub-corpora	Discipline	Degree of phrasal sophistication						Sentences complexity	
		CN/C		CN/T		VP/T		C/S	
		<i>mean</i>	(<i>SD</i>)	<i>mean</i>	(<i>SD</i>)	<i>mean</i>	(<i>SD</i>)	<i>mean</i>	(<i>SD</i>)
NNG	Edu	1.462	(0.039)	1.786	(0.034)	1.293	(0.124)	*0.962	(0.109)
	Elec	1.428	(0.007)	1.826	(0.007)	1.383	(0.068)	1.009	(0.019)
NG	Edu	1.662	(0.039)	2.125	(0.028)	1.372	(0.095)	1.019	(0.045)
	Elec	1.603	(0.060)	2.089	(0.043)	1.350	(0.085)	0.968	(0.039)

Note: Anglophone group = NG, non-Anglophone groups = NNG
Degree of variability *

Table 6. Indices of sophistication and sentence complexity

The writing excerpts (see qualitative phase) illustrate that electronics texts produce an average of 6-8 words before the main verb along with their sentences compared to the 3-4 words' occurrence before the main verb in sentences of education texts. Additionally, further analysis revealed that there are sentences of electronics abstracts in NG and NNG that contain more than ten words before the main verb, as in excerpts 9, 10, 11, and 12. These composing patterns make electronic texts challenging to read, as they tend to demand heavier processes on working memory to understand them (McNamara et al., 2014), which require the reader to make more inferences. In this way, Bailey and Butler (2007) highlighted that sciences texts had the most challenging difficulty to read and process due to their higher percentage of complex sentences. In other words, in order to understand the content, these abstracts need double reading to see how those sentences relate to each other.

Further data analysis found that most of the RAA is constituted of 5-8 sentences in a paragraph. Thus, sentence length ranged between 23-50 words. Accordingly, paragraph length in soft sciences "(education)" consisted of 93-375 words, whereas in hard sciences "(electronics)", it ranged between 92 and 340. Half of

the Ecuadorians' texts contain an average length of 168.27 tokens —words per abstract compared to 178.07 mean lengths of words in American texts.

Qualitative phase

This section describes the cross-linguistic and contrastive analysis of research article abstracts across disciplines. In this part, a qualitative technique supported on quantitative data allows classifying linguistic features used by writers in the two disciplines in NG and NNG. It was interesting to find out that most of the abstracts showed variability in their surface structures, such as left embeddedness —*the number of words before the main verb*. This writing construction, depending on the left-embeddedness words, makes sentences become more complex and, consequently, the text is challenging to read and process since this type of structure increases the load on the interpreter's cognitive resources to comprehend, as in the examples below, particularly in sentences 9, 10, 11, and 12.

1. Site-specific on-demand cooling of hot spots in microprocessors **can reduce** peak temperature and achieve a more uniform thermal profile on chip, thereby improve chip performance and increase the processor's lifetime. (Electronics– NG, 8 words before the main verb)
2. In this work, a stretchable conductive sensor **has been developed** using single-walled carbon nanotubes (SWCNTs) and monofunctional acrylate monomers (cyclic trimethylolpropane formal acrylate and acrylate ester). (Electronics– NG, 7 words before the main verb)
3. The decrease of interfacial electrical conductance **triggers** steady-state temperatures over copper and silicate melting points and, in consequence leads to temperature high enough to explain the physical degradation. (Electronics– NNG, 6 words before the main verb)
4. One of the main objectives of traffic studies **is to determine** the existing vehicular traffic on a particular stretch of highway under study. (Electronics– NNG, 8 words before the main verb)
5. Together, these orientations **revealed** a sense of agency over student engagement that was more evident in the comments of more engaging teachers. (Education– NG, 3 words before the main verb)
6. Little research currently **examines** language instruction educational programs (LIEPs) in states with a more recent growth of the Latino English learner population (Education– NG, 3 words before the main verb)
7. The criterion validations process **was carried out** by judges who were researchers that recommended the use of specific learning strategies for each style. (Education– NNG, 4 words before the main verb)
8. The results only **established** significant correlations between the socio educational variable high school (educational institution of origin) and the Dimension 1 named Cognitive and learning control strategies (Education– NNG, 3 words before the main verb)
9. Reflection high-energy electron diffraction and atomic force microscopy of Ge annealed under similar conditions on GaAs and Al_{0.3}Ga_{0.7}As surfaces **revealed** the gradual suppression of QD formation with decreasing Al-content of the buffer. (Electronics – NG, 16 words before the main verb)
10. Interdot transitions in the emission spectra of a quantum dot molecule **may be used** as a sensitive nanoscale probe to measure electric fields. (Electronics– NG, 11 words before the main verb)
11. The design procedure of the fuzzy controller as declaring variables, the fuzzification, the inference mechanism, the knowledge base and defuzzification of the system **are discussed** in this work. (Electronics– NNG, 22 words before the main verb)
12. The temperature response of the model to independent variation of electrical resistivity of studied materials, and interfacial electrical and thermal conductance between the copper contact and its diffusion barrier **were obtained**. (Electronics– NNG, 29 words before the main verb)

Research outcomes show that the syntactic complexity of NG and NNG significantly varies in terms of structure and that variation may be affected by several factors, such as, L2 knowledge, discursive practices, and context-related publication. This finding is in line with those of Ai and Lu (2013) and Lu and Ai (2015) who discovered that native English texts were found to be more syntactically complex than non-native ones. From the writing excerpts, we can point out that in terms of "syntactic complexity and writing quality, writers with different L1 backgrounds, even for those at the same or comparable proficiency levels, may not develop in the same ways" (Lu & Ai, 2015, p.26). Accordingly, when contrasting native and non-native English

groups, it is essential to take several context-related factors into account, as genre, discipline, academic or non-academic writing, publication setting, among others (e.g., Lu, 2011; Ortega, 2003). It is also because authors tend to follow the writing conventions and practices of those institutionalized discourse communities to accomplish the communicative purpose and position the work within the research communities (Hyland, 2004).

As mentioned in the literature review, empirical studies have claimed the probable relationship of syntactic complexity indices to language proficiency impacting writing quality. For example, Ortega (2003) reported correlations of syntactic complexity with L2 proficiency, whereas Lu (2010, 2011, 2017) informed relationships of syntactic complexity to L2 writing quality. It is known that complex ideas can be hidden in a coherent and cohesive text by using well-organized syntactic structures, combining clauses and phrases. In contrast, some long and embedded sentence constructions indicated a lack of clarity in the present study, which created a complex structure and increased readers' cognitive resources for its interpretation. Therefore, syntactic complexity indices are a good indicator of the quality of writing since they award the sentence constructions as comprehensible. Although writing quality is impacted by other factors besides syntactic complexity such as L2 knowledge, sentence patterns of academic writing would have practical pedagogical implications for less experienced writers. It is because they reflect the linguistic features for conveying meaning effectively (Biber & Gray, 2016).

Discussion

Lately, and more than ever before, genre-based article abstracts have become the data for analysis on contrastive studies. The descriptive statistics presented in Table 1 show abstract variation in the two disciplines. It was observed that the abstracts in both AEJ journals usually included six sentences. The mean length of abstracts (192.03 words), and the mean number of sentences per abstract (8.23) in electronics texts were longer compared to education journal abstracts.

Nonetheless, abstracts published in Ecuadorian journals outperformed American ones in sentence length. In this way, education texts with the mean of 38 words reported having the highest frequent length of sentences. Although the differences were statistically significant, they do not affect the linguistic and disciplinary analysis. Rather, research outcomes suggest that such variability may respond to different requirements set by the journal publishers and different discursive practices across disciplines. This result is in connection with those of Tseng (2011) and Amnuai (2019), who noticed similar statistical differences in native and non-native texts. Writers, therefore, anticipate what readers and scientific discourse communities expect from a text. Apart from this, writers, for example, must build up the content of the abstracts, cutting out any redundancy, ambiguity, and lack of clarity. In this way, authors are likely to respond to these requirements using rhetorical and writing conventions, grammatical features, and argument structure, indicating membership of the academic community of those disciplines (Hyland, 2014). These linguistic differences are inextricably bound up with language use; these allow us to understand the preference for lexico-grammatical choices and the way writers present the information to the audience. Furthermore, the differences between the writing compositions of the two groups of the RA articles could account for the linguistic or discursive convention discrepancies or both.

The results discussed above indicate that although the average length of thirty words is the standard in a sentence (Biber & Conrad, 2009), written RAAs on education ($M=38$, $SD=7.68$) and electronics ($M=35.2$, $SD=14.88$) published in Ecuadorian journals contain longer sentences than the American journals, and compared to other disciplines, for instance, linguistics in Holtz (2011), where 24–27 words was the norm. Thus, abstracts written in Ecuadorian journals, in the current study, outdid the average of thirty words. Because sentence length is a good indicator of syntactic complexity (Biber & Conrad), exceeding the average word count made these abstracts grammatically complex. On the other hand, electronics texts indicated more left embeddedness ($M=6.233$), words before verb of the main clause, compared with education ones ($M=4.280$), in both NG and NNG; see examples 1, 2, 3, 4, 5, 6, 7 and 8. Accordingly, both education and electronics abstracts in NG reported robust indices of writing sophistication but a difficulty in processing or reading comprehension since their average 2.107 complex nominals per T-unit (CN/T) outperformed NNG texts (1.806).

Abstracts in NG counted for more CN/T than their NNG counterpart. Even though electronics abstracts published in an Anglophone and non-Anglophone speaking context slightly varied the average use of CN/T (1.958) in comparison with education texts (1.955), no significance was found at the $p 0.05$ level. Nonetheless, when comparing electronics texts in NNG between education ones in NG, statistical differences

emerged ($t = -7.73$, $p = 0.01$). Similarly, there was a significant difference ($t = 11.08$, $p = 0.004$) between electronics abstracts in NG and education ones in NNG. It can be concluded; therefore, that electronics abstracts in terms of length of the production unit and phrasal sophistication are syntactically complex. That is, the high syntactic complexity indices and the high information content, tend to make these abstracts challenging to process.

Syntactic complexity, defined as “the range of forms that surface in language production and the degree of sophistication of such forms” (Ortega, 2003), has been widely investigated. For instance, Ai and Lu (2013), gauging the syntactic complexity of native and non-native written English compositions, reported significant differences in these dimensions, CP/T and CP/C. The findings of Ai and Lu result from the nativeness and non-nativeness of the texts, yet the current study is connected to those research outcomes; this is because both studies account for writing divergence. The present research and the prior work cited above demonstrate that variation in academic writing is affected not only by the nativeness of the text, but by the context of publication. Nonetheless, the nativeness of the writers may be mistrusted since there is always the probability of the abstracts being totally written or at least edited by a native or a native-like English speaker. Thus, the research findings in the present work are taken as the result of the product of different probable actual stages of production and not as the revision of native or native-like minds.

Although it might be possible that journals from another English-speaking country (England, Canada, Australia) could have different cultural norms in writing that would impact syntactic complexity in abstracts, the current research takes American journals as the conventional source to compare English abstracts published there with those published in a non-anglophone speaking country. The content and structure employed in the texts may respond to the cross-cultural variation between Anglophone and non-Anglophone speaking countries. Although it is far from conclusive, stating that culture is the only factor determining academic writing, communicative-situated language use suggests that language conventions influence organizing the information content. Culture remains a controversial term; nonetheless, the cross-cultural difference allows describing the merits commonly attributed to the abstracts published in Ecuadorian journals and the possible demerits of those published in Ecuador. In this way, the generic conventions may not simply be seen as a blueprint for further replication in similar rhetorical contexts but more as a resource to understand the generic conventions that make the genre possible (Bhatia, 1999).

Conclusions

Writing should be accurate, clear, concise, engaging, objective, coherent, and readable (Hacker & Sommers, 2016) so that readers can understand and choose from the hundreds of published articles published in the scientific community. Research outcomes provide valuable insights into how the information content, including syntactic complexity, should be perceived and employed in English as a Foreign Language (EFL) writing instructions and pedagogy. Even though the study does focus on the context-related publication of journal research article abstracts, the lexical knowledge, grammatical and syntactic processing seemed to be indirectly affected by the discursive practices, cultural and linguistic L1 backgrounds of the writer. Therefore, a solid understanding of writing-related differences in the rhetoric and syntactic complexity is necessary to develop appropriate pedagogical interventions for non-anglophone speakers of the language benefit, not only from writing instructions but from issues to context-related factors.

Awareness of such writing-related differences helps language instructors to understand that composing patterns are not essentially indicative of L2 proficiency in the same way for learners with different L1 backgrounds and L2 knowledge. Nonetheless, these differences may be seen as linguistic features that bring writing and pedagogical implications by addressing non-casual links between texts published in (non-)Anglophone-speaking countries. Accordingly, writing-related differences are the scaffolding to explore whether such composing dissimilarities do exist, describing what they are and how they are built up in the texts. It is hoped that these findings can guide further research on other disciplines and contribute to suggesting writing training and teaching programs, which can be adapted to the curricula and language teaching. It is also hoped that the research outcomes have linguistic implications and pedagogical applications for L2 writers' benefit from writing instructions and learning materials to produce accurate texts in terms of content and structure.

We can conclude that the generalizability of these results is subjected to certain limitations; for instance, sample size and two chosen disciplines. Nonetheless, these findings allow the description of the lexical and grammatical choices that tend to recur in genre-specific texts and how the surface representation of such lexico-grammatical choices reflect the linguistic features of the language in use. Future studies could

investigate composing patterns of journal RAAs in different disciplines to see the extent to which those results will match with the current study or vary within a large sample size of abstracts.

Reference

- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). John Benjamins. <https://doi.org/10.1075/sci.59.15ai>
- Amnuai, W. (2019). Analyses of rhetorical moves and linguistic realizations in accounting research article abstracts published in international and Thai-based journals. *SAGE Open Journals*, 9(1), 1-9. <https://doi.org/10.1177/2158244018822384>
- Bailey, A. L., & Butler, F. A. (2007). A conceptual framework of academic English language for broad application to education. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 68–102). Yale University Press.
- Bhatia, V. K. (1999). Integrating products, processes and participants in professional writing. In C. N. Candlin, & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 21-39). Longman.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- Bondi, M., & Lorés Sanz, R. (2014). *Abstracts in academic discourse: Variation and change*. Peter Lang.
- Can, S., Karabacak, E., & Qin, J. (2016). Structure of moves in research article abstracts in applied linguistics. *Publications*, 4(3), 1-16. <https://doi.org/10.3390/publications4030023>
- Cancillería de Ecuador (2008). *Sección octava: Ciencia, tecnología, innovación y saberes ancestrales* [Section eight: Science, technology, innovation and ancestral knowledge.] *Constitución del Ecuador*. https://www.cancilleria.gob.ec/wp-content/uploads/2013/06/constitucion_2008.pdf
- Common Core State Standards Initiative. (2010). Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects: Appendix A. http://www.corestandards.org/assets/Appendix_A.pdf
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16. <https://doi.org/10.1016/j.jslw.2016.01.003>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Frantz, R. S., Starr, L. E., & Bailey, A. L. (2015). Syntactic complexity as an aspect of text complexity. *Educational Researcher*, 44(7), 387–393. <https://doi.org/10.3102/0013189X15603980>
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, 115(2), 210–229. <https://doi.org/10.1086/678293>
- Hacker, D., & Sommers, N. (2016). APA version: A pocket style manual (7th ed.). Macmillan Education Imprint.
- Holtz, M. (2011). *Lexico-grammatical properties of abstracts and research articles: A corpus-based study of scientific discourse from multiple disciplines*. [Unpublished doctoral dissertation], Technischen Universität Darmstadt. <https://bit.ly/2LV1k9k>
- Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English*, 43(7), 732-739. <http://www.jstor.org/stable/41386067>
- Huang, J. C. (2018). Marine engineering and sub-disciplinary variations: A rhetorical analysis of research article abstracts. *Text & Talk*, 38(3), 341–363. <https://doi.org/10.1515/text-2018-0002>
- Halliday, M. A. K. & Hasan, R. (1989). *Language, context and text: Aspects of language in a social-semiotic perspective* (2nd ed.). Oxford University Press.
- Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*: University of Michigan.
- Hyland, K. (2014). English for academic purposes. In Leung, C. & Street, B.V. (Eds.) *The Routledge companion to English studies* (pp. 392-404). Routledge.
- Instituto Nacional de Estadística y Censos (2018) *Encuesta nacional de empleo, desempleo y subempleo (ENEMDU)*. [National survey of employment, Unemployment and underemployment]. <https://www.ecuadorencifras.gob.ec/enemdu-2018>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Lee, J. (2018). A Coh-Metrix analysis of lexical, syntactic and discourse aspects in the newspaper articles of Korean and British university students. *Modern English Education*, 19(4), 17–26. <https://doi.org/10.18095/meeso.2018.19.4.17>
- Lorés, R. (2004). On RA abstracts: from rhetorical structure to thematic organisation. *English for Specific Purposes*. 23(3), 280-302. <https://doi.org/10.1016/j.esp.2003.06.001>
- Lorés, R. (2014). Lost (and Gained) in translation: A contrastive (English/Spanish) analysis of rhetorical and lexico-grammatical patterns in sociology RA abstracts. In M. Bondi & R. Lorés Sanz (Eds.), *Abstracts in academic discourse: variation and change*, (pp. 84–109). Peter Lang.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493–511. <https://doi.org/10.1177/0265532217710675>
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27. <https://doi.org/10.1016/j.jslw.2015.06.003>
- Martín Martín, P. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, 22(1), 25-43. [https://doi.org/10.1016/S0889-4906\(01\)00033-3](https://doi.org/10.1016/S0889-4906(01)00033-3)
- McNamara, D. S., Graesser A. C., McCarthy P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Meara, P. (2005). Lexical frequency profiles: A Monte Carlo Analysis. *Applied Linguistics*, 26(1), 32–47. <https://doi.org/10.1093/applin/amh037>
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to Investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. <https://doi.org/10.1093/applin/amp044>
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>

- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills*. (2nd ed.). University of Michigan.
- Swales, J. M. , & Feak, C. B. (2009). *Abstracts and the writing of abstracts*. The University of Michigan.
- Tankó, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes*, 27, 42–55. <https://doi.org/10.1016/j.jeap.2017.04.003>
- Tovar-Viera, R. (2019a, November 23-24). *Rhetorical organization and linguistic realizations of moves in English research article abstracts* [Paper presentation]. International Congress on the Didactics of the English Language-PUCESE, Esmeraldas, Ecuador.
- Tovar, R. (2019b). Rhetorical move structure in abstracts of research articles published in Ecuadorian and American English-speaking contexts. *Arab World English Journal*, 10(4) 74-87. <https://dx.doi.org/10.24093/awej/vol10no4.6>
- Tseng, F.-P. (2011). Analyses of move structure and verb tense of research article abstracts in applied linguistics. *International Journal of English Linguistics*, 1(2), 27-39. <https://doi.org/10.5539/ijel.v1n2p27>
- Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53–67. <https://doi.org/10.1016/j.jslw.2015.02.002>