

Summer 8-20-2019

Differences in Syntactic Complexity in the Writing of EL1 and ELL Civil Engineering Students

Santiago Gustin
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [First and Second Language Acquisition Commons](#), and the [Syntax Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Gustin, Santiago, "Differences in Syntactic Complexity in the Writing of EL1 and ELL Civil Engineering Students" (2019). *Dissertations and Theses*. Paper 5129.

<https://doi.org/10.15760/etd.7008>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Differences in Syntactic Complexity in the Writing of
EL1 and ELL Civil Engineering Students

by
Santiago Gustin

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Arts
in
Teaching English to Speakers of Other Languages

Thesis Committee:
Susan Conrad, Chair
Lynn Santelmann
Linnea Spitzer

Portland State University
2019

© 2019 Santiago Gustin

Abstract

Traditional studies in syntactic complexity consider increased clausal complexity to be characteristic of development, proficiency and growth in written language production. However, this stereotypical view ignores two important facts. First, complexity differs by register (i.e. daily speech versus formal writing). Second, as the proficiency of writers increases, their complexity in formal writing changes from clausal complexity to phrasal complexity (i.e. lower-proficiency writers have more subordinate clauses whereas higher-proficiency writers tend to have more noun phrases). Therefore, in this study, I argue for the need to consider not just clausal complexity but also phrasal complexity measures when assessing development and performance in second language (L2) writing production. In addition, this study addresses two important gaps that remain understudied in the literature of syntactic complexity. First, there are few studies that analyze changes in syntactic complexity of first-language (L1) Spanish English Language Learners (ELL)'s writing. A few studies have analyzed writers' L1 background as an influential factor in complexity, but an important language such as Spanish has been ignored. Additionally, most studies focus on general academic writing (i.e. argumentative essays), but there are no studies that investigate syntactic complexity in other registers and English for Specific Purposes (ESP) areas. For instance, there are no studies in syntactic complexity that focus on civil engineering, which is an area where writing plays a vital role. Hence, this study intends to fill these gaps by looking at the syntactic complexity of civil engineering student writing, including Spanish L1 writers.

The present study investigated syntactic complexity in the writing of English-as-their-first-language (EL1) and English-language-learner (ELL) civil engineering student

writing. Taking a contrastive corpus-based approach, I used the L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010) to analyze measures of clausal and phrasal complexity. In particular, I used two measures of clausal complexity (clauses per sentence and dependent clauses per clause) and three measures of phrasal complexity (mean length of clause, coordinate phrases per clause, and complex nominals per clause). The analysis was focused on a total of 74 samples of student writing: 30 ELL low-level texts, 14 ELL high-level texts, and 30 EL1 texts. The quantitative analysis consisted of non-parametric statistical tests applied between groups (i.e. ELL-low vs ELL-high, ELL-low vs EL1, and ELL-high vs EL1).

The statistical analysis indicated that the writing of both ELL student groups was significantly more clausally complex than the writing of EL1 students on both clausal complexity measures. No differences were found in phrasal complexity, and no developmental trends were found in relation to levels of proficiency among writers. All groups exhibited high levels of internal diversity and lack of within-group consistency.

The pedagogical implications of this study include familiarizing ELL students with the characteristics of professional engineering writing as a way to break the stereotype that more clausally complex sentences entail more advanced and more proficient writing. ESP instructors should try to identify characteristics of the syntactic complexity particular to their field so that they can provide appropriate feedback to their students. Moreover, ESP programs with Spanish-speaking students should pay attention to clausal complexity as potential linguistic transfer from students' L1 into the writing production in the L2.

Acknowledgements

The completion of this thesis would have not been possible without the support and encouragement of important people in my life.

I want to thank my advisor Susan Conrad for introducing me to the world of corpus linguistics and for teaching me valuable skills in research and writing. Thank you for your patience and your advice. All your guidance and lessons are tools that will be very useful for my future professional career.

Thank you, Lynn and Linnea, for taking the time to read my work and for your feedback during the final stages of this process.

Thank you to all my friends, colleagues and students who made a contribution throughout this whole learning experience. Thanks for your helpful comments and ideas.

Thank you to all my family both in Colombia and Oregon for having my back when I most needed it. Thanks mom and dad for your encouraging words and for cheering me up when I felt stressed and frustrated. Thank you, Terri, Richard and the Sun family for opening your homes to me.

Byron – Thank you for your love and care and for always inspiring me to be a better person.

Thank you all so much!!

Table of Contents

Abstract.....	i
Acknowledgements.....	iii
List of Tables.....	vi
List of Figures	vii
Chapter 1: Introduction	1
1.1 Overview of the thesis	4
Chapter 2: Literature Review	6
2.1 The Importance of Syntactic Complexity in L2 Writing.....	6
2.2 Syntactic Complexity as an Indicator of L2 Proficiency	10
2.3 Syntactic Complexity in Civil Engineering Student Writing.....	13
2.4 ELT in Colombian Higher Education.....	17
2.5 The Present Study.....	18
2.6 Research Questions.....	19
Chapter 3: Methodology.....	20
3.1 Context of the study.....	20
3.1.1 Civil engineering in Portland State University.....	20
3.1.2 Civil engineering in Universidad Mariana.....	21
3.2 The Corpora.....	23
3.3 The L2 Syntactic Complexity Analyzer (L2SCA)	25

3.3.1. The measures for the present study.....	27
3.4 Analytical Procedures.....	32
Chapter 4: Results.....	34
4.1. Sentence Complexity Ratio	34
4.2 Dependent Clauses per Clause.....	37
4.3 Mean Length of Clause.....	39
4.4 Coordinate Phrases per Clause.....	41
4.5 Complex Nominals per Finite Clause.....	42
4.6 General Patterns of Syntactic Complexity in Civil Engineering Student Writing.....	44
Chapter 5: Discussion and Conclusions.....	48
5.1 Research Questions and Summary of Findings.....	48
5.2 Pedagogical Implications	53
5.4 Limitations and Future Research.....	57
5.5 Final Thoughts.....	61
References.....	62
Appendix: IRB Approval.....	70

List of Tables

Table 3.1 Summary of Corpora	25
Table 3.2 L2 Syntactic Complexity Measures in the L2SCA	26
Table 3.3 Measures Used in the Present Study.....	29
Table 4.1 Medians for Sentence Complexity Ratio.....	34
Table 4.2 Medians for Dependent Clauses per Clause.....	37
Table 4.3 Medians for Mean Length of Clause.....	39
Table 4.4 Medians for Coordinate Phrases per Clause.....	41
Table 4.5 Medians for Complex Nominals per Clause.....	42
Table 4.6. Differences in Complexity Values among the Groups.....	45
Table 5.1 Pairwise Comparisons for Clausal Complexity.....	49

List of Figures

Figure 3.1 Summary of Analytical Procedures.....	32
Figure 4.1 Box Plots for Sentence Complexity Ratio.....	35
Figure 4.2 Box Plots for Dependent Finite Clauses per Clause.....	38
Figure 4.3 Box Plots for Mean Length of Clause.....	40
Figure 4.4 Box Plots for Dependent Finite Clauses per Clause.....	41
Figure 4.5 Box Plots for Complex Nominals per Finite Clause.....	43
Figure 5.1 Trends in Measures of Clausal Complexity in All Groups.....	52
Figure 5.2 Trends in Measures of Phrasal Complexity in All Groups.....	53

Chapter 1: Introduction

As a former instructor for civil engineering in an English for Specific Purposes (ESP) program in Colombia, I used to have a clear set of criteria to evaluate my students' speaking skills. To determine how proficient their speaking was, I would assess their fluency, accuracy, comprehensibility, pronunciation and vocabulary. However, criteria were not as clear when it came to writing skills. I usually focused on general aspects of writing such as spelling and punctuation, word choice, organization, and grammar errors. I also believed that students' writing had to be syntactically complex in order to be considered proficient and advanced. This belief was merely based on intuition and the stereotype that the more complex your sentences are, the more sophisticated and advanced your language must be. Then, when I began studying for my M.A. TESOL degree, I learned how wrong my belief was.

Analyses of naturally occurring language through corpus-based research tell us that academic writing does not necessarily have to be syntactically complex to be considered advanced. In fact, research shows that syntactic complexity, traditionally defined as variety and diversity of grammatical structures, length of production units, and the degree of sophistication of linguistic resources (Wolfe Quintero, Inigaki & Kim, 1998; Ortega, 2003), varies between speech and writing. On the one hand, writing is characterized by being clausally simpler (Biber, Gray, & Poonpon, 2011; Parkinson & Musgrave, 2014). In other words, sentences in writing tend to have one clause per sentence, low levels of embedding and subordination, and more sophisticated phrases (e.g. noun phrases modified with pre or post-modifiers). On the other hand, speech usually carries high levels of dependent clauses and amounts of subordination. Therefore,

the same measures of complexity cannot be applied to describe different registers as complexity is multidimensional (Bulté & Housen, 2012).

Following the distinction of complexity in speech and writing, the study of syntactic complexity has become an important area of research in applied linguistics. Within the context of L2 writing research in particular, syntactic complexity plays a vital role because this construct can be used as an index of development, proficiency and performance in the target language. Thus, an objective and reliable method to measure and evaluate syntactic complexity in L2 writing has great implications for research, language teaching and language assessment.

Defining universal measures of syntactic complexity has been a difficult task in the field, especially because of the disagreements in the way the production units are defined (Lu, 2010; Bulté & Housen). For example, some authors only consider clauses as those structures with a subject and a finite verb (Hunt, 1965; Polio, 1997; Polio, 2017; Lu, 2010). Other authors define a clause as a phrase dominated by a verb phrase or subject (Bardovi-Harlig & Boffman, 1989). In addition, some studies prefer to study growth in complexity through length-based and subordination-based measures (e.g. length of sentence, clauses per T-unit) (Hunt, 1965), while other studies measure complexity through sophistication-based measures (i.e. complex nominals per clause, or passives per production unit) (see Wolfe-Quintero et al. 1998). Nowadays, with the advent of corpus-based research, and with the rising popularity of automatic analysis tools, such as the L2SCA (Lu, 2011) and CohMetrix (Crossley & McNamara, 2012), the construct of complexity has undergone a critical reexamination that has caused tensions

regarding the most appropriate measures of what it means for a text to be syntactically complex.

Overall, research shows that subordination-based measures are not the best candidates to predict development in L2 writing. In fact, academic writing tends to shift from clausal complexity (e.g. more than one clause per sentence and more high levels of subordinate clauses) to phrasal sophistication or phrasal complexity (e.g. complex noun phrases, coordinate phrases). These changes in the dimensions of complexity are susceptible to several factors such as levels of expertise or proficiency in the target language, task demands, modality and context of writing (Ortega, 2015). This characteristic of writing being clausally simpler than speaking, however, has generally been studied in general academic writing usually through the analysis of argumentative essays.

In addition, several studies have analyzed learner-related variables as potential factors that influence variation in syntactic complexity. One of those factors is learners' L1 background. For example, Lu & Ai (2015) studied the differences in syntactic complexity of academic essays written in English by learners with different L1s. Their study indicated that each L1 background group employed different levels of complexity even among groups who shared similar proficiency levels. This important finding suggests that changes in syntactic complexity in L2 writing are subject to writers' L1 background. However, the main language backgrounds that have been studied include Chinese, French, and German, but there is a gap in our understanding of other ELL's L1 backgrounds as is the case of Spanish, which seems to be understudied in the literature.

Moreover, another important gap that seems to be ignored is concerned with the register of writing. Most of the studies in syntactic complexity focus primarily on general academic writing (i.e. argumentative essays), but there are no studies that explore the syntactic complexity varieties in ESP contexts. For example, there are no studies that look at the syntactic complexity in civil engineering writing, which is a field where effective writing is important for the industry and therefore, worthy of systematic study.

Motivated by these gaps in the literature (i.e. scarce studies that analyze L1 Spanish ELLs, and scarce studies in civil engineering writing), I decided to investigate the writing of my former civil engineering students to better understand their grammatical choices in relation to their proficiency levels and their L1 background. Specifically, the purpose of this study was to determine if there are any significant differences between writers with different L1 backgrounds (English L1, Spanish L1), and if there are any developmental changes in civil engineering students' syntactic complexity as levels of proficiency in English increase.

1.1 Overview of the Thesis

Chapter 2 begins with a discussion of the importance of syntactic complexity, presents common measures to analyze this phenomenon and describes the uses and applications of the study of complexity in L2 writing research and pedagogy. The second chapter also reviews previous relevant literature to establish the need and context of my study and states the research questions. Chapter 3 presents the methodological design that I used in analyzing syntactic complexity. I also describe the context, corpus, instrument and analytical procedures. Chapter 4 reveals the results of this study in two parts: the first part presents the results for each of the syntactic complexity measures, and the second

part pulls together all the results to provide a general understanding of how all the measures interact with each other and within their specific context. Chapter 5 reviews the research questions and their answers, discusses the meaning of the findings, and concludes the thesis with its implications, limitations and suggestions for future research.

Chapter 2: Literature Review

This chapter reviews the relevant literature that establishes the context of the present study. I start by discussing why it is important to investigate syntactic complexity in L2 writing. I argue that there is a need to expand this field of research in areas that have been under-studied such as ESP contexts and non-essay types of writing. In the next section, I discuss the relationship between syntactic complexity and proficiency. Then, I examine the current state of affairs in civil engineering student writing. In this section I explain why it is important to consider writers' proficiency level in the target language as well as their L1 background, and how those variables influence their written performance. Finally, I contextualize the case of Colombian civil engineering students and their L2 writing in the discipline.

2.1 The Importance of Syntactic Complexity in L2 Writing

Over the past decades, research in second language acquisition (SLA) has paid considerable attention to syntactic complexity, a construct usually characterized as the degree, range, variety, and diversity of grammatical structures employed in language production and the level or amount of sophistication, elaboration or depth of those structures (Lu and Ai, 2015; Bulté & Housen, 2014; Ortega, 2003; Wolfe-Quintero et al., 1998).

Within the field of L2 writing in particular, syntactic complexity has been a major area of interest because it has long been recognized as an indicator of learner language development and proficiency (Yoon & Polio, 2017; Staples, Egbert, Biber & Gray, 2016; Biber, Gray & Staples, 2016; Mazgutova & Kormos, 2015; Crossley & McNamara, 2014; Ai & Lu, 2013; Lu, 2011; Beers & Nagy, 2009). At the same time, research in L2 writing

is interested in answering a variety of questions to understand the different forces that influence writers' use of linguistic resources in writing production. In a general sense, some of these questions include: How is syntactic complexity different in formal writing when compared to informal speaking? How does syntactic complexity vary across academic writing types, genres and styles? To what extent do levels of literacy and language proficiency distinguish different representations of syntactic complexity in writing?

Attempts to answer these questions have resulted in many applications of syntactic complexity in research and pedagogy. Among these applications are the development of fine-grained indexes to measure the construct itself (Bulté & Housen, 2014; Cooper, 1976; Hunt, 1965), the creation of corpus linguistics tools to improve investigation in L2 writing (Crossley & McNamara, 2014; Lu, 2010), instructional interventions to influence the development of grammar, writing ability or both (Polio, 2017; Mazgutova & Kormos, 2015; Parkinson et al., 2014); and the implementation of the construct as an indicator of writing quality through formal assessment and standardized tests such as the TOEFL or the IELTS (Lu, 2017).

Traditional studies that analyze syntactic complexity in L2 writing measure this linguistic property either through length-based measures or subordination-based measures (Liu and Li, 2016; Ortega, 2003; Wolfe-Quintero et al., 1998). The former measures typically include metrics such as mean length of sentence (MLS) or mean length of T-unit (MLT), whereas the latter refer to metrics such as dependent clauses per clause, or clauses per T-unit. As Liu and Li (2016) note, these two approaches to measure syntactic complexity were ubiquitous in past research because there was a common belief that

longer production units and higher amounts of subordination reflected higher levels of fluency in the language user. This belief was sustained under the assumption that subordination encompassed intricate argumentation and reasoning. In addition, subordination was presumed to be a predominant characteristic of formal writing because the task of writing involves time to prepare and revise (Biber, 1992). However, focusing exclusively on subordination and length-based metrics only reflects a misunderstanding of the nature of syntactic complexity (Norris & Ortega, 2009).

As popularity has risen in studies focused on syntactic complexity, so have the ways in which researchers analyze and measure this important construct. The urgency to reconsider the fact that complexity is such as multidimensional construct (Biber, 1992; Norris & Ortega, 2009; Bulté & Housen, 2014) has led scholars in the field to examine not only length and subordination measures but also, complexity at the phrase level along with nominal complexity and coordination (Polio, 2017). The possibility to expand research in different syntactic complexity features has been made possible, in part, thanks to the creation of popular corpus-linguistics tools such as the L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010) or the Coh-Metrix (Crossley & McNamara, 2014).

Studies that have explored non-clausal areas of complexity include those by Biber (1988), Biber & Gray (2011) and Biber, Gray & Ponpoon (2011). Generally, the previous studies have been able to identify significant differences of syntactic complexity between informal speaking and formal writing. For instance, Biber et al. (2011) found that subordination is a prominent characteristic of informal speaking but not common in academic writing. In fact, the written register has the characteristic of being phrasally complex, with a high density of noun phrases being pre- or post-modified. This

contradicts previous beliefs that subordination and clausal complexity were a prevalent aspect of formal writing.

Empirical evidence shows that complexity in academic writing shifts from clause embedding to noun phrase complexity. This change in writing takes place as writers develop along an academic path (see Lu & Ai, 2015; Lu, 2011; Biber, Gray & Poonpon, 2011). One of the reasons why academic writing tends to have more noun phrase complexity may be that writers are usually constrained by institutional requirements as is the case of journal publications which require authors to limit their contributions to a certain number of pages or words. Therefore, writers are obligated to limit their writing by packaging ideas into reduced expressions and more phrasal complexity. In addition, from the standpoint of systemic functional linguistics (SFL) (see Halliday & Martin, 1993) academic language seems to favor the use of nominalizations as grammatical metaphors to efficiently deliver relatively difficult concepts through reduced units of expressions. Grammatical metaphor, as Halliday et al. (1993) explain, should be understood as “a substitution of one grammatical class, or one grammatical structure, by another” (p. 87). For example, consider *a* and *b* below:

a. How quickly cracks in glass grow

b. Glass crack growth rate

In these examples (taken from Halliday et al., 1993), we see how the noun phrase in *b* expresses the same meaning in *a* by restructuring all the parts of the sentence. The verb *grow* in *a* for instance, has been replaced by the noun *growth* in *b*, and the phrase *how quickly* has been substituted by the noun *rate*. The result is then, a clause that has been reduced to a noun package that fulfils the purpose of expressing concise information

for more efficient communication for experts in the field. The ability to use this type of discourse in the writing domain appears to be “a question of maturity” (Halliday, 1993). When writers develop higher levels of literacy, they acquire the ability to substitute structures such as subordinate clauses with noun phrases to express complex ideas faster and more effectively.

Other factors associated with the way writers choose a particular linguistic resource (clause-level or phrase-level structures) include the communicative demands of a certain task (Ryshina-Pankova, 2015). It seems that the particular task requirements have a great impact in the way writing narratives are constructed. For example, Conrad (2017) and Conrad & Pfeiffer (2011), found that practicing engineers tend to use a high density of noun groups in their workplace writing when describing locations, amounts and objects. These descriptions were phrasally complex because writers needed to provide precise and unambiguous information required for successful completion of engineering projects.

2.2 Syntactic Complexity as an Indicator of L2 Proficiency

Research in SLA has traditionally recognized the triad of accuracy, fluency and complexity as the three best predictors of proficiency in language production (Wolfe-Quintero et al., 1998). Note that proficiency refers to the skills and competence that a language user may have at a certain point in time. It should not be confused with development, which refers to observable changes over time, normally examined in longitudinal studies (Polio, 2017).

Several attempts to explain how syntactic complexity can gauge levels of proficiency have been made over time (Liu, 2017; Ai & Lu, 2013; Kuiken & Vedder,

2012; Lu, 2010; Ortega, 2003). Research shows that as writers become more proficient, they modify their syntactic repertoire by moving from the use of dependent clauses to non-finite dependent clauses to eventually use dependent phrases with a high density of noun phrase modifiers (Ai & Lu, 2013; Wolfe-Quintero et al., 1998). Parkinson and Musgrave (2014) note that the use of clauses per sentence, and dependent clauses per clause and per T-unit decrease when users reach a higher level of literacy. These syntactic changes in academic writing have been confirmed both in L1 learners and L2 learners (see Biber et al. 2011). As Polio (2017) suggests, these changes of syntactic complexity at different proficiency levels are highly associated with changes over time due to instruction in the target language, context and purpose for writing.

Studies that have examined syntactic complexity taking into account learners' proficiency levels are common in literature although mainly focused in general academic writing (i.e. argumentative essays written by university students) (Ai & Lu, 2015; Lu, 2011; Lu, 2017; Lu, 2010; Lu, 2009; Lu & Ai, 2015; Kuiken & Vedder, 2012; Crossley & McNamara, 2012). For instance, Lu (2011) found that complex nominals per clause were one of the best predictors of written complexity. Lu investigated 14 measures of syntactic complexity in a large-scale corpus of L2 student writing from four different college-level groups of English Language Learners (ELL). His findings documented seven measures (i.e. mean length of clause, sentence, T-unit, coordinate phrases per clause and per T-unit, complex nominals per clause and per T-unit) to be highly correlated with proficiency and school level. Among the above measures, complex nominals per clause and mean clause length increased linearly and differentiated significantly among the four school levels that he analyzed. These findings, as mentioned

earlier, confirm the fact that use of phrasal complexity develops along with proficiency level and school level.

In a similar study, Ai & Lu (2015) examined the differences in syntactic complexity in the writing of native speakers (NS) and non-native speakers (NNS) of English in an EFL context. Their study included 10 measures of syntactic complexity taken from the L2SCA (Lu, 2011). The results indicated significant differences in 8 of the 10 measures analyzed. A more interesting finding was that the mean values for each measure increased in a linear fashion from NNS-low, through NNS-high to NS with the NS group using more phrasal complexity than the two NNS groups.

Parkinson & Musgrave (2014) examined the writing of two groups of EAP students from different proficiency levels at a higher-education context. Their study was interested in determining the different ways in which syntactic modification at the noun phrase level occurred. This study looked at whether noun phrases were made complex by using pre-modifiers (e.g. attributive adjectives) or post-modifiers (prepositional phrases). It was found that the lower-proficiency group used more attributive adjectives to complexify noun phrases, whereas the higher-proficiency group relied more heavily on nouns as pre-modifiers and prepositional phrases as post-modifiers. These findings suggest that modification in phrasal complexity also varies and develops linearly as proficiency levels increase.

Even though researchers in corpus linguistics have recognized that phrasal complexity is a distinctive characteristic of formal writing, and that variations in syntactic complexity occur across proficiency levels, most researchers have focused their studies on general academic writing (Lu, 2017; Lu & Ai, 2015; Mazgutova & Kormos, 2015;

Crossley & McNamara, 2014; Ai & Lu, 2013; Lu, 2011. Discipline-specific areas tend to be under-studied.

Additionally, another gap often found in the literature is concerned with the kind of texts that are analyzed. It appears that most researchers, as the ones mentioned above, have a preference for essay types of writing. Literature on syntactic complexity that focuses on more discipline-specific writing is scarce. In an attempt to fill these gaps, I extend this line of research by examining the case of civil engineering student writing. Recent studies in this particular area have pointed out the need for exclusive inquiry in syntactic complexity.

2.3 Syntactic Complexity in Civil Engineering Student Writing

Recent studies in civil engineering writing recognize the importance of developing professional writing skills to match practices in the industry (Conrad & Pfeiffer, 2011). Within the context of civil engineering, effective writing skills play a vital role in the execution of projects, in maintaining client satisfaction, and perhaps most importantly, in avoiding any potential liability (Conrad & Pfeiffer, 2011). Accordingly, in their training for the future profession, civil engineering students are constantly completing a variety of class assignments that require some kind of writing production. Apart from typical academic essays on an engineering topic, civil engineering student writing includes very specific writing genres: laboratory reports, technical memoranda, project-related emails, site visit reports, plan sheet notes, among others (see cewriting.org for more information on a project about writing in civil engineering) (Conrad, 2017).

Literature in linguistic analysis of civil engineering writing is scarce, but recent studies by Conrad (2017, 2018; Conrad et al., 2011) have examined several aspects of

writing such as sentence structure, genre organization, word choice, grammar and punctuation errors and writer's motivations. While acknowledging the fact that each of the above aspects are relevant to understand civil engineering writing, the present study is only concerned with syntactic or grammatical features to extend the scope of syntactic complexity in other areas of ESP.

Conrad's (2017) study, although not exclusively focused on complexity, looked at the sentence structure of civil engineering students and practitioners. The study found that overall, students use far more complex sentences than practitioners. However, the results in this study cannot be generalized to all levels of complexity given that it only focused on clausal complexity using subordination-based measures. The study defined complex sentences as "dependent clause structures" and non-complex sentences were operationalized as sentences that did not have "dependent structures" (p. 198). Even though Conrad did not use fine-grained measures of all types of syntactic complexity, the results indicated that students' writing had more grammatical complexity at the clause level in comparison to practitioners. Students had more embedded clauses and expressed more than one idea per sentence. In contrast, practitioners used less embedding in their sentences, and the syntactic complexity of their writing took place at the noun phrase level. These findings are consistent with a preliminary study by Conrad et al (2011) which also found that practitioners built their complexity at the phrase level by using prepositional phrases, and a "simpler sentence structure" (p. 5). Consider the following examples taken from Conrad's (2017) study.

a. simple sentence - no embedding (practitioners): *the existing bridge is a 9-span timber trestle bridge with a concrete deck.*

*b. complex sentence - embedding (students): departures tended to have less pronounced localized peaks than arrivals, **suggesting that departures are slightly less dependent on class time, as well as may account for the varying duration of class times.***

The practitioner's sentence in *a* shows complexity at the phrase level as it can be seen in the underlined section. The noun phrase *bridge* is made complex by adding the pre-modifying phrase *9-span timber trestle* and post-modified by the prepositional phrase *with a concrete deck*. In general, practitioners' writing did not exhibit subordination, which is consistent with previous claims that advanced writers do not rely on clause-complexity (Lu, 2011). A closer look at syntactic complexity in civil engineering writing using fine-grained measures would probably provide more information about the diversity and range of syntactic structures used by writers. It can be seen in *b* for example, that a student's sentence employs both subordination (in bold) and phrasal complexity (underlined).

In an attempt to extend previous research in civil engineering writing, Conrad (2019) conducted a multidimensional analysis of student, practitioner and academic engineering writing. She primarily examined the use of passives and impersonal style features, but the study also showed evidence of the variation that exists in writers' syntactic complexity. The most relevant results for the present study concern the fact that practitioners and academic engineers used simpler sentence structures than student writing, and the two more proficient groups exhibited higher frequencies of reduced clauses. These results also agree with previous findings in L2 writing research that has established the stages through which writers move as they progress in their proficiency

levels (i.e. from long sentences to subordination to reduction of clauses to phrases).

Earlier research has suggested that expert writers reach a point where they change their discourse to have more phrasal complexity and reduced clauses due to the need to be precise and unambiguous. These findings are also consistent with academic writing research that has shown that phrasal complexity is more common with advanced writers.

In short, these recent studies in civil engineering writing demonstrate the importance of analyzing syntactic complexity in different proficiency levels, under the lens of fine-grained measures and considering the specific communicative purpose and context in which writing is used. In the case of civil engineering, for example, it appears that overall, phrasal complexity is more valued by practicing engineers. Short and simple sentences seem to be more effective because they avoid vague and imprecise meaning, facilitate reading in clients and among colleagues, and limit the firm's potential liability. Students, however, may not be aware of the importance of developing effective writing skills for future professional practice, and thus, this creates pedagogical implications in engineering education.

Even though Conrad's studies provide a general idea of how syntactic complexity appears in engineering writing, an analysis through more fine-grained measures may provide additional information into the way syntactic complexity is expressed in this particular discipline. Additionally, there is a need to understand if civil engineering student writing differs when looking at writers' individual differences. For example, Conrad's studies did not distinguish writing based on writers' L1 background. This variable was not controlled for given that most of the texts analyzed were composed by groups, not individual writers.

Linguistic studies in civil engineering that consider students' L1 background are scarce or limit their analysis to other features unrelated to the construct focus of this study (Nausa, 2017; Nanwani, 2009; Luzon, 2009). Additionally, studies that have investigated differences in syntactic complexity among writers with different L1 backgrounds also seem to have ignored L1 Spanish speakers (Lu & Ai, 2015, Ai & Lu, 2013).

To begin filling this gap, the present paper proposes to examine Spanish-speaking students of civil engineering within the context of Colombian universities. The following section will contextualize ELT education in Colombia.

2.4 ELT in Colombian Higher Education

In Colombia, learning English as a foreign language is so important that, for more than a decade, the Ministry of National Education has institutionalized a bilingualism plan (i.e. Spanish and English) in all levels of formal education (Ministerio de Educación Nacional [MEN], 2009). The Colombian government's interest in bilingualism is based on the need to increase the insertion of the national human capital into the global economy (Gómez Sará, 2017).

With such an ambitious goal in mind, the MEN has established parameters for EFL education which are mainly based on the Common European Framework of Reference (CEFR) (Obando Guerrero & Sánchez Solarte, 2018). In turn, as a way to assess the progress of the bilingualism plan established by the government, the MEN has standardized national tests which are obligatory for graduation in any level of education. In this sense, with the need to raise institutional scores in the foreign language section of the tests, universities in particular put a great emphasis in language teaching, because

better evaluation results mean access for more institutional accreditation and better national rankings. Consequently, many universities across the country seek to improve language testing scores by creating obligatory language courses which sometimes include general language learning courses, bilingual (Spanish and English) discipline-specific courses, or even discipline-specific courses where English is the medium of instruction.

The pressure put on students to become proficient language users creates a big challenge for both teachers and learners (Obando Guerrero et al, 2018; Nanwani, 2009). In particular, use of English as second language becomes problematic when students are assigned discipline-specific writing tasks. Academic writing already poses a challenge for students regardless of their L1 background, and it becomes even more difficult when they have to write in a second language (Nanwani, 2009). Thus, a linguistic analysis of the way Spanish-speaking students write in the English may be useful in understanding the pedagogical needs to help students develop this important skill.

2.5 The Present Study

As shown so far, researchers in L2 writing have made great efforts in explaining the way syntactic complexity varies in relation to personal factors such as writer's levels of proficiency, levels of literacy, L1 backgrounds, task demands, and purpose and context of writing production. Many answers have been provided, and important findings have contributed to our global understanding of how writers use their linguistic resources. However, there are several gaps that seem to be unresolved. Mainly, most of the research focuses on general academic writing (i.e. argumentative essays). There is little knowledge about the way syntactic complexity is represented in other areas where English is used for specific purposes, as is the case of civil engineering. Additionally, even though

researchers have established that learners' variables such as proficiency levels and L1 background may influence syntactic structures, important languages such as Spanish have been under-studied.

Therefore, in light of these observations, the purpose of the present study is to determine the differences of syntactic complexity in the writing of civil engineering students: EL1 and Spanish-speaking ELL. In particular, the present study intends to answer the following research questions.

2.6 Research Questions

1. Is there a statistically significant difference in the features of syntactic complexity in engineering papers written by ELL lower-level, ELL higher-level, and EL1 students?
2. Is there overall development from clausal to phrasal complexity as levels of language proficiency increase from ELL-lower-level to EL1?

Chapter 3: Methodology

This chapter describes the methodology used for this study to analyze syntactic complexity in L2 writing. The following sections describe the contexts of the programs from which I gathered student writing, the corpora used in the study, the instrument for the automatic analysis of complexity features, the complexity measures employed, and the analytical procedures.

3.1 Context of the study

The present study, conducted at Portland State University (PSU), in Portland, Oregon, compared the writing produced by English L1 (EL1s) students enrolled in a civil engineering program at PSU, and English Language Learners (ELLs) in a civil engineering undergraduate program at a university in southwestern Colombia. In the following subsections, I describe the universities and their civil engineering programs, and the language learning program in the Colombian university.

3.1.1 Civil engineering in Portland State University. Portland State University (PSU) is a public research university located in the metropolitan area of Portland, Oregon. The university is 73 years old, has over 27,000 students and offers undergraduate, graduate and postgraduate programs. The university occupies the 230th position in the National Universities list (out of 301 institutions) in the Best Colleges U.S. News Rankings (2019).

The undergraduate civil engineering program at PSU is part of the Department of Civil and Environmental Engineering, and it offers a Bachelor's of Science degree in civil engineering with a 4-year curriculum. Graduates have an 85% pass rate on the Fundamentals of Engineering Exam, (FE), a national exam taken at the end of the

Bachelor's degree and the first step to become professional licensed engineer in the USA (NCEES, n.d.). This percentage of pass rate on the FE exam is considerably above the national average.

In the civil engineering program, all students are required to take general university courses that include courses in written and spoken communication skills, but they are not required to take a technical writing course in order to graduate (PSU Bulletin, 2017-2018). Civil engineering classes at PSU may include homework and projects with significant use of writing skills. Common assignments that require writing include laboratory reports, technical memos, proposals, and site descriptions, among others. The department has participated in the Civil Engineering Writing Project (Conrad, 2017), a research project that included the creation of instructional materials that help students develop effective writing skills for practice in the industry.

3.1.2 Civil engineering in Universidad Mariana. Universidad Mariana (UM) is a private catholic university located in the southwest region of Colombia. The university is 50 years old, has over 7000 students and offers undergraduate, graduate and postgraduate programs. The university is ranked 44th out of 287 national universities (Webometrics, 2019). Universidad Mariana was founded in 1970, and it offers undergraduate, graduate and postgraduate degrees.

The undergraduate civil engineering program is part of the School of Engineering and offers a Bachelor's of Science degree with a 4-year curriculum. Graduates from engineering programs in Colombia need to take the national higher education tests established by the Ministry of National Education (ICFES, n.d.). These exams are

required for graduation, but there are no national exams exclusively to get a professional engineering license in Colombia.

The School of Engineering at UM has its own program for teaching communication skills in English as a foreign language. Each semester, students take a general course with four sub-components: grammar and vocabulary, listening and speaking, presentation skills, and engineering writing. Each component takes two hours of instruction (i.e. 8 hours of language learning per week). Apart from this, each semester, students take one or two courses from their engineering program taught half in Spanish and half in English.

The communication program includes 5 courses. Four of the courses are based on the Common European Framework of Reference (CEFR): A1, A2, B1, B2. Students who pass the B2 level take a fifth course, International Test Training (ITT). A1 level is considered beginner. A2 is considered elementary. B1 is considered intermediate, and B2 is considered upper-intermediate (Council of Europe, 2001). To determine which of the above courses to take, students must take an institutional placement test that is equivalent to the CEFR standards. All students need a B2 level to fulfill their English language requirement, and a B2 level in an international test is the minimum to meet the graduation requirements.

As part of the requirements in the civil engineering program, students need to develop a project known as Proyecto Integrador or PI Project every semester during their engineering training. The goal of this project is to give students the opportunity to integrate several courses of each semester into a conceptual or investigative project. Depending on the semester, students can choose to do a variety of engineering projects

such as structural analysis reports, laboratory analyses, software implementation, or project proposals. A requirement of this project is to write 50% of its content in Spanish and 50% in English. This condition intends to give students practice with writing in the field both nationally and internationally.

3.2 The Corpora

All texts that make part of the corpora were written by undergraduate engineering students. The corpus included three subcorpora: the EL1 group, the ELL higher-level group, and the ELL lower-level group. As I compiled the subcorpora, I selected papers so that the genre, topics, and writing conditions (such as final drafts) were as consistent as possible across the groups, since those factors could affect syntactic complexity.

Both groups of ELL texts were taken from the Proyecto Integrador (PI) archive at the Civil Engineering Department at Universidad Mariana. The archive has copies of all the submissions of students' projects, including three versions of each document (i.e. first draft, second draft, final graded submission). I used only final drafts that received a passing grade. I used A1 and A2 level papers for my ELL low-level group, and B1 and B2 for my high-level group. All of these projects were of the same genre: proposals. Topics varied slightly depending on the level, as described further below.

As English is the first language of students in the EL1 group, I considered this group to be in the higher levels of proficiency based on CEFR standards. In other words, this group is in C1 and C2 levels (i.e. C1 means advanced, and C2 means proficient). Texts for this group came from a corpus that already existed as part of the Civil Engineering Writing Project (CEWP) (see Conrad, 2017). This corpus includes ten genres and the writing of students from several universities and practitioners in the USA.

For my study, I selected student papers that were written at PSU and were the same genre as the UM papers (proposals). I included only papers that had a passing grade.

The subcorpora were well controlled for the genre. Project proposals from both universities are characterized by having an introduction to the problem, a purpose statement, a series of specific objectives, and a description of the work plan with phases and procedures to complete the proposed project. In addition, all texts at both universities were written in groups. This made the subcorpora similar, but individual differences such as age or gender could not be investigated.

All proposals in all groups addressed infrastructure projects, but the topics varied slightly. ELL lower-level texts are either about construction of a bridge or building. ELL higher level papers, on the other hand, are about hydraulic structures. The EL1 group wrote about building urban transportation systems, such as highways and bridges. I considered this variation of topics in my interpretation of the results.

All texts in both corpora were collected after receiving approval from the Institutional Review Board at Portland State University. Permission slips were obtained from the Colombian university to access the archive where the students' texts are saved. My target was to analyze 30 texts from each group. However, only 14 higher-level texts were available in the UM archive. Consequently, I analyzed 30 lower-level ELL students (A1, A2), 14 higher-level ELL students (B1, B2) and 30 EL1 texts (C1, C2). Table 3.1 summarizes the data used for the analysis.

Table 3.1 Summary of Corpora

	ELL Lower Level	ELL Higher Level	EL1 group
Number of texts in corpus	30	14	30
Total words in each corpus	19338	18999	36748
Average words per text	645	1357	1225
Genre	Project Proposals	Project Proposals	Project Proposals
Topic	Building a structure (e.g. bridge, building, roadway)	Hydraulic structures	Urban transportation systems

Original texts in all three groups contained between 100 and 15,000 words. Given that the analytical tool I used only processes a maximum of 2,000 words per text, I took the first 2,000 words of each text that had more than 2,000 words. A weakness of this design is that the syntactic complexity might vary across sections, and that variation is not included in the scope of this study.

3.3 The L2 Syntactic Complexity Analyzer (L2SCA)

I used a corpus linguistics tool for the analysis of syntactic complexity features known as the L2 Syntactic Complexity Analyzer (L2SCA). This tool, developed by Lu (2010), automatically counts syntactic features and retrieves frequencies of 14 syntactic complexity measures distributed among five main categories: (1) length of production units, (2) amount of coordination, (3) amount of subordination, (4) degree of phrasal sophistication, and (5) overall sentence complexity (see Table 3.2).

The L2SCA completes the analytical process in the following stages. First, it uses the Stanford parser (Klein & Manning, 2003) to identify the syntactic structures of the sentence in the samples, which are tokenized and tagged for parts of speech (POS) with the same parser. The result of this first pre-processing stage is a series of parse trees. In the next step, the L2SCA calls Tregex (Levy & Andre, 2006) for identification and counting of the syntactic structures and production units based on Tregex patterns. The

final outputs are frequencies of nine production units (i.e. words, sentences, verb phrases, clauses, T-units, dependent clauses, complex T-units, coordinate phrases, and complex nominals) and the fourteen complexity measures seen in table 3.2.

Table 3.2 L2 Syntactic Complexity Measures in the L2SCA

Type	Measure	Definition
Overall sentence complexity	1. Sentence complexity ratio	# of clauses / # of sentences
Length of production unit	2. Mean length of clause	# of words / # of clauses
	3. Mean length of sentence	# of words / # of sentences
	4. Mean length of T-unit	# of words / # of T-units
Amount of subordination	5. T-unit complexity ratio	# of clauses / # of T-units
	6. Complex T-unit ratio	# of complex T-units / # of clauses
	7. Dependent clauses ratio	# of dependent clauses / # of clauses
	8. Dependent clauses per T-unit	# of dependent clauses / # of T-units
Amount of coordination	9. Coordinate phrases per clause	# of coordinate phrases / # of clauses
	10. Coordinate phrases per T-unit	# of coordinate phrases / # of T-units
	11. Sentence coordination ratio	# of T-units / # of sentences
Degree of phrasal sophistication	12. Complex nominals per clause	# of complex nominals / # of clauses
	13. Complex nominals per T-unit	# of complex nominals / # of T-units
	14. Verb phrases per T-unit	# of verb phrases / # of T-units

According to Lu (2010), the system achieved high levels of reliability given that the identification of the structures by human annotators was very similar to the identification made by the system. In addition, the error analysis reported by Lu (2010) indicates that errors learners make in their writing (e.g. issues with collocations, errors with determiners or agreement) do not cause problems in parsing or identifying the units of production and syntactic structures being analyzed. However, punctuation errors need to be considered given that the parser identifies sentences when they are delimited by a punctuation mark that signals the end of the sentence (i.e. period, question mark, exclamation mark, quotation mark or ellipsis)

The system can be used to analyze a single text file or multiple texts at a time. For the analysis, texts need to be formatted to plain text. The outputs are CSV files that can be imported into Excel or statistical software such as SPSS. This tool was chosen for this study because of its free availability, user-friendly design, and automatic analysis of syntactic complexity measures. However, there are certain considerations that need to be addressed when using the tool for research, some of which are discussed below.

3.3.1. The measures for the present study. The fourteen indices of syntactic complexity in Lu's analyzer are the result of a compilation of common measures used by several scholars in previous literature. The L2SCA, then, was created to give researchers the option to choose the measures that best fit the purpose of their particular study.

It is important to note that some of the measures in the above table are redundant because they measure the same things. This redundancy has been reported by Lu and Ai (2015), who used all 14 measures in their study and found that three subordination measures: complex T-units per T-unit, dependent clauses per clause, and dependent clauses per T-unit exhibited identical patterns across all the groups of writers they analyzed. For this reason, I narrowed measures to answer my research questions without redundancy.

In selecting measures, I first omitted the T-unit measures because of numerous discussions of T-units in the literature that discuss limitations of this type of measure. For instance, the T-unit has been criticized for its construct validity, as it is defined differently across multiple studies (Wolfe-Quintero et al., 1998). For example, Hunt (1965) defines a T-unit as a main clause plus any subordinate clauses. Bardovi-Harlig & Bofman (1989 in Wolfe-Quintero et al., 1998) expands that definition to include segment

fragments that are punctuated as sentences by the writer, but Ishikawa (1995 in Wolfe-Quintero et al., 1998) specifies that a T-unit does not include sentence fragments. Also, the T-unit appears to be unreliable when measuring texts produced by low-level students because they produce more errors, fragments and under-punctuated sentences, thus segmentation of the T-unit becomes subjective (Gaies, 1980). Finally, I avoided this type of measures because as mentioned before, the T-unit measures are redundant with other measures in the analyzer. For example, mean length of sentence and mean length of T-unit are both a measure of number of words, and in both cases, it is not possible to determine whether they show clausal or phrasal complexity as the units can be made longer by either phrases or clauses, and this distinction is not accounted for with either measure.

Since the purpose of this study is to understand how much variation there is in terms of clausal and phrasal complexity, I also selected the measures that provided information in these areas of complexity. I selected five measures from the L2SCA that best capture phrasal and clausal complexity (Table 3.3). One limitation of all the measures is that in the L2SCA a clause is defined only as a structure with a subject plus a finite verb. Non-finite clauses are counted as verb phrases. They can be part of phrasal complexity if they modify nouns; for instance, the sentence *the bridge built over 200 years ago will be replaced* has a complex noun phrase (underlined) However, adverbial non-finite clauses count as neither clausal nor phrasal complexity (e.g. *in order to complete this project*, or *although completed on time*). It is possible to measure non-finite structures manually (see Yang, Lu & Weigle, 2014), but for the scope of this study, I chose automatic measures from the L2SCA that are effective in determining general

differences between clausal and phrasal complexity. Following the descriptions in Lu’s analyzer, the following operational definitions should be taken into account when interpreting the results of this study.

Table 3.3 Measures Used in the Present Study

Type	Measure	Definition
Clausal complexity	1. Sentence complexity ratio	# of finite clauses / # of sentences
	2. Dependent finite clauses ratio	# of dependent finite clauses / # of finite clauses
Phrasal complexity	3. Mean length of clause	# of words / # of finite clauses
	4. Coordinate phrases per clause	# of coordinate phrases / # of finite clauses
	5. Complex nominals per clause	# of complex nominals / # of finite clauses

Clausal complexity. A sentence is complex if it has at least one main clause and at least one finite dependent clause. For example, *although the couplet currently operates with synchronized signal timing, the timing is coordinated for 25 mph.* This sentence is clausally complex because it has a main clause (in italics) and a subordinate or dependent finite clause (underlined). In contrast, the following sentence is not considered complex because it only has one finite clause: (...) *the consultant (PSU) will manage all aspects of the project defined above, including scheduling and task coordination.* The underlined section, even though it is a non-finite clause, does not make the sentence clausally complex according to the above operational definition.

Phrasal complexity. A sentence will be phrasally complex if within a finite clause there is at least a coordinate phrase (i.e. adjective, adverb, noun or verb phrase), and/or at least a complex nominal (i.e. noun plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, nominal clause, and gerunds and infinitives in subject position). For example: *the physical properties of the site, including sight lines,*

elevations and layout will be observed directly. Notice that this sentence is phrasally complex because it has a long coordinated noun phrase (underlined).

To determine the variation in clausal complexity, the two measures, sentence complexity ratio and dependent finite clauses per clause, will tell me both the number of finite clauses per sentence and the amount of subordination in a sentence. These two measures are useful because they can provide different perspectives on clausal complexity. Consider the following examples:

Sentence Complexity Ratio	Dependent Finite Clause Ratio	Example
1:1	0:1	<i>The project will replace the existing bridge.</i>
2:1	1:1	<i>The project, which was proposed by the New Horizons firm, will replace the existing bridge.</i>
2:1	0:2	<i>The project will replace the existing bridge and it will improve the fish passage greatly.</i>

As seen in the examples above, then, the two measures for clausal complexity are useful because they show different aspects of clausal complexity. In addition, these two measures avoid redundancy because they show different values and they tell different characteristics of complexity. The first one tells us the number of finite clauses in a sentence, and the second one provides information about subordination.

To investigate phrasal complexity, the measures I chose are mean length of clause, coordinate phrases per clause, and complex nominals per clause. The first measure, mean length of clause, which appears in the category of length of production unit in the L2SCA is clearly a measure of phrasal complexity. Consider the following examples.

1. *The bridge will be replaced.* (Mean length of clause: 5 words/1 finite clause)
2. *The bridge built over 200 years ago will be replaced* (Mean length of clause: 10 words/1 finite clause)

The previous examples contrast in that the second sentence is more phrasally complex as the noun phrase *the bridge* is post-modified by the verb phrase *built over 200 years ago*. Notice that the underlined structure, although traditionally considered a non-finite clause, is counted as a verb phrase following the operational definitions mentioned earlier. Consequently, by adding the post-modifier, the word count increases and thus makes the clause longer. Therefore, an increase in the value of number of words per clause should be interpreted as an index of phrasal complexity. It is important to consider also, that an increase in length does not translate to clausal complexity. For instance, a sentence may contain a relative clause (i.e. a finite clause) embedded in a complex nominal phrase as in the example: *The objective of this study is to provide **signal timing recommendations for N. Williams Street [that are more efficient for cyclists without significantly reducing level-of-service for vehicular traffic]**.* The complex nominal is the line in bold, and the relative clause in brackets modifies the head noun *recommendations*.

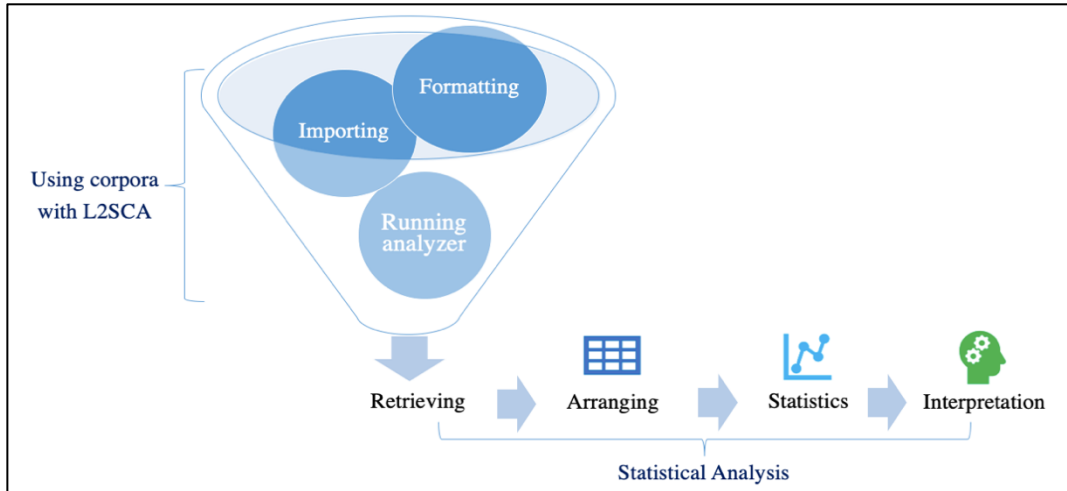
In respect to the other two measures of phrasal complexity, the measure coordinate phrases per clause will provide information about the amount of coordination within finite dependent clauses. A higher value in this measure will suggest a higher level of phrasal complexity. The final measure will indicate the number of complex nominals present in each finite clause as in the highlighted section of the following example: *The engineers analyzed **the structure built in 1975**.* In this example, the noun head *structure* is made complex because it is post-modified by a non-finite clause (underlined).

3.4 Analytical Procedures

Figure 3.1 summarizes the main steps taken during the analytical procedures.

Each step is described in sequence below.

Figure 3.1 Summary of Analytical Procedures



The first step was formatting the original texts from the ELL corpus. I deleted irrelevant information for the analysis such as images, figures, references and mathematical expressions. I took plain text from pdf versions and edited sections that had the wrong punctuation. In some texts, students wrote a comma at the end of a sentence, and so I modified this kind of case to make sure that the results from the L2SCA were accurate. Texts from the PSU corpus were already formatted.

The next step was importing the texts in the L2SCA and running the measures. After obtaining the values for each measure, I imported all the values generated by the L2SCA into SPSS for statistical analysis.

First, I employed box plots to describe the distribution of values for all variables in all groups. The use of boxplots allowed me to have a better visual understanding of the distribution of the data, its central tendency and its variability. I also decided to use

boxplots because I wanted to display the distribution of the three groups at the same time for each measure. This allowed for comparison of data from each group and made it easier to determine the statistical tests to perform. In addition, given that the sample sizes are small, I assumed there would be non-normal distribution, and I could confirm this with the box plots.

Because there was non-normal distribution of the data, I performed a non-parametric test, the Kruskal Wallis one-way analysis of variance test, for each of the measures. This test is the equivalent of an Analysis of Variance. Because I performed five comparisons, I used a Bonferroni correction factor. That is, to have an alpha for the study that was set to .05, I had to divide by 5, so I set alpha for each comparison at .01. Pair-wise comparisons were used to determine which groups were significantly different from the other groups.

After completing the statistical analysis, I compared the results for each variable and established whether there was a developmental pattern across the three groups. I looked at whether the results increased or decreased in a linear fashion from lower-level to higher level groups or vice-versa. I also analyzed the significantly different features in the context where they were used in the texts.

Finally, in the interpretation of my results, I compared my findings with the results obtained in earlier literature that analyzed differences between native versus non-native speakers' writing. I also compared my results with the ones obtained in past research in civil engineering writing.

Chapter 4: Results

The first part of this chapter addresses the results for each of the complexity measures included in my study. For each measure, I first review descriptive statistics, telling the central tendencies and describing the distributions with box plots. Then I report the statistical results and compare groups that were significantly different as well as those that did not exhibit differences. For each measure and comparison, I provide illustrative examples. The second part summarizes the findings of all the analyzed features across the three groups.

4.1. Sentence Complexity Ratio

Table 4.1 displays the median values for each group for sentence complexity ratio. From this table, it is obvious that there is little difference in the central tendencies for the two ELL groups. However, the box plots showed a great deal of variability in these groups (Figure 4.1).

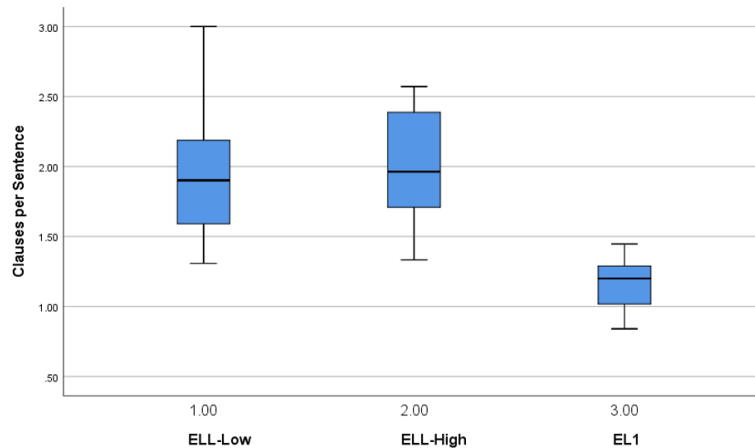
Table 4.1 Medians for Sentence Complexity Ratio

Group	Median
ELL-Low	1.90
ELL-High	1.96
EL1	1.20

Box plots show a five-number summary: minimum, first quartile, median, third quartile, and maximum. As seen in Figure 4.1, each plot shows a box in the middle which represents the interquartile range (i.e. the middle 50% percent of the data), a line in the middle of the box that represents the median, and “whiskers” which are lines on each side of the box that represent the ranges for the bottom 25% and the top 25% of the data values. The length of the vertical lines represents how much variation or how spread out the data are between the maximum and minimum values in comparison to the median.

When a case is more than three times the height of the box, it is considered an outlier, and it will appear as a dot in the box plot.

Figure 4.1 Box Plots for Sentence Complexity Ratio



With respect to sentence complexity ratio (i.e. finite clauses per sentence) specifically, Figure 4.1 indicates that the three groups had different levels of variability. On average, group EL1 had fewer finite clauses per sentence in comparison to the two ELL groups. Group ELL-low had a wider range with skewness to the upper side. In that group, there were texts that had up to three clauses per sentence. Both ELL groups have a similar median value, so there was not much difference between these two groups. However, the higher-level group showed less variation and more consistency than the lower-level group. This consistency can be seen in the length of the whiskers for the ELL-high plot which are shorter than those in the ELL-low plot.

To determine whether there was a statistically significant difference in clauses per sentence among the three groups, I used the Kruskal-Wallis test, which showed a significant difference of medians ($H = 49.824$, adjusted $p < .05$). I then conducted post hoc tests to test pairwise comparisons. There were significant differences between the ELL-low group and the EL1 group ($p < .05$) and between the ELL-high group and the EL1

group ($p < .05$). There was no significant difference between the ELL low and high groups. In other words, both the ELL groups had significantly more clauses per sentence in comparison to group EL1.

The difference in the ELL-low group's complexity is exemplified in the following sample, where numerous finite clauses are in the same sentence (finite clauses are delimited by brackets, highlighted and numbered for easier identification):

[1]Today there are many areas of difficult access [2]where the roads [3]that connect one place with another [4]have a long journey [5]which has made man have the need to build bridges to facilitate transport between two places and to improve communication between them, [6]so that the community of these sectors benefit, but also be able to provide care for tourist places such as rivers, streams, among others. [7]

Similarly, the ELL-high group wrote sentences such as:

[1]One cause of erosion is the silting caused by a precipitation (rain) [2]that creates nonexistent rivers with sediment affecting and/or polluting the water resources, [3]which lead us to take new strategies [4]that minimize the impacts on the environment. [5]

This sentence has a total of 4 finite clauses in one sentence. The sentence also includes some non-finite clauses (e.g. *caused by...*), but remember that those do not affect the value of clauses per sentence according to the system employed in the analysis.

In contrast, sentences in the EL1 group exhibited fewer clauses per sentence.

The following example for instance contains only one finite clause:

[1]North Williams Avenue and North Vancouver Avenue operate as a one-way couplet for approximately two miles between CityName1's central city and the neighborhoods on the southern banks of the Columbia River [2].

In short, based on the statistical results for clauses per sentence, the two ELL groups exhibited significantly more clausal complexity than the EL1 group, as measured by the number of clauses per clause. No significant differences were found between ELL-low and ELL-high groups.

4.2 Dependent Clauses per Clause

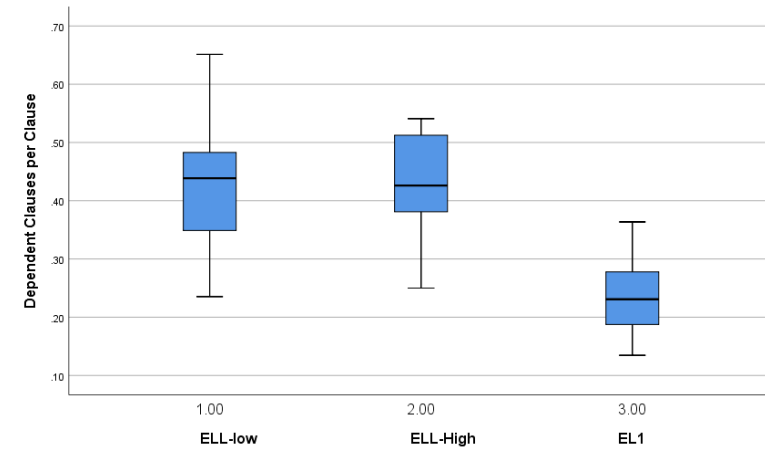
Table 4.2 summarizes the median values for dependent clauses per clause. As seen in the table, results for this measure are similar to the previous feature in the sense that both ELL groups have very similar median values. However, the two ELL groups showed a considerable amount of internal variation as seen in Figure 4.2.

Table 4.2 Medians for Dependent Clauses per Clause

Group	Median
ELL-Low	.43
ELL-High	.42
EL1	.23

The box plots in Figure 4.2 show that the ELL-low group had the most variation among the three groups and it also had the highest number of dependent finite clauses per clause in comparison to ELL-high and EL1. Based on the diagram, it is also possible to say that there were no large differences in means between the two ELL groups. In addition, the EL1 group differed drastically from the other two groups. This third group had less variation, and 50% of the data appeared evenly distributed in the first and third quartile. In contrast, the values for the two ELL groups were non-normally distributed as can be seen by the skewness in each plot.

Figure 4.2 Box Plots for Dependent Finite Clauses per Clause



The Kruskal-Wallis test showed a significant difference of means ($H = 42.146$, adjusted $p < .05$). Pairwise comparisons indicated similar results as for clauses per sentence. That is, the group EL1 was significantly different from ELL low and high groups ($p < .05$). No significant differences were found between groups ELL-low and ELL-high.

These results are additional evidence that the two groups of ELL writers had more clausal complexity in comparison to the EL1 group. In other words, ELL writers had more dependent finite clauses in their sentences. For example, the ELL-low level wrote sentences such as:

[₁As it has been seen₁], bridges are structures [₂that can change the lives of human beings₂] [₃because they mean more than access to a territory initially divided by geographical characteristics₃]

In this sentence, the writers employed an adverbial clause (yellow, subscript 1), a relative clause that modifies *structures* (green, subscript 2), and a subordinate clause (in blue, subscript 3). Writers in the EL1 group, in contrast, wrote fewer dependent finite clauses

per sentence as in the following example which presents a complement clause (in brackets) which is a relative clause modifying the head noun *plan*:

Portland State University (PSU) is pleased to submit this work plan [that documents the tasks and scope of the proposed project.]

4.3 Mean Length of Clause

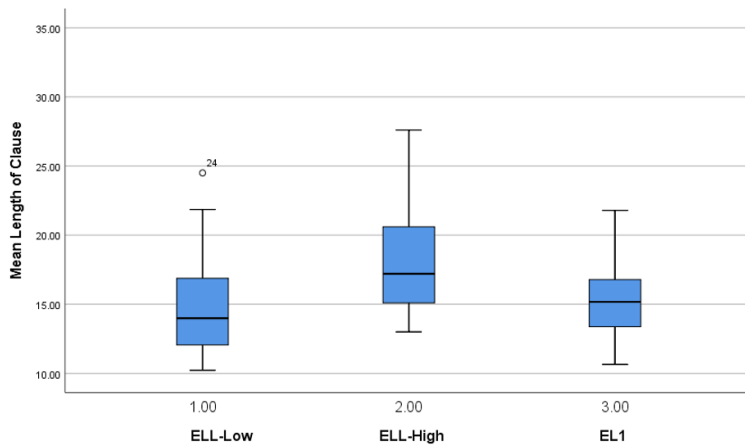
Table 4.3 indicates the values of central tendency. The median values show that the three groups had results close to one another. The median values for the two ELL groups indicate that there was an increase in the mean length of clause as levels of proficiency increased, but there was no linear relationship among the three groups. In addition, the three groups exhibited great variability as shown in the box plots (Figure 4.3).

Table 4.3 Medians for Mean Length of Clause

Group	Median
ELL-Low	13.98
ELL-High	17.20
EL1	15.20

As seen in the figure, the ELL-high group had more words per clause with a median of 18 words per clause. The EL1 group followed with a median value of 15 words per clause. In last place, the ELL-low group exhibited the lowest median value at 13 words per clause. The difference in medians among groups is only of two to three words as was shown in table 4.3. For the ELL-low group, there was an extreme value (shown with a dot). For this extreme case, the genre and topic are the same as the other texts in the group, but this text had longer clauses because it wrote full sentences for the sections where it describes objectives and purposes. On the other hand, most of the other texts in that group had objectives written with non-finite infinitive clauses (e.g. *to build a cover for a sports facility; to conduct an analysis...*)

Figure 4.3 Box Plots for Mean Length of Clause



The Kruskal-Wallis test did not find statistically significant differences ($H = 6.161$, adjusted $p > .05$) among the means. This indicates that the three groups had similar levels of phrasal complexity as per the definition of this measure in this study (i.e. words per finite clause). However, more of the ELL high group had higher values (as seen in the long tail above the median), and even the lowest of the higher-level group was close to the median of the ELL low group. For instance, the following sentence from an ELL-low text has a total of 18 words in one finite clause (e.g. *a project has been proposed to generate greater security when crossing the bridge that connects the Pasto-Ipiales*). In the next example taken from an ELL-high group, the number of words goes up in comparison to the ELL-low text, with a total of 29 words in one finite clause (e.g. *during the development of the semester a detailed analysis of different factors involved in the dam was carried out in the channels that evacuate the water through a landfill*). Finally, the EL1 texts range between 11 and 23 words per finite clause as seen in the following example (e.g. *TriMet's transit Route 4 is a popular transit route and offers 15 minute or better headways throughout the day*).

4.4 Coordinate Phrases per Clause

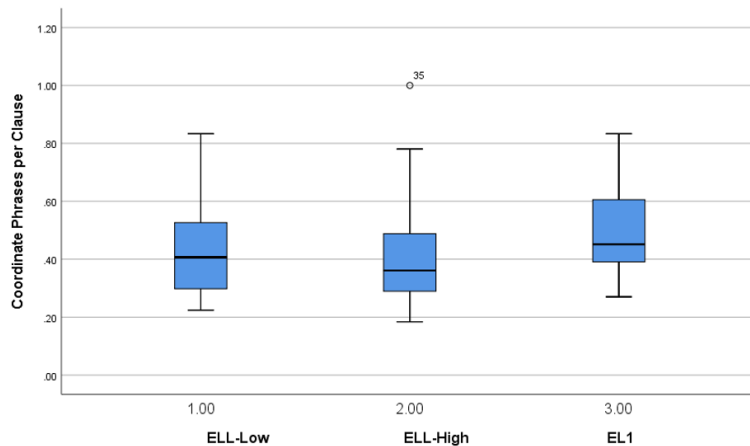
The median values for coordinate phrases per clause indicate a small decrease from the low to high group, and a decrease from the ELL high group in comparison to the EL1. There was internal variation within each group as seen in Figure 4.4.

Table 4.4 Medians for Coordinate Phrases per Clause

Group	Median
ELL-Low	.40
ELL-High	.36
EL1	.45

Boxplots in Figure 4.4 indicate non-normally distributed values for the three datasets. However, the diagram indicates similar levels of variation in the three plots. All the groups have some students who are higher, but none of them were consistently using coordinated phrases. Based on the diagram and the median values (table 4.4), EL1 writers employed slightly more coordinate phrases per clause and the ELL-high group exhibited the lowest median value for this feature of complexity.

Figure 4.4 Box Plots for Coordinate Phrases per Clause



As with mean length of clause, the Kruskal-Wallis showed no significant differences for coordinate phrases per clause ($H = 3.821$, adjusted $p > .05$). This result

suggests that writers at different levels of proficiency used about the same amounts of coordinate phrases per clause, and there was no difference between ELL and EL1 groups as seen in the following examples.

Our project consists of the seismic events, which are natural phenomena whose occurrence, momentum and location aren't controllable by man.

In the previous example taken from an ELL-low text, writers employed a coordinate noun phrase in one clause. In this other example, taken from an EL1 text, writers also employed one coordinate phrase per clause:

a site survey, geotechnical survey, and a meeting with the client will be scheduled.

4.5 Complex Nominals per Finite Clause

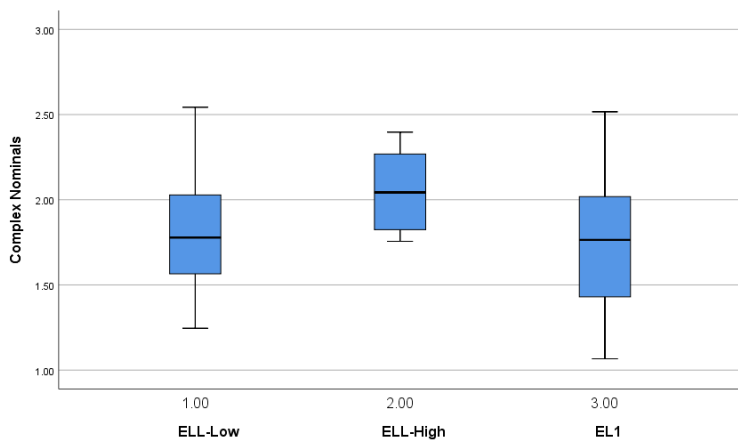
For this last feature of complexity, the median values indicated that the number of complex nominals per clause increased in relationship to proficiency level from low to high groups, but it did not increase for the EL1 group. In fact, the ELL-low group exhibited similar median values as the EL1 group. The box plots in Figure 4.5 provide more information on the distribution.

Table 4.5 Medians for Complex Nominals per Clause

Group	Median
ELL-Low	1.77
ELL-High	2.04
EL1	1.76

As seen in Figure 4.5, on average, group ELL-high had the highest value of complex nominals per finite clause in comparison to the other two groups. ELL-low and EL1 had very similar median values. The EL1 group had a great deal of variability compared to the ELL groups. The ELL-high group had the least variation, which means that this group used complex nominals more consistently than the other two groups

Figure 4.5 Box Plots for Complex Nominals per Finite Clause



The Kruskal-Wallis test showed no significant differences among the groups ($H = 8.819$, adjusted $p > .05$). This result is evidence that groups did not differ significantly in their amount of complex nominals. However, there is a linear increase between proficiency levels in the ELL groups. As the boxplots show, the number of complex nominals between the two ELL groups increases as writers achieve a higher proficiency level. Notice also that the ELL-high group has less variation and more consistency. This same reduction of variation seen between the ELL groups was seen in clauses per sentence and dependent clauses per sentence.

To exemplify some of the results, consider the following sentence taken from the EL1 group: *The surrounding neighborhood is [a new **subdivision** under construction with only a handful of structures built to date].* This example, there is one complex nominal (surrounded by brackets) in one finite clause. The noun in bold is the head of the complex nominal structure which has a pre-modifying adjective (in blue and underlined) and two prepositional phrases that describe conditions of the head noun (in yellow, and in green respectively, double underlined).

From the ELL-high group, the following example illustrates a different way in which writers employed complex nominals:

[₁The *problem* presented at the beginning of this study₁] is [₂the *erosion* for silting that occurs in the micro watershed of Pasto₂].

Writers in the ELL-high group wrote more complex nominals per clause on average, as seen on the diagram and as seen in the previous example. This sentence has two complex nominals with the head nouns *problem* and *erosion*. The first one is only made complex by adding a non-finite clause as a post-modifier (in yellow, underlined), whereas the second one is made complex with two post-modifiers being the prepositional phrase (in green, double underlined) and the relative clause (in blue, bold underlined). The only difference is that the second example did not include a pre-modifier.

4.6 General Patterns of Syntactic Complexity in Civil Engineering Student Writing

In general, the analysis and comparison of all sets of data indicated that statistically significant differences were only present between ELL groups altogether and the EL1 group. No significant differences were found between the ELL-low and ELL-high groups. The differences were only present in the measures of clausal complexity as seen in table 4.6. Both ELL groups had significantly more clausal complexity than the EL1 group. That is, both ELL groups had more clauses per sentence and more dependent finite clauses per clause. The tests did not show any difference in terms of phrasal complexity among groups. This result indicates that on average, the three groups employed about the same amount of words per clause, coordinate phrases per clause, and complex nominals per clause.

Table 4.6. Differences in Complexity Values among the Groups

Kind of Complexity	Measure	ELL-low vs ELL-high	ELL-low vs EL1	ELL-high vs EL1
Clausal Complexity	Complex Clauses per Sentence	-	*	*
	Dependent Finite Clauses per Clause	-	*	*
Phrasal Complexity	Mean Length of Clause	-	-	-
	Coordinate Phrases per Finite Clause	-	-	-
	Complex Nominals per Finite Clause	-	-	-

* indicates a statistically significant difference (adjusted $p < .05$); - indicates a non-significant difference (adjusted $p > .05$)

In addition, based on the comparison of all boxplots, and the values of central tendency, there was a reduction of variation between the two ELL groups when levels of proficiency increased. That is, ELL-high texts had more consistent types of clausal complexity (i.e. clauses per sentence and dependent finite clauses per clause) and complex nominals per clause.

Finally, the statistical results and the comparison among groups did not provide enough evidence to determine an overall development from clausal complexity to phrasal complexity from ELL-low to ELL-high to EL1.

In order to have a better understanding of how all the features work together at the discourse level, the following examples illustrate how the three groups employed syntactic complexity. Notice in particular how examples 1 and 2 have features of clausal complexity (i.e. clauses per sentence and dependent clauses per clause). Note also how examples 1 and 2 include significantly higher clausal complexity in comparison to example 3. All examples are marked with brackets to indicate how many clauses there are in each sentence.

Example 1. ELL-Low

In the municipality of San Bernardo, [1]there is a problem of sports training [2]because the current sport center is not enough to meet the needs of the entire population.[2] [1]. [3]The municipal administration believes [4]that it's convenient to build a citizen integration center[4] [5]so that all inhabitants can develop their sporting activities and lead a healthy life[5] [3].

Example 2. ELL-High

[1]The erosion is the degradation of the soil caused by natural factors or processes, [2]which in critical cases, can cause the loss or partial destruction of a land.[2] [1] [1]One cause of erosion is the silting caused by a precipitation (rain) [2]that creates nonexistent rivers with sediment affecting and / or polluting the water resources [3]that lead us to take new strategies [4]that minimize the impacts on the environment. [4] [3] [2] [1]

Example 3. EL1

[1]North Williams Avenue and North Vancouver Avenue operate as a one-way couplet for approximately two miles between CityName1's central city and the neighborhoods on the southern banks of the Columbia River.[1]

As can be seen from examples 1 and 2, ELL writers tend to use more than one clause per sentence to convey their ideas. In contrast, as seen in example 3, writers in the EL1 group tend to use only one clause per sentence.

Notice also how even though there were no developmental patterns in the phrasal complexity measures, all examples show a great deal of nominal complexity. Example 1 for instance, shows some noun phrases modified with prepositional phrases (e.g. *the needs of the entire population*). Example 2 employed noun phrases modified not just by prepositional phrases but also with relative clauses (e.g. *the degradation of the soil caused by natural factors or processes, [2]which in critical cases, can cause the loss or partial destruction of a land[2]*). Finally, example 3 employed a similar number of complex

nominals as the EL1 group because it used more prepositional phrases to modify head nouns (e.g. *two miles between CityName1's central city and the neighborhoods on the southern banks of the Columbia River*).

Another interesting result worth mentioning is that all texts included high numbers of non-finite clauses as in example 2 (e.g. *caused by natural factors or processes; caused by a precipitation rain; affecting or polluting the water resources*). Even though these structures were not accounted for in the analyzer, they affect the word count in each sentence and they contribute to other dimensions of complexity not seen with the instrument I used. This issue is discussed further in the next chapter.

Now that the results have been presented and some examples have been analyzed, the following chapter will review the research questions of the present study and the answers to them.

Chapter 5: Discussion and Conclusions

In the previous chapters, I reviewed the literature on syntactic complexity in L2 writing and its relation to proficiency and L1 background. I also presented the goals of my study and the methods to investigate the differences of syntactic complexity between ELL and EL1 students' writing in civil engineering. Then, I presented and described the results of my analysis. In this chapter, I summarize my findings and discuss how they helped answer my research questions. Next, I discuss the pedagogical implications for civil engineering student writing as well as ESP-based language programs. Finally, I close the chapter by discussing the limitations found in the study and provide some recommendations for further research.

5.1 Research Questions and Summary of Findings

The main purpose of my study was to investigate the differences in syntactic complexity in the writing of EL1 and ELL (Spanish L1) civil engineering students. I also investigated the relationship between writers' proficiency and levels of syntactic complexity: clausal complexity and phrasal complexity. My study addressed two questions, whose answers I summarize and discuss below.

Research Question 1. Is there a statistically significant difference in the features of syntactic complexity in engineering papers written by ELL lower-level, ELL higher-level, and EL1 students?

As table 5.1 shows, significant differences were found only for the two measures of clausal complexity (i.e. clauses per clause and dependent clauses per clause) between the two ELLs and the EL1 group. No significant differences were found between ELL-low and ELL-high groups.

Table 5.1 Pairwise Comparisons for Clausal Complexity

Groups	Adjusted Significance ($p < .05$)
ELL-Low vs EL1	Significantly different
ELL-High vs EL1	Significantly different
ELL-Low vs ELL-high	Not significant

Moreover, no statistically significant differences were found for the measures of phrasal complexity (i.e. coordinated phrases per clause, complex nominals per clause and mean length of clause). These findings mean that the writing of Spanish-speaking ELL writers was more clausally complex than the EL1 group. Overall, ELL texts altogether had more than one dependent finite clause per sentence and more clauses per sentence. The EL1 group had typically one dependent finite clause per clause.

In terms of phrasal complexity, there were great levels of variability for the three groups, but the median values did not differ significantly. For mean length of clause, the higher-level ELL group had the highest value and the greater variability. However, for complex nominals per clause, the higher-level ELL group had the most consistency in comparison to the other groups. These two measures are an indication that there was a lot of phrasal complexity involved in the higher ELL group.

The lack of significant differences in clausal complexity between the two ELL groups, and the lack of significant differences in all the measures of phrasal complexity can also be explained with the variability that was exhibited within each of the groups as was seen in the box plot figures in the previous chapter. For the three groups, there were extreme values that were far from the central values in all the complexity measures. This variability could be a sign that better instructional practices are needed so that those writers who had very long and complicated sentences develop phrasal complexity and express their ideas more clearly and concisely. For example, this sentence written by a

higher-level ELL student group has several ideas in the same sentence, which make the text look unclear and complicated:

This research presents the approach of a possible solution to the problems found in Genoy, Nariño in which a stagnation of water in a stream is generated, causing overflows that cause the continuous deterioration of the soil present in the area and its deterioration is intensified with the raising of livestock in the area.

Students who tend to use this type of writing in which they express multiple ideas with complex structures could benefit from clear instruction. In particular, attention should be focused on writing simple sentences (i.e. one clause per sentence) so that they are easier to follow and to understand. Technical information could be expressed via complex phrases such as in the following example:

The Sisga reservoir is a dam made of concrete located in the course of the Sisga river in the municipality of Choconta in the department of Cundinamarca.

The previous example clearly indicates that it is possible to have one simple sentence (clausally speaking) and which carries important information expressed through phrasal complexity feature. For example, the prepositional phrases in the sentence provide exact and precise information about where the reservoir is located.

In regards to the differences in clausal complexity between ELLs and EL1s, a possible explanation could be based on linguistic transfer. It is often mentioned that Spanish written discourse produces heavy levels of subordination and clausal embedding both in novice and expert writing (Neff, Dafouz, Díez, Prieto & Chaudron, 2004; Ortega, 2003). In my data, the amount of clausal complexity revealed in the ELL texts might be

due to a linguistic transfer from their L1. Consider the following example taken from one of the higher-level ELL texts that was partially written in Spanish:

[₁La dosificación de la mezcla es la fase final del diseño]₁] [₂pues se proporcionan las cantidades de los diferentes materiales en base al cemento]₂] [₃y que además cumplen con las características de diseño preestablecidas.₃]

ENG: [₁The dosage of the mixture is the final phase in the design]₁] [₂as the quantities are proportionate to the different materials based on cement]₂] [₃and which also meet the pre-established design characteristics.₃]

Notice the sentence in Spanish has a high level of clausal complexity because it has three finite clauses and two finite dependent clauses. This kind of writing is typical in Spanish. Therefore, a linguistic transfer from Spanish into English is likely, especially if students write their project first in Spanish and then translate it directly without considering the effects it will have in terms of conciseness and ambiguity. Such effects would be an interesting aspect to include in a future study.

Research Question 2: Is there overall development from clausal to phrasal complexity as levels of language proficiency increase from ELL-lower-level to EL1?

To answer this second research question, I looked at the statistical tests and the medians to determine if there was a shift from clausal to phrasal complexity as levels of proficiency increased (i.e. ELL-low to ELL-high to EL1). Based on previous literature (Mazgutova & Kormos, 2015, Parkinson & Musgrave, 2014), it was expected that as levels of proficiency increase, clausal complexity decreases and phrasal complexity increases.

As expected, and as seen in Figures 5.1 and 5.2 levels of clausal complexity (shown with blue lines) do decrease in the more proficient group (i.e. EL1). However,

this decrease was not continuous from level to level. In other words, there was no decrease from the lower-level ELL group to the higher-level ELL group.

Moreover, even though the EL1 group exhibited lower levels of clausal complexity, the EL1 group did not decrease. As the lines in Figure 5.2 show, the median values slightly increased from low-level to high-level, but went down to EL1. All groups appeared to use similar levels of phrasal complexity. There was no significant distinction in the way they employed complex nominals or coordinate phrases per clause.

In short, the results indicated that there was no overall development from clausal to phrasal complexity. There was not a consistent decrease in clausal complexity and increase in phrasal complexity as proficiency levels increased. This finding differs from what was found by Lu and Ai (2015). In their study, which compared the English written production of learners with different L1 backgrounds, Lu and Ai found that as levels of proficiency increased, so did the levels of phrasal complexity, and the levels of clausal complexity decreased.

Figure 5.1 Trends in Measures of Clausal Complexity in All Groups

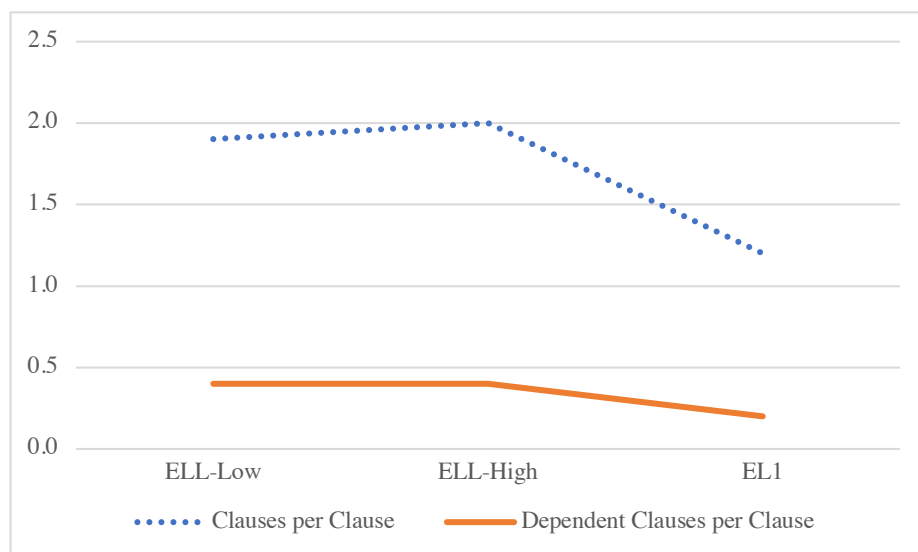
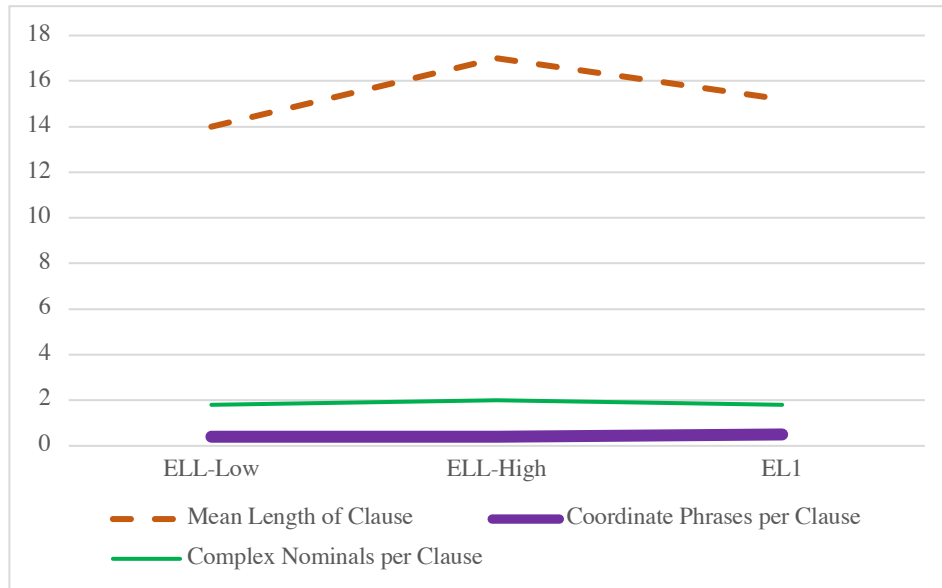


Figure 5.2 Trends in Measures of Phrasal Complexity in All Groups



A possible explanation for why ELL students relied on high amounts of clausal complexity is that students needed to meet certain criteria to complete their PI Project (Integrative Project). For example, based on my personal experience teaching these students, a requirement for the projects was to have a minimum of three pages written in English. It is possible that they had a tendency to draw on heavy amounts of embedding and clausal complexity in general to fill the word count and to increase their number of pages in their projects. In addition, the same way I had the belief that clausal complexity was a skill that students needed to develop, it is possible that other teachers emphasized this linguistic feature in their writing classes, and these instructional decisions could have had an impact in the grammatical choices that students made.

5.2 Pedagogical Implications

Even though there were no significant differences between the two ELL groups, it is important to take into account that these two groups together differed significantly from the EL1 group in the two measures of clausal complexity. This is an important

aspect that needs to be considered and paid attention to because of the stakes that clausal complexity has in engineering writing. As Conrad (2017) demonstrated, professional engineers in the industry tend to use more phrasal complexity and value having one clause per sentence given that in this way the ideas are more clear, concise and unambiguous. Therefore, if ELL engineering students have the belief that having more clausally complex sentences means being more proficient in the language, it might create a mismatch between their way of writing and what is typically done in actual engineering texts. In addition, if linguistic transfer is an actual factor that is causing students to be more clausally complex in their writing, instructors should address this issue by raising awareness about these differences, and that using linguistic features from the L1 might not be a good strategy for effective writing in engineering.

Therefore, there is a need for instructors to challenge the stereotype that advanced writing is complex at the clause level, and expose students to samples of professional engineering writing that exemplifies the syntactic structures typical in this discipline. Especially, students should be aware that writing long and complicated sentences may not be useful and effective in the civil engineering industry if they plan to work abroad or in an international firm where they have to use English as their principal medium of communication. For instance, a quality that is highly valued in engineering practice involves providing precise and accurate information. Syntactically speaking, this includes reverting to simple sentence structure (i.e. one finite clause per sentence) and reverting to complexity at the phrase level to provide exact information (e.g. *the difference in grade between the inbound and outbound directions makes the inbound downhill easier for cyclists than the outbound uphill direction*).

In addition, even though there were no significant differences in clausal complexity between ELL-low and ELL-high, there were individuals who clearly would benefit from instruction. The large variation found within the two ELL groups is a sign that they need explicit instruction on L2 writing. A great place to start would be to dispel myths about complexity in writing, and to raise awareness about the importance of developing effective writing skills for the workplace. Feedback should be provided to students in such a way that they recognize the usefulness of the language skills they are learning. Language programs in EFL contexts usually teach general grammar components which then students have to figure out how to implement in their future careers. But if they know from the beginning why they are learning something, they might benefit more from the feedback they are provided in their classes.

Higher education institutions where English is not the primary medium of instruction and which are interested in adopting ESP-based curricula should also raise awareness about the different levels of syntactic complexity, and emphasize on the idea that clausal complexity is not the only goal in developing writing skills in a foreign language. Particular attention should be given to the specific areas of study of each program that adopts an ESP approach so that they can determine better practices that aim at developing the levels of grammatical and syntactic complexity that match the needs of their specific discipline, context or area of expertise.

ELL civil engineering students in an ESP program might benefit if they were exposed to authentic materials as those used in the real practice of their careers. Assigning students general academic writing tasks (i.e. argumentative essays) may not be the most productive way of teaching them how to use a language for their specific

purpose. Rather, students would enjoy their time learning a language much more if they were learning how to write for the different tasks an engineer does such as technical memoranda, project proposals, site descriptions, structural analysis, and email writing.

In addition, civil engineering programs in the US or in contexts where English is the main medium of instruction, should consider that their ESL students might need particular feedback in their writing if it does not meet the expectations of the field. Students with different L1 backgrounds other than Spanish might have different types of linguistic transfer, and even though civil engineering instructors are not necessarily experts in writing or language instruction, they could show samples of effective writing in the workplace and raise awareness about the grammatical components that are characteristic in professional writing in civil engineering. At the same time, even though Spanish L1 ELL students are more likely to have clausal complexity, EL1 students could also benefit from distinguishing between texts with heavy levels of clausal complexity and texts that employ clausally simple sentences but are phrasally complex.

Finally, it is important to recognize that syntactic complexity is only one dimension in L2 writing. There are other areas that instructors need to pay attention to when teaching writing skills. For instance, texts are more effective for readers if they have good information flow (e.g. known information before new information), appropriate word choice for the genre and style, and important elements such as coherence, cohesion, and effective spelling and punctuation.

5.4 Limitations and Future Research

Several factors limited how well I could meet the goals I set out for my study. In addition, there are goals that I never considered but that would be worth exploring in future research.

One of the first limitations that I considered from the beginning was the size of my data sets. With only 74 texts, the results I found are not representative of all civil engineering students nor all Spanish-speaking ELL students. Originally, I wanted to have at least 30 texts in three different proficiency levels for the ELL groups (i.e. beginner, intermediate and advanced), but I could only use the ones that were available in the archive from Universidad Mariana, which only contained 30 texts in the beginner level and 14 texts in the advanced level. I was able to collect texts from another Colombian university. However, the texts I collected had a very different topic and genre in comparison to the texts from Universidad Mariana and the ones from the CEWP corpus. Therefore, as I was trying to control genre and topic to be as similar as possible, I could not use the texts from the second Colombian university. Consequently, a future study could have larger sets of data to have a more representative sample. Also, it would be necessary to have all texts in the same genre and topic because this could potentially affect the way writers change their syntactic structures.

Another important limitation was concerned with the instrument I used. The L2SCA only accounted for structures that contained finite clauses, ignoring non-finite clauses (e.g. *This bridge was designed to benefit two important towns in southwest Colombia*). This particular issue with the instrument could be one of the reasons why it was not possible to see differences and linear relationships between groups. When

inspected individually, most texts contained high levels of non-finite structures. Thus, having a more finely tuned analysis of grammatical features could provide a better insight into all the dimensions of syntactic complexity. The L2SCA serves the purpose of analyzing large sets of data automatically, which is an advantage for large-scale studies, but the operational definition of clauses is an issue that needs to be considered in future studies that intend to use this instrument. Perhaps, a future study could create their own operational definitions that include more finely grained measures of complexity. Nonetheless, this would require researchers to write their own software and it would be time consuming and demanding. Another option could be to have texts coded manually, a task that would also be demanding and tedious, but worth considering, especially if it is only with a small sample.

Another limitation based on the features of the instrument was that I only analyzed the first 2000 words of texts that had more words than this number. I did not take samples from sections across the texts. That is, I did not include samples from the middle or end sections. Therefore, studying complexity across sections could be a decision to consider for a future study.

It would also be interesting to have data from different Spanish-speaking universities and even from different Spanish-speaking countries although this could create issues with selection criteria. For instance, as mentioned earlier, I had collected texts from another Colombian university, but the texts were on a completely different topic and genre, and were written individually. Therefore, they were not comparable with the other corpora as topic and number of writers had to be similar. Researchers interested

in using data from multiple sources should consider that it might be difficult to find already existing corpora with similar criteria.

Another limitation is that I used texts that were written in groups. A future study could come up with a better design where several variables are controlled for. For instance, students could first be assessed on their proficiency level before and then be assigned a timed writing task where they work individually. This would control for things we cannot see behind the process of already existing data. One problem with the texts that I used from Universidad Mariana, for example, is that it is possible that students wrote their texts in Spanish first, and then used a tool or dictionary to translate sentence by sentence. This is knowledge I have based on personal experiences while teaching at that university to civil engineering students. To avoid students using direct translation then, I would have all my participants in the same time and place to make sure they are writing in the target language, on the same topic and the same genre.

Also, I only analyzed papers that had a passing grade. I did not analyze texts based on the exact score they received and did not analyze papers that did not have a passing grade. A future study could investigate whether grades correlate with the different measures of syntactic complexity. Similarly, I did not analyze texts that did not have a passing grade, because I did not intend to analyze student errors. However, a future study could analyze lower-scoring papers and investigate errors and examine the use of complexity features with more through a qualitative analysis.

In addition, a future study could contrast English language programs in EFL contexts. For example, researchers could collect data from general English acquisition programs and compare the analysis with data from ESP programs. I only analyzed data

from an ESP program for civil engineering students. An analysis of data from non-ESP programs could inform about the effect of instruction with this particular teaching approach in comparison to using other methods to teach English at the university level. Also, another study could analyze the writing of other ESP areas, not just civil engineering.

Moreover, in order to determine whether the L1 background has an effect in written syntactic complexity, a future study could analyze writing not just of the second language but also the first language. Researchers could collect texts written in both languages using the same conditions: topic, genre, time, task. This procedure would allow for a systematic understanding of the relationship between L1 and L2 and would more accurately inform whether there was linguistic transfer from one language to the other.

An additional idea for a future study is to analyze the writing of civil engineering in Spanish. Spanish is generally characterized by being clausally complex, but there are no studies that have looked at how syntactic complexity features are used in engineering writing in that language. Moreover, from a world Englishes perspective, it would be interesting to see how complexity features vary in the English writing production of engineers working in multinational firms.

Another interesting direction of research would be to take a different approach to determine development in syntactic complexity in civil engineering writing. The present study adopted a cross-sectional approach, but a future study could conduct longitudinal research where the same group of writers are studied through time. An analysis of this kind could inform the changes that happen over time, and could show whether there is a shift from one level of complexity to the other (i.e. clausal to phrasal complexity).

Finally, a future study could also analyze texts from the practitioners' collection in the CEWP corpus. An analysis that compares student writing with expert writing could tell us more information about the relationship between expertise and choices in grammatical structures in writing. A great advantage of using the CEWP corpus is that it has several genres. I only analyzed students' proposals, but there are other genres that could be analyzed and compared with ELL students' writing.

5.5 Final Thoughts

Even though much more work remains to be done, this study contributes to L2 writing research because it explored an understudied language background (i.e. Spanish), it provided a view of the variety of syntactic structures employed in an ESP context (i.e. civil engineering), and it employed data different from the typical argumentative essays that are analyzed in large-scale studies (i.e. civil engineering project proposals). More work needs to be done in other discipline-specific areas where writing plays an important role because we cannot assume that the writing trends found in general academic writing (i.e. argumentative essays) will be the same everywhere. Therefore, having an informed understanding of the grammatical patterns and syntactic characteristics of writing production in each discipline might benefit instructors when creating lessons and materials for teaching written composition in their specific field.

References

- Ai, H. & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Diaz Negrillo, N. Ballier, and P. Tompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249-264). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11(1), 17-34. doi:10.1017/S0272263100007816
- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, 22(2), 185-200. doi:10.1007/s11145-007-9107-5
- Berthouex, P. (1996). Honing the writing skills of engineers. *Journal of Professional Issues in Engineering Education and Practice*. 122(3), 101-110. doi:10.1061/(ASCE)1052-3928(1996)122:3(107)
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1992) On the complexity of discourse complexity: A multidimensional Analysis. *Discourse Processes*, 15(2), 133-163. doi:10.1080/01638539209544806
- Biber, D. & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics*, 15(2), 223-250. doi:10.1017/S1360674311000025

- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35. doi:10.5054/tq.2011.244483
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639-668. doi:10.1093/applin/amu059
- Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins Publishing.
- Best Colleges, US News Rankings, (n.d.). Portland State University. Retrieved from <https://www.usnews.com/best-colleges/portland-state-3216>
- Conrad, S. (2015). Dispelling student myths about writing in civil engineering. *Proceedings of the 2015 American Society for Engineering Education Conference, USA*. Retrieved from <https://www.asee.org/public/conferences/56/papers/12742/view>
- Conrad, S. (2017). A comparison of practitioner and student writing in civil engineering. *Journal of Engineering Education*, 106(2), 191-217. doi.org/10.1002/jee.20161
- Conrad, S. (2019). Register in English for academic purposes and English for specific purposes. *Register Studies*, 1(1), 168-198. doi:10.1075/rs.18008.con
- Conrad, S. & Pfeiffer, T. (2011). Preliminary analysis of student and workplace writing in civil engineering. In *Proceedings of the 2011 American Society for Civil*

- Engineering Conference, USA*. Retrieved from <https://peer.asee.org/preliminary-analysis-of-student-and-workplace-writing-in-civil-engineering>
- Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, 69(5), 176–183.
doi.org/10.1080/00220671.1976.10884868
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.
- Crossley, S. & McNamara, D. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity and conceptual knowledge. In *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 106-126). Bristol, UK: Channel View Publications.
- Crossley, S. & McNamara, D. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*. 26, 66-79. doi.org/10.1016/j.jslw.2014.09.006
- Gómez Sará, M. M. (2017). Review and analysis of the Colombian foreign language bilingualism policies and plans. *HOW*, 24(1), 139-56. doi:10.19183/how.24.1.343
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). International corpus of learner English (Version 2.0). Louvain-la-Neuve: Presses universitaires de Louvain.
- Halliday, M.A.K., & Martin, J.R. (1993). *Writing science: Literacy and discursive power*. London: The Falmer Press.

- Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (NCTE research report, no. 3) Champaign, Ill: National Council of Teachers of English.
- Hunt, K. W. (1970). *Syntactic maturity in school children and adults*. (Vol. 134). Monographs of the Society for Research in Child Development.
- ICFES, (n.d.). Pruebas Saber Pro. Retrieved from <http://www.icfes.gov.co/web/guest/acerca-del-examen-saber-pro>
- Klein, D. & Manning, C. D. (2003). Fast exact inference with a factored model for natural language parsing. In S. Becker, S. Thrun & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems, 15* (pp. 3–10). Cambridge, MA: MIT Press.
- Kuiken, F. & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-160). Amsterdam: John Benjamins Publishing.
- Levy, R. & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 2231-2234). Genoa, Italy: ELRA.
- Liu, L. & Li, L. (2016). Noun phrase complexity in EFL academic writing: A corpus-based study of postgraduate academic writing. *The Journal of Asia TEFL, 13*(1), 48-65. doi:10.18823/asiatefl.2016.13.1.4.48

- 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. doi: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. doi:10.1016/j.jslw.2015.06.003
- 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. doi:10.5054/tq.2011.240859
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), pp. (474-496)
- Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics*, 14(1), pp. (3-28). doi:10.1075/ijcl.14.1.02lu
- Luzón, M.J. (2009). The use of we in a learner corpus of reports written by EFL engineering students. *Journal of English for Academic Purposes*, 8 (3), 192-206. doi:10.1016/j.jeap.2009.04.001
- Mazgutova, D. & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, 29, 3-15. doi:10.1016/j.jslw.2015.06.004
- McNamara, D., Graesser, A., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

- Ministerio de Educación Nacional de Colombia. (2004). El programa nacional de bilingüismo. Retrieved from www.mineducación.gov.co
- Nausa R. (2017). Syntactic mechanisms in the transition from academic written to oral discourses. *Colombian Applied Linguistics Journal*, 19(2), 234-249. doi:10.14483/22487085.11765
- Nanwani, S. (2009). Linguistic challenges lived by university students in Bogotá in the development of academic literacy. *Colombian Applied Linguistics Journal*, 11, 136-148. doi:10.14483/22487085.160
- NCEES, (n.d.) FE Exam. Retrieved from <https://ncees.org/engineering/fe/>
- Neff, J., Dafouz, E., Prados, M., Prieto, R., & Chaudron, C. (2004). Contrastive discourse analysis: Argumentative text in English and Spanish. In C. L. Moder & A. Martinovic-Zic (Eds.), *Discourse across languages and cultures* (pp. 267-283). Amsterdam: John Benjamins Publishing Company.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555–578. doi:10.1093/applin/amp044
- Obando Guerrero, G., & Sánchez Solarte, A. (2018). Learners' satisfaction in two foreign language teacher education programs: Are we doing our homework? *How: A Colombian Journal for English Teachers*, 25(1), 135-155. doi:10.19183/how.25.1.436.
- Ortega, L. (2015). Syntactic complexity in L2 writing: Process and expansion. *Journal of Second Language Writing*. 25, 82-94. doi:10.1016/j.jslw.2015.06.008

- 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. doi:10.1093/applin/24.4.492
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48–59. doi:10.1016/j.jeap.2013.12.001
- Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching*, 50(02), 261–275. doi:10.1017/S0261444817000015
- Polio, C. (1997). Measures of Linguistic Accuracy in Second Language Writing Research. *Language Learning*, 47(1), 101-143. doi:10.1111/0023-8333.31997003
- Portland State University (n.d.) PSU Bulletin 2017-2018. Retrieved from <http://pdx.smartcatalogiq.com/en/2017-2018/Bulletin/Courses>
- Ranking Web of Universities, (2019). Universidad Mariana, country rank. Retrieved from <http://www.webometrics.info/en/search/Rankings/universidad%20mariana%20type%3Apais>
- Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing*, 29, 51–63. doi.org/10.1016/j.jslw.2015.06.005
- Staples, S., Egbert, J., Biber, D. E., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149-183. doi.org/10.1177/0741088316631527

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

Yang, W., Lu, X., & Weigle, S. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgements of writing quality. *Journal of Second Language Writing*, 28, 53-67. doi:10.1016/j.jslw.2015.02.002

Yoon, H., & Polio, C. (2017). The linguistic development of students of english as a second language in two written genres. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 51(2), 275-301. doi:10.1002/tesq.296

Appendix: IRB Approval

Human Subjects Research Review Committee (IRB)
Research Integrity, Research and Graduate Studies
PO Box 751, Portland, OR 97207
(503) 725-2227 hsrrc@pdx.edu



Date: November 6, 2018

To: Susan Conrad & Santiago Gustin,

From: Jack Barbera, IRB Chair

A handwritten signature in black ink that reads "Jack Barbera".

Re: IRB review of your amended protocol #184824, titled "Differences in Syntactic Complexity of Native and Non-Native Speakers' Writing in Civil Engineering"

Notice of Review and Exempt Determination - Amendment Exempt Review Category 4; as per Title 45 CFR Part 46

The amendment submitted on November 6, 2018 for the above protocol has been reviewed by the PSU Research Integrity office on behalf of the IRB (Human Subjects Research Review Committee). This is a minimal risk research protocol that continues to qualify for an exemption from IRB review under 45 CFR 46.

Please note that you will not be required to submit continuing reviews for this protocol; however, you must submit any changes to the protocol to Research Integrity for assessment to verify the protocol continues to qualify for exemption.

As a reminder, your responsibility as a Principal Investigator also includes:

1. Obtaining written documentation of the appropriate permissions from public school districts, institutions, agencies, or other organizations, etc., prior to conducting your research;
2. Notifying the Office of Research Integrity of any change in Principal Investigator;
3. Notifying the Office of Research Integrity of any changes to or supplemental funding;
4. Notifying the Office of Research Integrity within 5 days of any unexpected events;
5. Retaining copies of this determination, any signed consent forms, and related research materials for three years after conclusion of your study or the closure of your sponsored research, whichever comes last.

As with all human subjects research, exempt research is subject to periodic post approval monitoring and review. If you have any questions regarding your protocol or the review process, please contact the Research Integrity office in Research & Graduate Studies at hsrrc@pdx.edu or call (503) 725-2227.

Amended: **Submission of signed permission form as required during the initial contingent exempt determination.**

EXEMPT DETERMINATION

----- 11/06/2018 -----

Portland State University
Research Integrity Office