CHAPTER I
INTRODUCTION

1.1 Background

The emergence of Artificial Intelligence (AI) has made the public aware of its great potential in reducing workload, be it for company use or in academic settings. Specifically, AI allows workers to generate report, analyze data, handles repetitive tasks, all of which are intended to increase productivity, reduce costs, and overall efficiency. This practice is also done by university students in aiding them with homework, papers, presentation and exams, though as effective as it seems, it has a long-lasting effect of not helping in improving student's academic abilities.

One of the core skills of language learning is writing, and it has its own set of challenge in order to achieve the advanced level. With the help of AI, one would be able to generate a well-written essay with a noticeable rich vocabularies and well-supported arguments. Most importantly, students are able to perform it in a matter of seconds which is very much contrasting to the traditional way of constructing an essay.

While this method can be used in a good way because students could learn from a perfect sample, often that automation could lead to exploitation. They realize that they could do this to every single task given to them to get good grades especially when teachers are unable to distinguish AI-written and human-written texts. Moreover, sometimes teachers do not have a solid proof whether the text is truly made by human or AI, making this practice is used more and more because of little to no consequences.

One of teacher's principle that is clear is that fairness should always be ensured in learning environment, and they have to make sure that any form of cheating is highly discouraged. With AI, students are confident with their score because of its high accuracy, and efficiency, but this is not fair for both skilled and [illegible] students who relies on their own ability. Thus, Teachers should be [illegible] to anticipate this practice.

In this regard, studies have shown the confidence level of detecting AI-generated text. Interestingly, data have shown that experienced teacher have a better chance when detecting a high-level AI text, but it is not as successful when presented with low quality text produced by AI (Fleckenstein et al., 2024). Low quality text in this study is defined as texts that are equivalent with level 2 from the TOEFL iBT writing rubric.

Moreover, tools that are able to detect AI is also available, and teachers could also utilize this. The problem is that a study that measures the accuracy of six different AI detection tools shows only 63% success rate (Cooperman & Brandao, 2024). Though it seems that it could detect AI to some extent, there is still a chance where it could falsely label a human-written text.

As mentioned before, AI can serve as a valuable tool for improving writing skills, offering students opportunities to learn and analyze AI writings. This can be done by examining the distinct characteristics of Second Language writings and AI writings through linguistic features for writing quality and development. These linguistic features include Syntactic complexity, Lexical diversity, density, and sophistication, and cohesion, and they are all generally used for writing research (McNamara, Crossley, McCarthy, 2010).

These indices are able to provide measurement for writing proficiency, evidenced by several studies exploring their relationship with proficiency levels. Some studies found similar result for one index, while others may suggest varying interpretation. For example, some research suggests that certain syntactic complexity indices have either positive or negative relationship with writing proficiency. Positive relationship indicates that the higher the score, the more proficient the user is in writing. While negative relationship means that lower score reflects proficiency. However, there are some indices that have conflicting interpretations, with some researchers found negative relationship while other found positive. This uncertainty opens the possibility of theory expansion which is explored further in this research.

One way to obtain the score of these indices is through a series of coding thon programming, but there is one alternative option available which is this research. TAACO (Tool for the Automatic Analysis of Cohesion), (Tool for the Automatic Analysis of Lexical Diversity), TAALES (Tool for matic Analysis of Lexical Sophistication), and TAASSC (Tool for the

Automatic Analysis of Syntactic Sophistication and Complexity) are free and user-friendly application that could automatically calculate the indices score from given corpus.

Based on the explanation, this research investigated linguistic features from Second Language writings text, AI-generated text, with the addition of Low-level AI generated text. Though, there is an adjustment for Low-level text and the linguistic features selection. Since the participants in this study is unfamiliar with TOEFL iBT writing style and rubrics, this research used AI to mimic the participant's writing style. Thus, Low-level AI generated text is labeled as AI Human Mimic. Furthermore, this research does not investigate Lexical features. This is because lexical features are easy to detect to the point that it is obvious, such that AI text uses more advanced vocabulary that no English user at one particular language level would even use. Cohesion is not this research main focus because AI is able to generate context due to the fact that it is trained with large corpora. Moreover, this ability is natural to human, so there are minor difference in terms of overall cohesion between the two types of text. However, if AI fails to include proper context or cohesion, again, it would be obvious due to our natural ability to understand both within texts. Thus, this research focused on syntactic complexity indices with the main goal of expanding its existing theory while also focus on practical significance for improving student's writing.

Understanding Syntactic complexity in the context of AI and Second language writings could provide insights on advancing L2 writing research and instruction. AI-generated texts can be used as a benchmark for L2 writing improvement. AI could generate texts with varying levels, so a comparison between student's and AI texts could be made in order to identify areas for growth. This approach allows researchers and teachers to observe patterns in syntactic development and to recognize gaps in student's writing.

Therefore, it is important to position AI not as a shortcut, but as a learning tool that can support syntactic development. While this research does not assume that AI always writes better than humans, AI generated texts produced by Large Language Models are trained on large-scale corpora. This result in the model [illegible] in generating structurally fluent texts based on specific prompt [illegible]ns, which allows for clearer comparisons across different levels of writing [illegible]cy. Through comparison, teachers can identify which features are underdeveloped in student writing. This way, AI is not used to bypass writing process, but rather serves as a model that supports L2 syntactic development and writing proficiency.

1.2 Research Question

Based on the study background, this research focus on syntactic complexity from three types of text, Second Language Writings (SLW) texts, AI-written texts, and AI Human mimic texts. The formulated research questions are:

1. How do syntactic complexity indices relate to writing proficiency in SLW, AI-generated, and AI Human mimic texts?
2. To what extent are the syntactic complexity indices effective for L2 learners' writings

1.3 Research Objectives

Based on the research question, the research objectives are:

1. To expand the existing theories on the relationship between syntactic complexity indices and writing proficiency.
2. To provide an effective recommendation for L2 learners writings

1.4 Research Significance

1. Theoretical Significance

This study is expected to contribute to existing theories of syntactic complexity indices, specifically their relationship with writing proficiency by exploring SLW, AI-generated, and AI Human mimic texts. Previous findings suggest a fixed relation to writing proficiency for some indices, while there are other indices that suggest two distinct relations at the same time. Therefore, insights from the three groups would reveal new relations or confirm the existing one.

2. Practical Significance

Practically, the study findings can be used as a reference for writing instructions that emphasize in improving syntactic complexity in writing. The findings could suggest which area of sentence structure that needs to be practiced hat exercises are developed to improve writing skills. This could also lead ɘ structured materials for writing enhancement, such that creating levels difficulties in syntax in order to progressively build student's writing skill.

CHAPTER II
REVIEW OF RELATED LITERATURE

2.1 Previous Studies

There are studies that relate to this research. These studies rely on linguistic feature indices from NLP tools as a means to explain or solve an issue, and some of them are implemented in various areas of science.

The first study, conducted by Clarke, Foltz, & Garrard (2020), attempts to find a new approach to the early detection of Alzheimer's disease by investigating changes in patients' spoken and written language. The research focuses on natural language components such as lexical properties and richness, syntactic complexity, coherence and cohesion, entropy and perplexity, semantics, and sentiment. The large dataset was processed using a variety of Natural Language Processing Tools, including TAALES, TAALED, TAASSC, Automatic Speech Recognition, and others. The study suggests that this approach has potential for early detection. However, several issues must be addressed before it can be applied in clinical practice, including clinical acceptance, data ethics, and disease-specific diagnostics.

Liang et al. (2022) conducted a study to investigate the characteristics of language usage in patients with schizophrenia. They aimed to identify differences in fluency (Thought and Language Index), lexical cohesion (Givenness, repeated content words or lemmas), and syntactic complexity (MLS, MLT, MLC) between 66 patients who had just begun experiencing schizophrenia and 36 healthy control patients. They also measured the quantitative differences in glutamate levels in the brain. The results show that the group with near-normal cortical thickness patterns had similar language production to that of the healthy controls. Meanwhile, the group with widespread cortical thinning and higher glutamate levels were considered fluent but produced a reduced mean length of T-units (complexity) and fewer repeats of content words (lexical cohesion).

Another study, conducted by Mizumoto & Eguchi (2023), involved utilizing ıguage model for automated essay scoring. The study created seven ıf automated essay scoring: (1) GPT scores only, (2) GPT scores and easures, (3) GPT scores and syntactic complexity measures, (4) GPT scores and fine-grained syntactic dependency and verb argument construction, (5) GPT scores and cohesion measures, (6) GPT scores and all linguistic measures from models 2-5, and (7) linguistic measures only. The linguistic measures for models 2-7 were obtained from TAALED, TAALES, TAASSC, and TAACO. The results show that models 3, 6, and 7 outperformed model 1, suggesting that automatically scoring essays is more effective when GPT scores are used alongside linguistic features.

A study by Tabari & Johnson (2023) focuses on identifying the differences in cohesive devices used between narrative and argumentative essays written by L2 students. This study employed multiple regression analysis to determine whether cohesion indices from TAACO could recognize these variations and which cohesive devices were used between the two essay types. The results show that narrative texts used connective devices for cohesion, while argumentative texts relied on global-level repetition.

There is also a study that investigates between human texts and AI-generated texts. However, this study uses different aspect to characterize the differences. Sardinha (2024) studies the texts generated by AI and human texts (written and spoken) from the perspective of five dimensions identified by Biber. The five dimensions are: (1) Involved versus Informational Productions, (2) Narrative versus Non-Narrative Concerns, (3) Explicit versus Situation-Dependent Reference, (4) Overt Expression of Persuasion, and (5) Abstract versus Non-Abstract Information. The study uses a large dataset of texts from academic writing, Conversations, essays, and news articles, and they are analyzed using ANOVA and coefficient of Determination. The result shows that AI-generated texts could not fully imitates the natural patterns of human produced speech and writings. Specifically: (1) AI are less interactive and involved, (2) AI is more informational, yet seems to lack narrative style, (3) AI have difficulties in making explicit reference and using context-dependent language.

Meanwhile, this study focuses on exploring Syntactic Complexity indices from TAASSC in three different types of text, texts from second language students, texts generated by AI, and texts generated by AI that mimics the writing style of the ...anguage students. This study aims to expand the existing theories on the ...hip between syntactic complexity indices and writing proficiency and to ...nsights for writing instructions targeting specific syntactic complexity features. This study looks into syntactic complexity indices (Lu, 2010). These indices are analyzed quantitatively. The mean of the indices is compared using ANOVA and then the Effect size is determined using Eta squared. The limitations for this study are that, the students are from the second semester of college and only studies English up until high school. In regard to this, all students could only create one paragraph because of their inexperience in writing in English. Furthermore, prompts given to AI might not create the intended text quality. AI creates something based on the instructions given to it, so different instructions creates different result. All of these limitations open the possibility of further research, so that in the future, new insights could be discovered by addressing these limitations.

2.2 Theoretical Background

2.2.1 Theory of Syntax

Syntax refers to the set of rules and principles that guide the construction and organization of sentences. These rules dictate how words and phrases are combined according to grammatical standards, allowing us to form coherent and meaningful sentences. Interestingly, each language has its own distinct sentence structures, which prompts us to consider whether syntax alone accounts for all linguistic differences, or if it serves as a universal framework for understanding language structure. As Chomsky (2002) notes, the aim of studying syntax is to develop a grammar that can be used to construct sentences in a specific language. Ultimately, this analysis should contribute to a broader theory of linguistic structure, one that systematically examines and conceptualizes the methods used in various grammars without being limited to any single language. Such a theory would offer a comprehensive approach to selecting the appropriate grammar for any language based on a sample of its sentences.

By possessing a thorough comprehension of syntax, we are able to understand sentences and their correlation with significance and proficient communication. When words are organized differently, they can produce distinct meanings. However, the same concept can also be expressed through distinct ructures, which is particularly relevant when considering translations from languages. As an example, the sentence "I read the book" in English e subject-verb-object order, however in Japanese, it follows the subjectobject-verb order as "watashi wa hon o yomimasu," which can be roughly translated as "I-(watashi wa) the book-(hon o) read-(yomimasu)". Although conveying the same content, it is evident that the sentence construction differs.

a. Chomsky's View on Syntactic Structures

Chomsky's Syntactic Structures rejects the empirical methodology of structural linguistics and shifts the focus of analysis from observable behavior to the intuitions of native speakers about their language (Winograd, 1983:11). Chomsky argues that the structural analysis of texts or recorded collections of utterances is insufficient to completely comprehend the innovative nature of human language. Speeches are the result of the speaker's cognitive faculties. The most challenging aspect lies in explaining the grammar of a language, which encompasses the implicit understanding of its rules. As a result, it is necessary to provide clarification on people's intuitions about a word or sound sequence in a phrase, as well as the structural connections such as paraphrasing between sentences. The theory must postulate mental structures and processes as the minds of language users are not directly observable. Linguists study languages based on their innate understanding of their own language, rather than strict standards (Winograd, 1983:12).

Chomsky (2002) suggests that linguists have the potential to study the theoretical cognitive frameworks that form the basis of linguistic ability. Competence refers to the abstract evaluation of language proficiency, whereas performance pertains to the actual mechanisms that determine the speaker's utterances and their reception within a specific context. The notion of abstract ability is intricately linked to mathematical proof. Mathematicians clarify the symbols that represent valid claims in a set of formulas, using axioms and standards of reasoning. The equation $(x + 1)^2 = x^2 + 2x + 1$ is valid in basic algebra, although $(x + 2)^2 = x^2 + 4x + 2$ is not (Winograd, 1983:12). Generative linguistics regards language as a mathematical construct and develops theories that imitate mathematical axioms and inference methods. Grammatical statements can be generated from principles, similar to mathematical proofs.

le Syntax Structure

finite state grammar is considered as the simplest type of grammar where et of rules could generate an infinite number of sentences (Chomsky, 2002). The purpose is to find what sort of grammar needed to generate the infinitely available morphemes. The model can be represented by a state diagram below

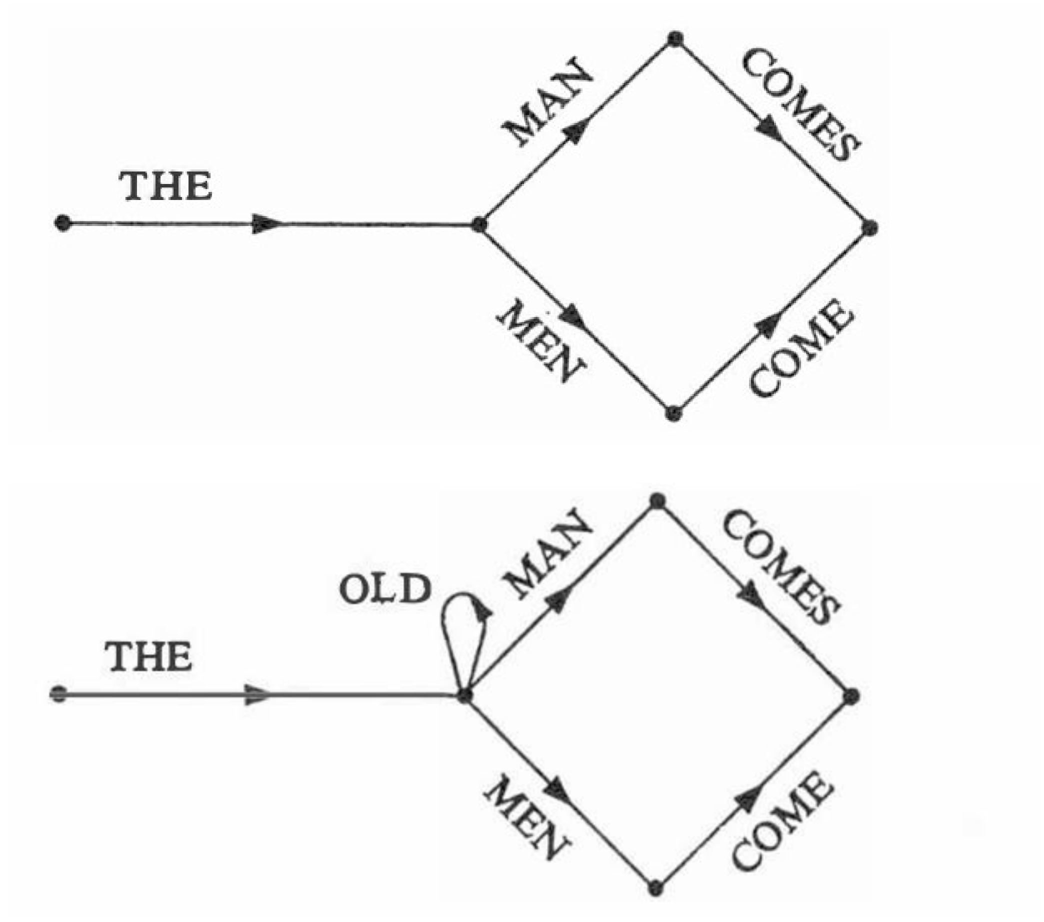

Figure 1 *Finite State Diagram*

The diagram models represent how language is processed in a step-by-step, linear manner. A sentence begins with a fixed starting point, such as "the", and follows a line to form basic structures. The diagram includes two sentences differentiated by the subject and the verb, the subject “man” with the appropriate verb “comes” while “men” is in line with “come”. The second diagram represent a case where it is possible to produce an infinite number of sentences by adding any number of close loop (Chomsky, 2002:19-20).

However, the model is rejected by Chomsky, He argues that language cannot be fully explained by the model, which treats sentences as sequences of nerated step by step. Instead, a more powerful and structured model such e structure grammar is necessary to accurately describe how sentences ed. He discusses how syntax is typically analyzed in linguistics using nt analysis (parsing), which breaks sentences into hierarchical units (phrases, clauses, etc.). The key point is that the grammatical model required for this type of analysis is more complex than the finite-state model.

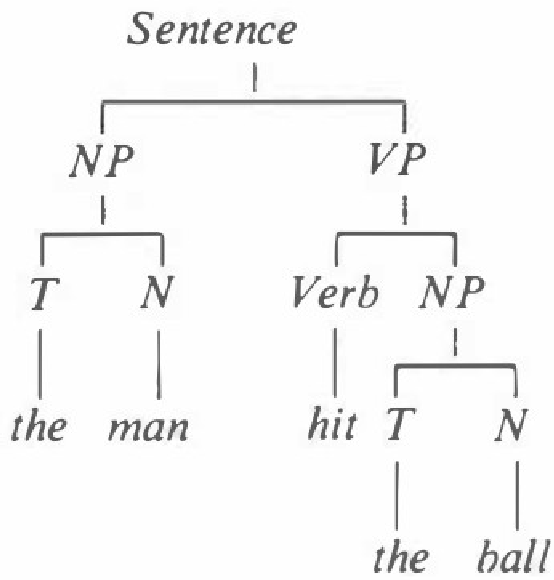

Figure 2 *Phrase Structure Diagram*

Chomsky (2002) explains that this diagram shows how sentence is structured hierarchically rather than as a simple sequence of words. The sentence is broken down into units using phrase structure rules, such as S → NP VP (a sentence consists of a noun phrase and a verb phrase), NP → T N (a noun phrase consists of a determiner and a noun), and VP → Verb NP (a verb phrase consists of a verb followed by another noun phrase). This allows for a more accurate representation of syntax compared to the Finite-State Model, which processes sentences in a strictly linear, left-to-right manner. Unlike finite-state grammar, phrase structure grammar can handle complex sentences. Phrases can be added within other phrases, allowing for an infinite number of grammatical sentences. This makes phrase structure grammar a more powerful and flexible model for analyzing natural language, which becomes the foundation for modern syntactic analysis and generative grammar.

c. Syntactic Complexity

Numerous syntactic complexity indices have been suggested and d in L2 writing studies, but only a limited number have been continuously Kyle, 2016:9). Wolfe-Quintero et al. (1998) proposed a commonly used complexity indices which are Mean length of clause (MLC), Mean length (MLT), Mean length of sentence (MLS), Complex T-units per T-unit

(CLT/T), T-units per sentence (T/S), Clauses per sentence (C/S), Clauses per T-unit (C/T), Dependent clauses per clause (DC/C), Dependent clauses per T-unit (DC/T), Coordinate phrases per clause (CP/C), Coordinate phrases per T-Unit (CP/T), Complex nominals per clause (CN/C), Complex nominal per T-unit (CN/T), Verb phrases per T-unit (VP/T), and Verb phrases per clause (VP/C).

Mean length of clause is measured from the average number of words per clause, and there is no distinction between independent and dependent clauses (Kyle, 2016). Kyle (2016) explained that its value increases because of several factors: (1) It could increase as phrasal coordination and modification increases as well, (2) Proper aspect use (If simple declarative clauses have no auxiliaries, perfect and progressives have one auxiliary, and perfect/progressive combinations have two auxiliary), and (3) Proper syntax structure (SV structure with only two words, and SVO structures with at least three words). Although, there are many reasons for a lengthy clause, If the proficiency level of a writer is high, their clauses length also increases (Knoch, Rouhshad, & Storc, 2014, as cited in Kyle 2016). This is supported by studies that have shown that there is a positive relationship between Mean length of clause and proficiency levels (Cumming et al., 2005; Ortega, 2003; Wolfe-Quintero et al., 1998, as cited in Kyle, 2016)

Mean length of T-unit is measured from any independent clause that has dependent clause attached to it, and it adds an extra level of specificity where dependent clauses are disambiguated (Kyle, 2016:10). For example, the sentence “The professor suggests to abandon this facility, and she insists on burning them immediately” has two independent clauses, so it has two T-units. Compared to the sentence “The professor suggests to abandon this facility because it is dangerous to continue researching here” while lengthy it only has one T-unit. Kyle (2016) also suggest that T-units length increases in conjunction with proficiency.

Mean length of sentence can be counted quickly because it is measured from the number of words in a sentence (Kyle 2016). There is a strong correlation of r = .907 between Mean length of sentence and Mean length of T-unit according to Lu (2010). Studies also shows a positive relationship between Mean length of sentence and Proficiency in Language (Wolfe-Quintero et al., 1998; Ortega, 2003, n Kyle, 2016).

omplex T-units per T-unit can be measured from a T-unit that includes pendent and a dependent clause (Casanave, 1994; Lu, 2011, as cited in

Kyle, 2016). A study from Lu (2011) shows no significant relationship between language development and Complex T-units per T-unit, while Casanave (1994) found positive trend between development and Complex T-units per T-unit with no statistical result to support his findings. However, Kyle (2016) hypothesized that learners might be able to use more combinations of independent/dependent clause, but might unable to determine the number or type of dependent clauses.

Kyle (2016) explained that T-units per sentence calculates the number of (independent) clausal coordination in a text where a score of 1 indicates no clausal coordination in the whole essay, and a score of 2 indicates that every sentence averagely includes one instance of clausal coordination. A study by Monroe (1975) who investigated French as the L2 language instead of English found a significant negative relationship between T-units per sentence and language proficiency. Clausal coordination decreased as proficiency decreased (Kyle, 2016).

The amount of clausal coordination and subordination in each sentence is measured to find the number of Clauses per sentence (Kyle, 2016). He explained that this index needs to be explored further because of these two studies: 1. A study by Ishikawa (1995) who found a positive relationship between Clauses per sentence and language development within three months. 2. A study by Lu (2011) who found a negative relationship between Clauses per sentence and school year. There is a clear contrast between these two findings, so the accuracy of findings from this index are still questionable. Kyle (2016) also suggest that this index needs more attention.

Clauses per T-unit is measured from the number of clausal subordinations found in a text, but there is no distinction made between types of subordination (Kyle, 2016). According to Kyle (2016) eighteen studies that investigated relationship between Clauses per T-units and language proficiency have been reviewed by Wolfe-Quintero et al. (1998); from six studies, there is a positive relationship found, one study found a significant negative relationship, while the rest of 11 studies, does not found any significant relationship between them. Recent studies by Cumming et al. (2005), Knoch (2014), and Lu (2011) also found no significant relationship.

yle (2016) suggests that Dependent clauses per clause have similarities uses per T-unit since both of them measures the amount of clausal ation in a text. The difference is that this index investigated the average number of dependent clauses found within total clauses in a sentence. There is a negative relationship between this index and 2-4 years of school level which suggest that as language proficiency increases, writers would use fewer dependent clauses (Lu, 2011, as cited in Kyle 2016).

Dependent clauses per T-unit index is measured as the name suggested, it counts the average number of dependent clauses found within one T-unit. A study shows a correlation of *r* = .922 between Dependent clauses per clauses and Dependent clauses per T-unit (Lu, 2010, as cited in Kyle 2016). Studies of this index shows a varied result: 1. A study by Homburg (1984) investigating Dependent clauses per T-unit and proficiency results in a positive relationship. 2. A study by Lu (2011) results in a negative relationship. 3. A study by Vann (1979) who reviewed two studies found no significant relationship.

Coordinate phrases per clause is measured from the number of phrasal coordination in a text (Kyle 2016) or more specifically in a clause. A study suggest that phrasal coordination increases as language learners develop, which is supported by a positive relationship found between this index and proficiency levels (years 1-3 and 1-4) (Lu, 2011, as cited in Kyle 2016).

Coordinate phrases per T-unit index also measures the number of phrasal coordination but it is measured per T-unit. Lu (2010) found a correlation of r = .945 between Coordinate phrases per T-unit and Coordinate phrases per clause.

Complex nominals per clause is measured from a variety of syntactic constructions: nominal clauses, infinitive or gerund as subject, nouns combined with adjectives, adjective clauses, appositives, prepositional phrases, and/or possessives (Cooper, 1976; Lu, 2011, as cited in Kyle, 2016). A further study by Lu (2011) found a positive relationship between all levels except years 3-4 and this index.

Complex nominals per T-unit is measured from the same syntactic constructions as the previous index per T-unit, both are correlated with *r* = .867 (Lu 2011, as cited in Kyle 2016). This index is considered inferior to the previous index because it creates an irrelevant distinction between coordination and subordination uintero et al., 1998, as cited in Kyle 2016). This idea is based from Lu ata where the use of complex nominals per T-unit showed a distinction the first year and years 2-4, but not for any other consecutive levels. On the other hand, the use of complex nominals per sentence distinguished between all levels except year 3-4.

Verb phrases per T-unit measures the total amount of verb phrases in one T-unit which includes finite and non-finite verbs. Kyle (2016) explained that only one study (Lu, 2011) within his knowledge that investigated Verb phrases per T-unit with proficiency which shows no relationship. Another Similar index is Verb phrases per clause with clause as its denominator instead of T-unit (Kyle 2016).

Below is the summary table for the 14 syntactic Complexity indices:

Table 1 *Syntactic Complexity Measures*

Measure	Code	Definition
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2: Sentence Complexity		
Sentence complexity ratio	C/S	# of clauses / # of sentences
Type 3 Subordination		
T-unit complexity ratio	C/T	# of clauses / # of T-units
Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4: Coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
Sentence coordination ratio	T/S	# of T-units / # of sentences
Type 5: Particular structures		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

Source: Lu, X. (2010). Automatic Analysis of Syntactic Complexity in Second Language Writing. *International Journal of Corpus Linguistics, 15*(4), 474-496. doi:https://doi.org/10.1075/ijcl.15.4.02lu

2.2.2 Second Language Writings

Second language (L2) writing refers to the process of creating texts in a language that is not a person's first language. It is often a complex task that goes ... simply translating thoughts from the native language into the second ... . L2 writers are expected to produce grammatically and syntactically ...xt while also demonstrating a wide range of vocabulary.

In general, L2 writers often experience both cognitive and linguistic difficulties. These include errors in grammar, word choice, and sentence structure, as well as trouble expressing complex ideas clearly and effectively. Their writing is often influenced by direct translation from their native language, which can lead to mistakes. These issues may affect the fluency and coherence of their writing, making it harder for them to write at the same level as native speakers.

a. Relationship with Syntax

Crossley (2020) describes the features of proficient writers based on findings from various studies. According to Myhill (2008, as cited in Crossley, 2020), skilled writers tend to produce texts with fewer finite verbs, subordinate clauses, and coordinated clauses. They also often show greater syntactic complexity and include more words before the main verb (McNamara et al., 2010, as cited in Crossley, 2020). Crossley also points out that lower-quality essays are usually linked to shorter sentence structures, while longer noun phrases and more pre-verbal elements tend to be associated with better writing (Crossley et al., 2011; McNamara et al., 2013).

The connection between syntactic complexity and writing quality is especially noticeable in L2 writing, with many studies showing consistent results (Crossley, 2020:9). In general, more proficient L2 writers produce texts with longer and more varied syntactic forms, often measured through T-units (Lu, 2011; Ortega, 2003; Wolfe–Quintero, Inagaki, & Kim, 1998, as cited in Crossley, 2020). Beyond T-units, research has shown that essays by higher-level L2 writers frequently include more clausal subordination and passive constructions (Grant & Ginther, 2000; Connor, 1990; Ferris, 1994, as cited in Crossley, 2020). More recent studies have also found that improved L2 writing tends to feature more advanced syntactic structures, such as dependent clauses, infinitives, and 'that'-clauses, all of which increase clause complexity (Crossley & McNamara, 2014; Friginal & Weigle, 2014, as cited in Crossley, 2020). As writing skills improve, phrasal complexity also increases (Taguchi, Crawford, & Wetzel, 2013, as cited in Crossley, 2020).

rossley (2020) also presents both cross-sectional and longitudinal on syntactic development in L2 writers: (1) Larsen-Freeman (1978) found ercentage of error-free T-units and their length were strong indicators of iency, (2) Ferris (1994) observed that higher-level L2 writers produced more complex structures, such as passives, nominalizations, and relative clauses, compared to lower-level writers, (3) Ortega (2003) showed that low- and high-proficiency L2 writers differed in their T-unit features, like average clause length and number of clauses per T-unit, (4) Lu (2011) confirmed that most T-unit features significantly vary across different levels of proficiency, (5) Casanave (1994) and Ishikawa (1995) found through longitudinal studies that L2 writers improve over time in their use of syntactic structures, including gains in length and accuracy, (6) Similar patterns were also seen in short-term studies, where increases in average words per T-unit and in the percentage of error-free T-units were signs of progress (Stockwell & Harrington, 2003; Crossley & McNamara, 2014).

b. Relationship with Lexical Quality

Crossley (2020) stated that lexical quality is often measured through lexical diversity, which refers to how many unique words are used, lexical density, which looks at the balance between content and function words, and lexical sophistication. These aspects are commonly used to evaluate the overall quality of a text. Among these, lexical sophistication is seen as the most informative metric for evaluating text quality. It refers to the proportion of advanced or complex words used in a text (Read, 2000, as cited in Crossley, 2020). Crossley (2020) noted that the definition of lexical sophistication has changed over time, as it can include many different word features. For example: (1) Laufer & Nation (1995) define it as the use of low-frequency words, (2) Coxhead (2000) describes it as words more commonly found in academic texts, (3) other researchers see them as words that are less concrete, imageable, and familiar (Crossley & Skalicky, in press; Salsbury, Crossley, & McNamara, 2011; Saito et al., 2016), (4) Balota et al. (2007) describe them as words with fewer phonological and orthographic neighbors, and which take longer to recognize in tasks like word naming or lexical decision, (5) Fellbaum (1998) sees them as more specific words, and (6) McDonald & Shillcock (2001) consider them as words used in more limited contexts.

Crossley (2020) also explained that the presence of sophisticated words in a text shows strong lexical knowledge and is linked to better writing performance. n on L1 writing has shown that high-quality academic texts tend to use ademic words (Douglas, 2013), more specific and imageable words, as ords that are less familiar or meaningful (McNamara et al., 2013), longer nd less familiar vocabulary (Crossley, Weston, McLain, & McNamara, 2011), and more infrequent words (McNamara, Crossley, & McCarthy, 2010). For L2 writing, Crossley (2020) also shared similar findings: L2 writers tend to use less frequent, familiar, and meaningful words (Crossley & McNamara, 2012), longer words with more letters or syllables (Grant & Ginther, 2000; Reppen, 1994), more specific words (Guo, Crossley, & McNamara, 2013; Kyle & Crossley, 2016), and less imageable vocabulary (Crossley, Kyle, Allen, Guo, & McNamara, 2014).

c. Relationship with Cohesion

Text cohesion is a crucial component of writing because it can reveal lexical, semantic, and argumentative relationships within a text. It is associated with the interconnection of text segments based on textual properties (Halliday & Hasan, 1976 as cited in Crossley, 2020). According to Crossley (2020) there is a common method in identifying cohesion which is to examine the connections between text segments that includes referencing previous elements (using pronouns), repeating lexical items, substituting lexical items, and using conjunctions to connect ideas. Coherence is based from text and make reference to the existing or absence of explicit cues in the text that facilitate connecting segments of text together (Crossley, 2020).

According to Crossley (2020) though rare, there are studies that investigate the relationship between text quality and cohesion features in L2 writings: (1) Adults L2 writers have similarities with L1 college level writers where proficient L2 writers have a tendency to produce less cohesive text as measured by lexical and semantic overlap across sentences (Crossley & McNamara, 2010; Engber, 1995; Grant & Ginther, 2000; Jarvis, 2002; Reppen, 1994)., (2) Proficient L2 writers produce text with less lexical overlap (Berman & Verhoeven, 2002; Crossley & McNamara, 2012; Engber, 1995; Grant & Ginther, 2000; Jarvis, 2002; Reppen, 1994)., (3) One study shows that proficient L2 writers produce more connectives (Jin, 2001; Connor, 1990)., (4) In contrast, a study found that high quality L2 text did not contain more connectives (Crossley & McNamara, 2012)., (5) Recent research indicates that sentence-overlapping pronouns and coordination conjunctions, but not function word overlap at the sentence level, are adversely d with essay quality in the context of local cohesiveness features. lly, there is a significant correlation between essay quality and two global features: nearby overlap within paragraphs for both function terms and rossley & McNamara, 2012).

2.2.3 AI-Generated texts

Natural language processing is a field that supports the development of artificial intelligence. It makes it possible to build advanced models that can produce text similar to how humans write. These models are usually trained on large datasets, allowing them to learn language patterns, grammar, and context so they can generate text that is both coherent and appropriate.

One of the main strengths of AI-generated text is its speed and consistency, often outperforming humans in productivity. It can produce a large volume of text much faster than a person and can be used for various purposes, such as content creation, summarization, and translation. These models can also adjust their output based on the style, tone, and complexity requested in the prompt.

AI text generation has practical uses across many fields, including journalism, marketing, and customer service. It can help draft articles, create personalized content, and support language learning. Its potential applications are broad and can cover almost any area that involves digital communication. In the future, the use of AI in text generation looks promising, with expected improvements in how it understands context, creates better content, and works together with human writers.

Even so, there are still challenges. These models are highly skilled at recognizing patterns, but they do not truly understand the content they produce. AI is strong in tasks that involve structure and rules, but it tends to struggle with things that require creativity, critical thinking, or deeper awareness of context.

a. Current AI models

According to Hagos, Battle, and Rawat (2024), Generative AI is a kind of artificial intelligence designed to produce new content that closely resembles the data it was trained on. These models work by recognizing patterns and structures in their training data, allowing them to generate outputs that are different but still similar. Generative AI is a powerful tool that can be used to create text, images, and other types of media through different approaches. Since the models are n large datasets, they can imitate the features of the content they learn eir main goal is to understand how the data is distributed so they can results that match it.

Modern Natural Language Processing (NLP) applications depend heavily on language models, which are the core of many AI tools (Hagos, Battle, & Rawat, 2024). One of the most widely used model types is the transformer. They explained that the transformer model uses the concept of attention, which helps it focus on the most important parts of the input when making predictions. Unlike traditional neural networks, transformers support parallel processing, which improves both training speed and accuracy. Because they can manage different text lengths and keep track of long-range dependencies, transformers are commonly used. They have also helped improve areas like machine translation by making the process faster and more accurate than older techniques.

Hagos, Battle, and Rawat (2024) also stated that using the transformer architecture makes it easier to build Large Language Models (LLMs) such as GPT. These models are trained on massive datasets, including text and code, which helps them understand and generate language that sounds natural and human-like. LLMs have brought big progress in Natural Language Generation (NLG) and Natural Language Understanding (NLU). Their abilities go beyond generating text to handling tasks like answering questions, analyzing sentiment, summarizing information, and translating languages. Thanks to their large training data and advanced learning methods, LLMs are more powerful and flexible compared to traditional language models.

Table 2 *A list of LLMs for a wide range of NLP tasks*

Year of Release	**LLMs**	**Number of Parameters**	**Number of Training Tokens**	**Learning Rate (Default)**	**Developer**
2017	Transformer	530 million	Not explicitly stated	$1x10^{-3}$	Google AI
2018	BERT	340 million	250 billion	$5x10^{-5}$	Google AI
2019	GPT-2	1.5 billion	40 billion	$1x10^{-5}$	OpenAI
2020	T5	11 billion	1 trillion	$5x10^{-5}$	Google AI
2020	GPT-3	175 billion	300 billion	$6x10^{-5}$	OpenAI
2020	Gopher	280 billion	300 billion	$4x10^{-5}$	Google AI
[illegible]	Jurassic-1 Jumbo	178 billion	300 billion	$6x10^{-5}$	AI21 Labs
[illegible]	Megatron-Turing NLG	530 billion	270 billion	$5x10^{-5}$	NVIDIA
[illegible]	Chinchilla	70 billion	1.4 trillion	$1.25x10^{-4}$	Deep Mind

2022	LaMDA	137 billion	768 billion	Not explicitly stated	Google AI
2022	GPT-3.5 (InstructGPT)	175 billion	Not explicitly stated	Not explicitly stated	OpenAI
2022	GPT-3.5 (ChatGPT)	175 billion	Not explicitly stated	$5x10^{-5}$	OpenAI
2022	PaLM	540 billion	780 billion	Not explicitly stated	Google AI
2023	LLaMa	65 billion	1.4 trillion	$1.5x10^{-4}$	Meta AI
2023	Llama 2	2 trillion	2 trillion	$1.5x10^{-4}$	Meta AI
2023	PaLM 2	340 billion	3.6 trillion	Not explicitly stated	Google AI
2023	GPT-4	1-1.76 trillion	Not explicitly stated	Not explicitly stated	OpenAI
2023	Gemini	Not explicitly stated	Not explicitly stated	Not explicitly stated	Google AI

Source: Hagos, D. H., Battle, R., & Rawat, D. B. (2024). Recent Advances in Generative AI and Large Language Models: Current Status, Challenges, and Perspectives. *IEEE Transactions on Artificial Intelligence.*

Hagos, Battle, and Rawat (2024) also explained several applications of Large Language Models (LLMs). First, LLMs are capable of understanding language. Interpreting and making sense of human language is a key function of LLMs, especially within the field of Natural Language Understanding (NLU). These models are widely used in many NLU tasks, such as identifying named entities and analyzing sentiment. LLMs play an important role in tasks that require a deeper understanding of text by processing and interpreting the context (Hagos, Battle, & Rawat, 2024).

Second, LLMs are useful for machine translation. They support the automatic translation of text across different languages. A well-known example is Google Translate, which can easily translate documents, web pages, and text usi LLMs. This ability is made possible through training on large multilingual showing how LLMs help reduce language barriers and support global cation. Their ability to handle and produce natural language is what eir translations both accurate and fluent (Hagos, Battle, & Rawat, 2024).

Third, LLMs are effective in question answering. These models can generate relevant and informative answers across a wide range of topics. This ability is commonly used in information retrieval systems, learning platforms, and virtual assistants. For example, Google's AI assistant shows how LLMs can provide accurate, context-aware answers on subjects such as science, history, and current events (Hagos, Battle, & Rawat, 2024).

Lastly, LLMs are capable of summarizing text. They are useful for generating summaries of news articles and documents. One study introduced a sequence-to-sequence pre-training model that showed strong performance in abstractive summarization. With their advanced NLP abilities, today's LLMs can understand a document's context and produce clear, well-structured summaries while keeping the original meaning intact (Hagos, Battle, & Rawat, 2024).

b. Ethical Concerns and disadvantages

AI has raised several concerns, especially in academic ethics. Because of this, some research publishers have introduced policies on the use of AI-generated text (Sinclair B.J., 2023:1), which include: (1) only using generative AI to improve language and readability, (2) using it under human supervision, (3) only crediting human authors, and (4) clearly stating the use of AI-generated content in a declaration included in the final publication. Using tools like ChatGPT is considered unethical and can be seen as a form of plagiarism, as it involves presenting someone else's work as one's own (Sinclair B.J., 2023:1). This is supported by Sinclair's own test, where he asked ChatGPT to summarize concerns and benefits of using ChatGPT in academic writing. The model was able to generate a summary, but its content was based on information developed by Elsevier and other publishers (Sinclair B.J., 2023:2). This shows that while limited use of AI is allowed, relying on it for content ideas is discouraged.

Poland and Kennedy (2023) also highlight several risks and weaknesses of using AI: (1) Even though AI can generate written content, it cannot ensure proper context, scientific accuracy, or validity, (2) It may produce fake references, (3) Mistakes in technical or scientific writing can be difficult to detect unless the reader [illegible]cient expertise, (4) There is a "halo effect" where people wrongly assume [illegible]ays correct and fully understands the subject, and (5) Relying too much on AI could weaken researchers' essential skills like reading, analyzing, and synthesizing academic literature.

Viglia, as cited in Dwivedi et al. (2023), warns that AI can negatively impact both students and teachers. Students, who already face issues like short attention span (Trinidad, 2020) and declining literacy (O'Connor, 2021), may become even more passive if they use AI excessively. On the other hand, teachers can only distinguish between student work and AI-generated content if they are trained in critical thinking and knowledgeable in the subject. Dwivedi et al. (2023:25) conclude that letting AI do everything can weaken creativity and critical thinking.

Barlette, also in Dwivedi et al. (2023), discusses several limitations of AI. While it can produce clean, well-structured, and grammatically correct writing, the result often feels bland and lacks personality (Whitford, 2022, as cited in Dwivedi et al., 2023). Barlette also notes that if multiple students use similar prompts, the outputs may turn out very similar. He adds that institutions may not be fully ready to handle cases where students use AI for their assignments.

In particular, institutions often lack proper systems to detect and discipline AI usage. According to Barlette, current plagiarism detection tools are not effective against AI because the generated content is too different from any source to be flagged as copied (Dwivedi et al., 2023:27). This creates an unfair situation where students who are not caught receive no consequences. In addition, some institutions do not yet classify AI-generated work as academic misconduct. In other words, because regulations about AI use are still unclear or undeveloped, its use is not yet considered illegal from a policy standpoint.

c. Potential benefits

Although AI has sparked many controversies, it also offers clear benefits. Sinclair B.J. (2023) points out that using AI can help him concentrate more on critical and creative thinking, as the writing itself is handled by the AI. He also mentions that ChatGPT can support writers by helping improve scientific writing, which is especially helpful for non-native English speakers.

...or non-native speakers, writing in English comes at a cost (Woolston & ...019, as cited in Sinclair B.J., 2023). It requires a lot of time, effort, and ...es money for professional editing services. Sinclair adds that by using ..., non-native scientists can focus on developing their ideas in point form and let the AI expand them into full paragraphs. This makes the writing process easier, with grammatically and syntactically correct results.

Poland and Kennedy (2023) also mention the benefits of using AI. It removes the need for time-consuming literature searches, helps summarize findings, and assists in drafting based on those sources. It also improves grammar, punctuation, and overall readability. In addition, AI can do all of this almost instantly, making the process more efficient and cost-effective. This helps reduce the time and resources needed to prepare academic manuscripts.

2.3 Conceptual Framework

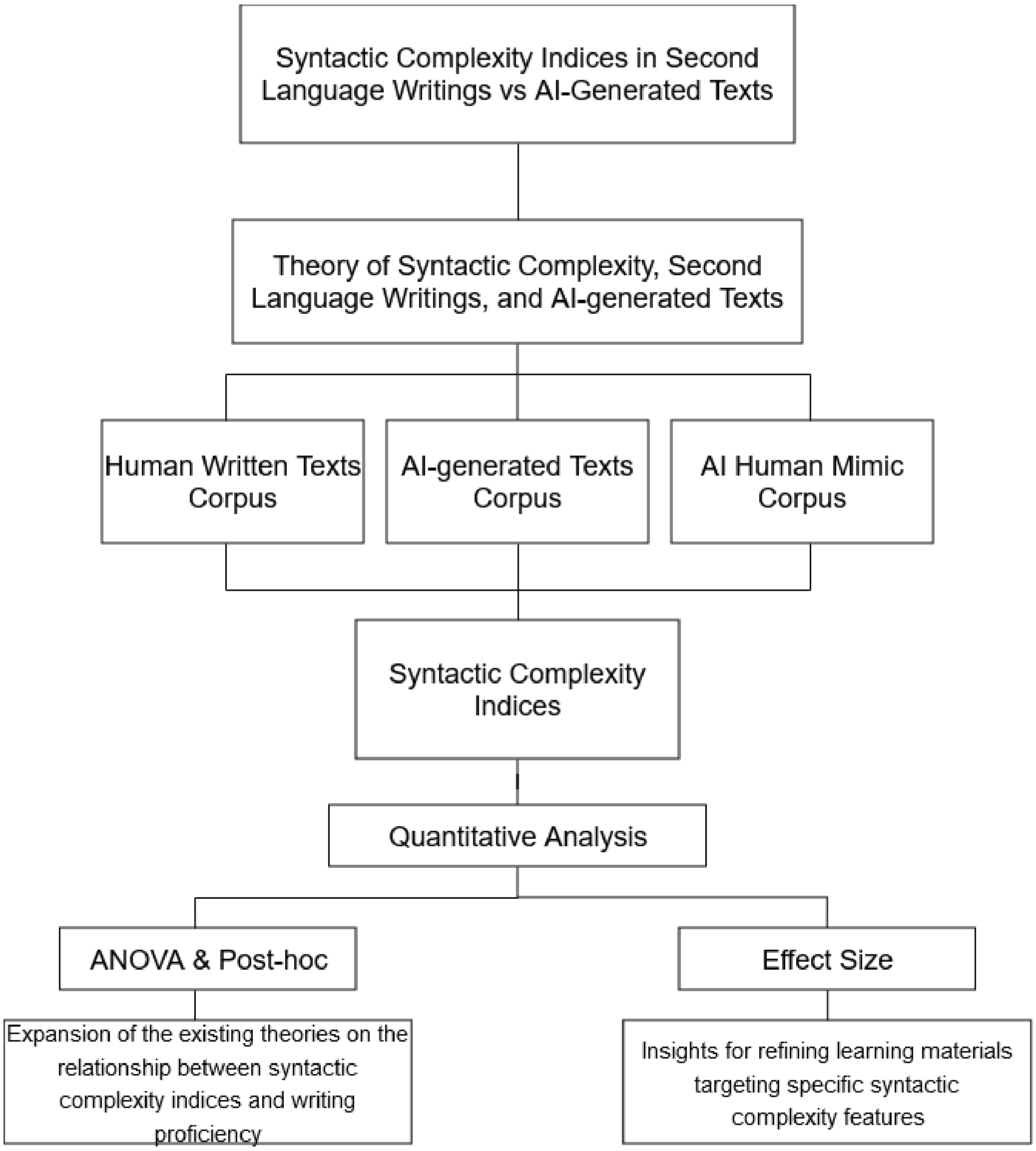

***Figure 3** Conceptual Framework*

As mentioned before, this study focuses on syntactic complexity indices found in the three corpora. The selected syntactic complexity indices are MLC, MLS, MLT, C/S, C/T, CT/T, DC/C, DC/T, CP/C, CP/T, T/S, CN/C, CN/T, VP/T. By investigating these 14 indices, this study aims to expand on existing theories regarding their relationship with writing proficiency using ANOVA and post-hoc result. The effect size result is used to provide insights for writing instructions to proficiency.