A Prelude to Determine 'Datum Point' for MA Writing in English: Comparing Syntactic Complexity of Inner-circles and Expanding-circles

Arab Journal of Applied Linguistics
e-ISSN 2490-4198
Vol. 06, No. 02, 2021, 31 – 70

http://www.arjals.com

Dr. Cuneyt DEMIR, ELT Department, Faculty of Education, The University of Siirt, Turkey

Abstract

Writing native-like has always been a primary purpose for non-native writers of English and, accordingly, many language components have been investigated to develop writers' fluency. One of these language components is syntactic complexity (SC), which is often regarded as a reliable way to grade any texts from easy to difficult. This study aims to compare native and non-native writers' fluency by measuring their SC and to create a reference SC point for non-native writers of English. To achieve this, the study comprised two groups: Group 1 was composed of native speakers from Inner-circle countries, namely USA, England, Canada, and Australia, and Group 2 was composed of non-native speakers from Expanding countries, namely Turkey, Saudi Arabia, Russia, and China. The data were composed of 200 MA dissertations in ELT equally collected from countries in Group 1 and Group 2. SC realisations were divided into 14 sub-categories. The 14 sub-categories constituted a taxonomy for SC and they were analysed through ANOVA, Kruskal-Wallis, Independent sample t-test, and Mann-Whitney tests. The results showed that there is a statistically significant difference in 9 categories in favour of Group 1 and that the average SC scores of Group1 are significantly high compared to Group 2. The study suggests that writers of MA dissertations pay particular attention to SC, if native-like fluency is sought because it was concluded that native writers of MA dissertations are prone to write with higher SC level when compared to non-native writers of MA dissertations.

Keywords: Syntactic complexity, academic writing, native writer, non-native writer

1. Introduction

The process of building a sentence requires a range of cognitive processes occurring either naturally (as in a native speaker's mind) or systematically (as with a non-native speaker) and these cognitive processes turn out language production in varying complexity levels. Linguistic Complexity is an indicator of overall development in Second Language Learning (SLL) and it determines that linguistic complexity may show variance depending on your stance on the issue; for example, Droop and Verhoeven (1998) correlated linguistic complexity and background knowledge while Gibson (1998) excluded any chance of positive correlation between lexical and linguistic complexity. The common sense as to linguistic complexity accumulates over what determines the complexity level (Klebanov & Flor, 2018) and how to measure its internal structure (c.f. Newmeyer & Preston, 2014).

Out of many linguistic factors determining text complexity, syntactic complexity (SC), i.e. sophistication degree of structures in discourse, is one of the linguistic subdomains that is frequently regarded as a reliable way to categorize the spoken or written productions. Accordingly, a growing body of research involves the measuring of SC to determine overall complexity level of particularly written works because different from written discourse there are many other linguistic and extra-linguistic factors that may convolute the processing of speech (Roll, Frid, & Horne, 2007). Therefore, compared to myriad factors affecting the complexity of spoken discourse, the measuring of SC provides much more elaborate and accurate results concerning linguistic maturity of writing. On this account, SC can function as the capacities of the language resources deployed by the writer to bolster the claims, which is particularly of importance for authors who write in English for Academic Purposes (EAP).

Building up texts with high SC necessitates linguistic competence and having such a linguistic competence may not come within the purview of EFL writers largely because they lack metalinguistic judgments (Walters & Wolf, 1996). To clarify, whereas native writers (NW) inherently have an upper hand in terms of SC thanks to their regular use of language even in the simplest daily work, non-native writers (NNW) lack this

opportunity. This insufficiency in the second language inevitably leaves a trace on their language production, particularly in contexts where a standard level of SC is needed. Authors in EAP are expected to reach a reasonable level of SC in writing. This is neither to say that writers need to have a degree of maximum SC nor to have a minimum SC, but a decent level that does not disrupt the flow of reading fluency. Concisely, it does seem that writing for EAP does not make a high SC level obligatory (see Ortega, 2003, for the relationship between SC and L2 proficiency) but underlines the importance of striking a balance.

1.1. Three Concentric Circle Model

When first introduced in 1978 to examine the regional concepts of English globally, the term *World Englishes* was probably not contemplated to earn so much attraction from researchers. Although the primary purpose was to investigate indigenized and localized varieties of English in the UK and the USA, the term crossed its borderlines and became one of the fundamentals to sociolinguistics studies. The term that was once for intra-national studies investigating regional dialects and accents by sociolinguists has started to be employed largely for international research to linguistically compare and contrast the differences of Englishes in countries where English is of concern. Meanwhile, the term needed to lay its theoretical structure on a concrete ground so that studies would be able to produce more objective and evidence-based conclusions. A TESOL conference in 1988 in Hawaii made inroads for the term to be used in Applied Linguistics and subsequently, Kachru (1992) proposed the *Three Concentric Circle Model* that would help researchers substantially categorize the countries where English was gaining ground. This model coined three terms as Inner, Outer, and Expanding-circles and placed countries according to the function of English. In other words, Inner-circle represents the Anglophonic countries where English is the primary official language (e.g. the UK, the USA, Australia) while outer-circle involves countries where English is not the native language but is rather crucial for historical reasons or other functional reasons such as its being the language of national institutions (e.g. India, Nigeria, Bangladesh). The third circle, Expanding, comprises countries where English has no governmental or historical

roles, but nevertheless English is used as a lingua franca that bears importance for better global communication or trade (e.g. Russia, Turkey, countries in the Arab League). One caveat to note is that although English holds limited purposes in Expanding-circle, the role of it is expanding enormously in those communities as it is understood from the name of the circle. The global spread of English is categorized as First Dispersal (the transfer of English to the new world) and Second Dispersal (the transfer of English to Asia and Africa); however, it seems that the third dispersal of English should be named for the Expanding countries where English comes into prominence at a rate of sweeping force.

1.2. Purpose and Research Questions

Writing in EAP necessitates a degree of SC that will neither disrupt reading fluency nor will it give an impression of oversimplification, i.e. a reasonable degree of SC. However, there is not a consensus in the literature on what constitutes a reasonable SC. Many studies simply make a comparison between NW and NNW to provide a general description of the language proficiency or they examine the correlation between SC and fluency in writing, but there remains a residue concerning to what extent an author of EAP, namely authors of MA dissertations in this study, need have SC in their writing; whether the SC level of their dissertation is sufficient; or simply whether there should be a concern about SC. Keeping these in mind, this study aims to create a reference SC point for MA dissertations in ELT. To achieve this, one hundred MA dissertations from four Inner-circle countries (USA, England, Canada, and Australia) were analysed through an index of SC on computer. The average of the dissertations was roughly considered to be the reference SC point for MA dissertations in ELT. Furthermore, the present study aimed to compare SC levels of MA dissertations of Inner-circles with Expanding-circles, namely Turkey, Saudi Arabia, Russia, and China to reveal whether there is a statistically significant difference between them in terms of SC in MA dissertation writing. Thus, this study aims to answer the following research questions:

1. What is the average SC level of Inner-circle and Expanding-circle in writing an MA dissertation in ELT?

2. Is there a statistically significant difference between Inner-circle and Expanding-circle in terms of SC in writing MA dissertation in ELT?

2. Literature Review

This section reviews research on SC complexity and writing, studies of SC among native and non-native EAP writers, and the different means utilized in measuring L2 SC.

2.1. Syntactic complexity and writing

In linguistic theories, syntactic complexity traditionally refers to compound and complex sentences, i.e., clausal complexity. In some linguistic traditions, the notion of syntactic complexity has not extended to phrasal complexity (e.g., Givo´n 2009; Givo´n & Shibatan, 2009; Gökmen, 2020). However, in another view emerging in L1 and L2 developmental studies focusing on syntactic maturity (e.g., Hunt, 1965; Cooper, 1976; Ravid & Berman, 2010; Crossley et al., 2011; Lu, 2011) and discourse analysis of texts in different genres (e.g., Biber, 2006; Biber et al., 2011), phrasal complexity (particularly noun phrase complexity) has been considered an integral part of syntactic complexity. What complicates the construct of syntactic complexity further is that the notion of clause has not been defined consistently across disciplines. Notably, linguistic theories of grammar (Halliday & Matthiessen, 2004; Langacker, 2008; Elcin, 2017) accept both finite and non-finite clauses as clauses. In writing research, however, following Hunt's (1965) definition, the term clause has been predominantly used to refer only to finite clauses. Therefore, when calculating an index such as the number of clauses per sentence as a syntactic complexity measure, any discrepancies in results may arise from the different definitions of the clause adopted. There is not an easy answer regarding which definition of clause is more appropriate, but we adopt the view that both finite clauses and non-finite elements should be examined as part of the construct. However, to maintain consistency with previous writing research, we use the term *clause* only to refer to finite clauses and we use the term *non-finite element* to refer to non-finite clauses.

SC has been a crucial construct in the teaching of L2 writing and learner's growth of syntactic repertoire is critical for L2 development (Ortega, 2003); depending on the SC level of a piece of text, expert readers may deduce enough as to the linguistic level of the author because it is important to measure L2 writing proficiency (Kyle & Crossley, 2018). Therefore, having a certain level of SC is a commonplace requirement for academic writing particularly if it addresses an audience with expertise in the field. However, there is not a standardised complexity level to ensure the suitability between writing and audience, which is why studies that measure SC and establish a link between SC and writing may be of importance to provide a basis for a commonplace guide for writers.

2.2. Comparison of native and non-native writers in academic English writing

Comparing and contrasting studies are largely preferred to see more specific results of two or more target groups, which is why they are widely used by researchers. As in other fields, scholars used comparing and contrasting techniques to investigate syntactic complexities of written or spoken texts and reached a wide range of results. Ai and Lu (2013) compared university students' writing in four areas of SC (length of production unit, amount of subordination, amount of coordination, and degree of phrasal sophistication) and found that NNW had lower SC levels in all four areas of SC when compared to NW. Not all researchers ended up with the superiority of NW in terms of SC; for example, Neff et al. (2004) revealed that NNW had higher levels of subordination complexity in their argumentative than essays of L1 English writers. Similarly, Eckstein and Ferris (2018) compared L1 and L2 groups from many angles and found that the groups did not differ significantly regarding SC. However, all these comparative studies collected data from BA students who wrote argumentative texts or free texts with writing prompts. Lexical diversity, linguistic accuracy, and SC may reflect differences at various points as L2 speakers' proficiency grows in time. Therefore, studies conducted at different education levels may not reveal similar results because SC may vary particularly at higher proficiency levels (Gyllstadt et al., 2014). In other words, the studies in the literature largely include BA students while this study was conducted on written works of MA students. Besides, rhetorical devices are said to have changed

across disciplines (Hyland, 2005); therefore, each scientific field may require a particular discourse, which might affect the writing style and the language. Different from other studies that did not provide contextual information or academic subject in which the research was implemented, the present study provides clear details about its data source: MA Dissertations in ELT.

2.3. Measuring L2 syntactic complexity

Measures of SC, with the increasing importance of second language writing, have caught attention over other linguistic measures and this steered the attention from crude and subjective measures to valid and objective developmental gauging procedures. All these strands of progressions aimed to neutrally measure the linguistic development of speakers so that the connection between proficiency and syntactic development could be elicited. Then on, cross-linguistic and comparative studies to measure SC proliferated in the literature; however, the question was how to measure SC objectively because it seemed that there was not a common consensus on how to reliably calculate SC in writing. Lu (2010) stated that very early SC measuring studies were labour-intensive because they would be based on node-counting algorithms that calculate the number of nodes in the phrase markers of syntactic constructions (e.g. Frazier, 1985; Yngve, 1960), and then new mathematical formulas were devised to calculate SC, which partly facilitated labour-intensive workload of calculations. These formulas were largely operating on length-gauging of production units such as morphemes (Miller & Chapman, 1981), clauses (Beers & Nagy, 2009) and sentences (Buhr & Zebrowski, 2009). Later on, computer-based indices, which are much more fine-grained when compared to labour-intensive indices, emerged thanks to the technological developments and they alleviated the labour-intensive workload of manual analyses that would take hours or even days depending on the amount of data.

It was in the 1970s that SLA witnessed the first usages of T-units which are used to measure the syntactic maturity of learners (Larsen-Freeman & Strom, 1977). Though

initially employed to evaluate writing development in L1, the T-units were later accommodated for L2 research. Having been used by SLA researchers extensively (e.g. O'Donnell, 1976), the efficiency of T-units to measure SC in writing began to be questioned by some researchers (e.g. Gaies, 1980; Lantolf, 1988) because it was thought to be focussed mainly on syntax, hence fails to deal with morphological errors of the learners (Barnwell, 1988). Another question as to its use in L2 writing was that the clausal subordinations gauged by T-units are largely valid for spoken discourse such as conversations whereas the characterization of academic writing is based on syntactical units such as the use of noun phrase constituents and complex clauses (Biber, Gray, & Poonpon, 2011). Although T-units are still used in the SC studies, new models aiming to eliminate the infelicities of T-units appeared, one of which is the Taxonomic Model by Bulte and Housen (2012) that provides a detailed description of L2 complexity. It firstly divides L2 complexity into two as Relative Complexity and Absolute Complexity, and then further divides Absolute Complexity till reaching Structure Complexity and categorizing it into four categories: Lexical, Morphological, Syntactic, and Phonological. Syntactic categorization mainly involves sentences, clauses, and phrases. Accordingly, studies largely base their ramifications on the measuring length of sentences, clauses, and phrases. In line with this, literature proposes a large variety of measures to categorize SC in second language writing. The majority of these measures gather around quantifying some units in writing; as shown in table 1 (see Wolfe-Quintero et al., 1998 for a comprehensive review).

Table 1.

Main categorization of language units for measuring SC.

Unit	Measure	Definition
Length of Production	1. Mean length of sentence (MLS)	Number of words divided by number of sentences
	2. Mean length of T-unit* (MLT)	Number of words divided by number of T-units
	3. Mean length of clause (MLC)	Number of words divided by number of clauses
Sentence Complexity	4. Clauses per sentence (C/S)	Number of clauses divided by number of sentences
Phrases	5. Verb phrases per T-unit (VP/T)	Number of verb phrases divided by number of T-units
Subordination (embedding)	6. Clauses per T-unit (C/T):	Number of clauses divided by number of T-units
	7. Dependent clauses per clause (DC/C)	Number of dependent clauses divided by number of clauses
	8. Dependent clauses per T-unit (DC/T)	Number of dependent clauses divided by number of T-units
Coordination	9. T -units per sentence (T/S)	Number of T-units divided by number of sentences
	10.Complex T-units ratio (CT/T)	Number of complex T-units divided by number of T-units
	11.Coordinate phrases per T-unit (CP/T)	Number of coordinate phrases divided by number of T-units
	12.Coordinate phrases per clause (CP/C)	Number of coordinate phrases divided by number of clauses
Nominals	13. Complex nominals per T-unit (CN/T)	Number of complex nominals divided by number of T-units Number of complex nominals divided by number of clauses
	14. Complex nominals per clause (CN/C)	

*T-unit as defined by Hunt (1965, p. 20).

Although T-unit is not the most promising measuring method, it still holds in studies aiming to investigate SC, particularly for the ones focussing more on accuracy than simply on complexity (which is one of the criticisms against the T-units indices). Other indices that are not specifically based on T-units are also used commonly to investigate SC in L2 writing. These indices are computer-based and measure the frequency of

clauses and phrases used, the length of syntactic structures, the range and types of phrasal units produced, the types and incidence of embedding, and the types and number of coordination between clauses (Ortega, 2003, as cited in Crossley & McNamara, 2014). For such automatic computational tools of indices, we can cite Coh-Metrix (Graesser et al., 2004), Crossley's and McNamara's SC Index (2014), L2 SC Analyzer (Lu, 2010), and Tool for Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) (Kyle, 2014). Each of these indices appears to have some setbacks when compared; for example, while one does not measure the sophistication of syntactic forms, the other entirely concentrates on phrase complexity (see Kyle, 2014 for comprehensive comparisons of SC indices). In a nutshell, various indices are widespread in the literature; therefore, researchers are not confined to a single index to measure SC.

3. Methodology

3.1. Corpus

This study constructed a corpus of over 1 million words taken from a total of 200 MA dissertations in ELT: one hundred MA dissertations from Inner-circles and one hundred from Expanding-circles. Each dissertation had a different number of words and this may disrupt the reliability of the data because the length of a text may affect the results while calculating SC levels. Therefore, the first 5000 words were picked from each dissertation (the number may slightly change depending on the last sentence) and a corpus of 500.275 words was built for Inner and 500.975 words for Expanding-circles. The corpora for all circles excluded direct citations because they did not belong to the authors. The data were collected from the dissertations published in the last 5 years (2016-20). The Inner-circle is composed of countries whose native tongue is English, namely the USA, England, Canada, and Australia while Expanding-circles include Turkey, Saudi Arabia, Russia, and China. The whole data was obtained from the dissertations written in English i.e., 25 dissertations from each country, but the number of universities showed a slight variance from country to country: 45 universities in total (Table 1).

Table 2.

The distribution of the corpus

Country	Circle	No. of dissertation	Average No. of words	No. of university
USA		25	5000	7
England	**Inner**	25	5007	5
Canada		25	5003	6
Australia		25	5001	5
Total		**100**	**500275**	**23**
Turkey		25	5009	8
Saudi Arabia	**Expanding**	25	5010	4
Russia		25	5015	4
China		25	5005	6
Total		**100**	**500975**	**22**
Grand Total		**200**	**1001250**	**45**

The dissertations were downloaded from the web pages of the universities or from the legal web pages that seek author consent; in other words, the dissertations were open-access and available for research. One problem emerged while collecting the corpus: how to be sure about the writers' native language. Because this study makes a comparison between NW and NNW, it is of importance to confirm the writers' native language firmly. The first criterion was related to the university from which the author graduated while the second criterion was with the supervisor of the dissertation. In other words, authors needed to graduate from their own countries, and their supervisors needed to have the same nationality. For example, while building the corpus for the Chinese, it was ensured that the author of the dissertation had graduated from a Chinese university under a Chinese supervisor. The third criterion in the construction of the corpus was related to the name of the author; it is not prevalent that a westerner may have an eastern name while an easterner may highly hold a western name. In line with this, author name was taken as an additional evidence for nationality while constructing the corpus for Expanding-circles; however, much more was needed to be done to determine

the nationality of the authors in outer-circles. In case there should be a western name used by an easterner, the content of the dissertation was scanned because some dissertations included the CV of the author. It was noted that 19 dissertations did not include any personal information of the author; therefore, the authors were contacted through e-mail and informed about the purpose of the study. Fourteen of the 19 authors responded to the email and provided the necessary information while 5 did not reply to the email. Three of those who answered the email were not Anglophonic; therefore, they and those who did not provide a reply (8 in total) were replaced by other authors and the same procedure was repeated for them to gather data for outer-circle.

3.2. Analyses and procedure

The respective values, composed of 14 categories as shown in table 1, were inserted into the Syntactic Complexity Analyser (L2SCA) which is a computational system for the automatic analysis of syntactic complexity developed and reliability-tested by Lu (2010). This analyser measures the SC of texts by dividing them into two: syntactic structures and syntactic complexity indices. Syntactic structures present descriptive statistics of the texts by counting basic language components, namely word count, sentence, verb phrase, clause, T-unit, dependent clause, complex T-unit, coordinate phrase, and complex nominal; on the other hand, Syntactic complexity indices make more sophisticated calculations to measure Mean length of sentence (MLS), Mean length of T-unit (MLT), Mean length of clause (MLC), Clause per sentence (C/S), Verb phrase per T-unit (VP/T), Clause per T-unit (C/T), Dependent clause per clause (DC/C), Dependent clause per T-unit (DC/T), T-unit per sentence (T/S), Complex T-unit ratio (CT/T), Coordinate phrase per T-unit (CP/T), Coordinate phrase per clause (CP/C), Complex nominal per T-unit (CN/T), and Complex nominal per clause (CN/C).

The outputs from L2SCA were inserted into SPSS to make statistical analyses, but before starting the analyses the type of test-parametric or non-parametric needed to be decided on. Therefore, a normality test of Skewness and Kurtosis was used to determine whether they had a normal distribution (See Appendix for detail). The test results showed that 5

of 14 categories were not distributed normally (MLS, MLT, MLC, CN/T, and CN/C) while the rest had a normal distribution (the average skewness and kurtosis values were between -1 and +1); however, skewness and kurtosis results might not be sufficient to reach an absolute conclusion on data distribution. Therefore, histogram graphics, normal Q-Q and detrended normal Q-Q graphics, and box-blot graphics were manually controlled and it was concluded that the data were normally distributed for the categories of CN/T and CN/C, and thus the number of categories non-distributed normally fell from 5 to 3, namely MLS, MLT, and MLC. Finally, based on the results of normality tests, this study used ANOVA for normally distributed categories and Kruskal-Wallis test for non-normally distributed categories.

This study conducted analyses in two similar aspects: the first one was to obtain country-based SC levels of data while the second one was for group-based (Figure 1).

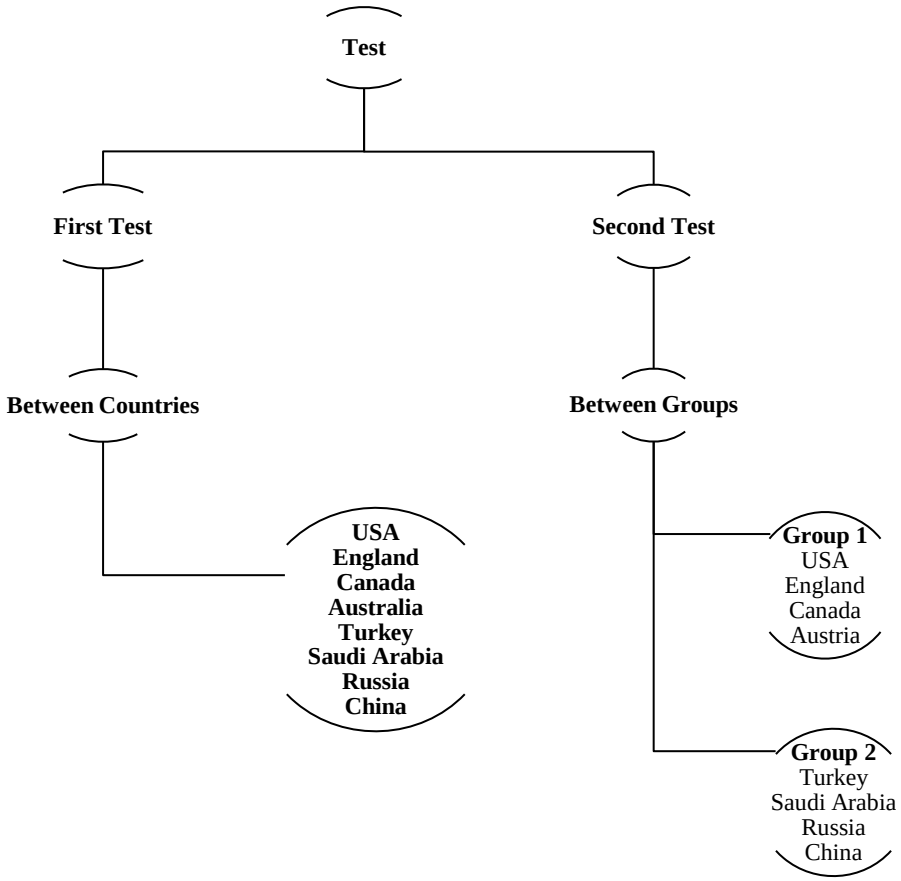

Figure 1. Graphic depicting the analyses.

To explain the figure, this study aimed to reveal whether SC levels of MA dissertations in the research countries vary in terms of 14 categories. Beside country-based SC differences, this study sought to find group-based differences; accordingly, the countries were divided as Inner-circle, henceforth Group 1, and Expanding countries, henceforth Group 2, and then the average SC scores of the groups were calculated. Because the population was reduced in number through dividing into groups and taking the average of the scores, the tests to be implemented were changed as summed up in Table 3.

Table 3.

The tests that were implemented in the study.

Test	Data	Test Type	Test Name
Between countries	Normally distributed	Parametric	ANOVA
	Non-normally distributed	Non-parametric	Kruskal-Wallis
Between groups	Normally distributed	Parametric	Independent Sample T-test
	Non-normally distributed	Non-parametric	Mann-Whitney

ANOVA and Kruskal-Wallis tests were employed to investigate the differences between countries and Independent-Samples t-test and Mann-Whitney tests to investigate between-group differences.

4. Results

Analyses were conducted to show whether each country differs from the other countries (first tests) and the differences that each group shows variance (second tests) and accordingly, the results were presented.

4.1. Results of the first test

To show whether the countries differ from one another significantly, ANOVA for the normally-distributed data and Kruskal-Wallis for the non-normally distributed data were employed and the results were provided under the relevant subtitles.

4.1.1. ANOVA results

Eleven categories, which the skewness tests proved normality of data distribution, were analysed through ANOVA and the results were tabulated in table 4.

Table 4. ANOVA results

		Sum of Squares	df	Mean Square	F	*p*
	Between Groups	580869606.435	7	82981372.348	8.292	
C/S	Within Groups	1921482964.960	192	10007723.776		**.000***
	Total	2502352571.395	199			
	Between Groups	361425072.560	7	51632153.223	3.532	
VP/T	Within Groups	2807021683.760	192	14619904.603		**.001***
	Total	3168446756.320	199			
	Between Groups	200880308.035	7	28697186.862	5.716	
C/T	Within Groups	964005276.960	192	5020860.818		**.000***
	Total	1164885584.995	199			
	Between Groups	19707121.795	7	2815303.114	4.374	
DC/C	Within Groups	123583333.360	192	643663.195		**.000***
	Total	143290455.155	199			
	Between Groups	169725746.160	7	24246535.166	5.086	
DC/T	Within Groups	915290024.960	192	4767135.547		**.000***
	Total	1085015771.120	199			
	Between Groups	24578442.275	7	3511206.039	4.134	
T/S	Within Groups	163061008.480	192	849276.086		**.000***
	Total	187639450.755	199			
	Between Groups	61660687.280	7	8808669.611	5.926	
CT/T	Within Groups	285408369.440	192	1486501.924		**.000***
	Total	347069056.720	199			
	Between Groups	54544944.880	7	7792134.983	.974	
CP/T	Within Groups	1536142931.200	192	8000744.433		.452
	Total	1590687876.080	199			
	Between Groups	43452445.675	7	6207492.239	2.018	
CP/C	Within Groups	590545667.200	192	3075758.683		.055
	Total	633998112.875	199			
	Between Groups	570960809.600	7	81565829.943	1.230	
CN/T	Within Groups	12730781908.480	192	66306155.773		.288
	Total	13301742718.080	199			
	Between Groups	325750833.180	7	46535833.311	2.013	
CN/C	Within Groups	4439623544.400	192	23123039.294		.055
	Total	4765374377.580	199			

*significant *p* value.

It was found that some of the categories are statistically significant which are C/S [F(7,192)=8.292, p<.05]; VP/T [F(7,192)=3.532, p<.05]; C/T [F(7,192)=5.716, p<.05]; DC/C [F(7,192)=4.374, p<.05]; DC/T [F(7,192)=5.086, p<.05]; T/S [F(7,192)=4.134, p<.05], and CT/T [F(7,192)=5.926, p<.05] while the rest did not yield statistically significant results: CP/T [F(7,192)=.974, p>.05], CP/C [F(7,192)=2.018, p>.05], CN/T [F(7,192)=1.230, p>.05], and CN/C [F(7,192)=2.013, p>.05]. Scheffe, as post hoc multiple comparisons, was used for the equal variance assumed data that were statistically significant and the results were provided in table 5.

Table 5. Scheffe post hoc multiple comparisons results.

Categories	Variables	Mean Difference	Std. Error	p
C/S	USA – Australia	4140.920	894.773	.004
	USA – Turkey	4072.560	894.773	.006
	USA – Russia	4754.680	894.773	.000
	USA – Saudi Arabia	4940.960	894.773	.000
	USA – China	4598.600	894.773	.001
	Canada – Russia	3465.480	894.773	.041
	Canada–S. Arabia	3651.760	894.773	.024
VP/T	USA – Russia	4319.120	1081.477	.030
C/T	USA – Australia	2547.560	633.774	.028
	USA – Turkey	2470.080	633.774	.039
	USA – Russia	3015.920	633.774	.003
	USA – S. Arabia	2580.240	633.774	.024
	USA – China	2866.720	633.774	.006
DC/C	USA – China	887.400	226.921	.037
DC/T	USA – Russia	2589.200	617.552	.017
	USA – China	2681.960	617.552	.011
T/S	USA – S. Arabia	1123.480	260.657	.012
	Canada – S. Arabia	1041.560	260.657	.030
CT/T	USA – Russia	1446.400	344.848	.017
	USA – S. Arabia	1373.480	344.848	.031
	USA – China	1387.280	344.848	.028

Data from the USA differs from other countries in all categories. To sum up the table, CS is the category where the differentiation expands at most (seven significances) while DCC is the least (one significance). The USA statistically differed from all other countries

but Saudi Arabia and UK in the category of CT proved a variance with USA and Canada in TS.

4.1.2. Kruskal-Wallis results

Three categories that the skewness tests did not prove normality of data distribution were analysed through Kruskal-Wallis and the results were tabulated in table 6.

Table 6.

Kruskal-Wallis test results

Categories	N	Mean	Std. Deviation	df	x^2	*p*
MLS	200	294958.06	84003.529	7	21.424	**.003***
MLT	200	260924.12	70794.391	7	9.460	.221
MLC	200	155372.94	37897.125	7	5.973	.543

*significant *p* value.

A Kruskal-Wallis H test showed that there was a statistically significant difference in MLS score between the countries [$\chi^2(7) = 21.424$, $p < .05$], but the result did not yield statistically significant difference for the MLT [$\chi^2(7) = 9.460$, $p > .05$] and MLC scores [$\chi^2(7) = 5.973$, $p > .05$]. To find between which countries the difference exists in MLS, Mann-Whitney U test was used and the results were shown in table 7.

Table 7.

Mann Whitney U test results

Variables	Mean Rank	U	*p*
USA	30.92	177.000	.009
Australia	20.08		
USA	31.88	153.000	.002
Turkey	19.12		
USA	30	200.000	.029
Russia	21		
USA	31.26	168.500	.005
Saudi Arabia	19.74		
UK	29.94	201.500	.031
Turkey	21.06		
Australia	20.04	176.000	.008
Canada	30.96		
Canada	32.66	133.500	.001
Turkey	18.34		
Canada	29.92	202.000	.032
Russia	21.08		
Canada	31.08	173.000	.007
Saudi Arabia	19.92		

The test results showed that USA differs in the category of MLS from Australia [U=177, p<.05]; from Turkey [U=153, p<.05]; from Russia; [U=200, p<.05], and from Saudi Arabia [U=168.500, p<.05] while Turkey also differs from UK [U=201.500, p<.05] and Canada [U=133.500, p<.05]. Furthermore, a statistically significant difference was found between Australia and Canada [U=176, p<.05]; Canada and Russia [U=202, p<.05], and finally between Canada and Saudi Arabia [U=173, p<.05].

4.2. Results of second test

The second tests include Independent sample t-test and Mann-Whitney U test. Different from the first tests, Independent sample t-test was employed for the normally-distributed data and Mann-Whitney U test for the non-normally distributed data.

4.2.1. Independent sample t-test results

Eleven categories with normal data distribution were analysed through independent sample t-test and the scores were tabulated in table 8.

Table 8.

Independent sample t-test results

Variables	t	df	Mean Difference	Std. Error Difference	*p*
C/S	5.527	198	2586.350	467.952	**.000***
VP/T	1.827	198	1024.880	561.017	.069
C/T	4.630	198	1508.550	325.840	**.000***
DC/C	4.276	198	492.190	115.110	**.000***
DC/T	4.511	198	1421.960	315.255	**.000***
T/S	3.747	198	498.450	133.036	**.000***
CT/T	4.919	198	869.480	176.746	**.000***
CP/T	.122	198	49.080	400.828	.903
CP/C	-1.426	198	-359.070	251.772	.155
CN/T	.543	198	628.640	1158.281	.588
CN/C	-2.195	198	-1504.960	685.501	**.029***

*significant *p* value.

Out of eleven categories, seven categories had significant *p* value which are C/S [$t(198)=5.527$, $p<.05$]; C/T [$t(198)=4.630$, $p<.05$]; DC/C [$t(198)=4.276$, $p<.05$]; DC/T [$t(198)=4.511$, $p<.05$]; T/S [$t(198)=3.747$, $p<.05$]; CT/T [$t(198)=4.919$, $p<.05$], and CN/C [$t(198)=-2.195$, $p<.05$] whereas the rest, four categories, did not yield a significant *p* value which are VP/T [$t(198)=1.827$, $p>.05$]; CP/T [$t(198)=.122$, $p>.05$]; CP/C [$t(198)=-1.426$, $p>.05$]; and CN/T [$t(198)=.543$, $p>.05$].

4.2.2. Mann-Whitney test results

Before starting the testing, the average SC scores of each country were calculated by taking mean scores of each category and the results were provided in table 9.

Table 9. Country-based SC mean scores

Category	USA	England	Canada	Australia	Turkey	S. Arabia	Russia	China
MLS	**332.51**	300.10	332.03	272.56	266.24	273.30	290.25	292.72
MLT	**281.36**	265.71	281.12	243.68	238.79	254.46	260.21	262.08
MLC	149.59	151.61	158.32	150.78	146.77	154.87	165.54	**166.53**
C/S	**22.46**	19.87	21.17	18.30	18.39	17.51	17.70	17.86
VP/T	**26.19**	23.39	25.17	22.6	24.721	23.74	21.87	22.93
C/T	**18.87**	17.54	17.83	16.3	16.40	16.29	15.84	16
DC/C	**0.433**	0.410	0.425	0.371	0.379	0.368	0.351	0.344
DC/T	**3.392**	1.918	2.324	1.031	0.632	1.375	1.0	1.40
T/S	11.57	11.32	**11.83**	11.20	11.22	10.43	11.18	11.15
CT/T	**0.568**	0.537	0.548	0.450	0.471	0.431	0.432	0.429
CP/T	2.10	3.46	2.24	**4.26**	4.10	2.99	2.74	3.32
CP/C	0.402	0.451	0.429	1.33	**0.507**	0.506	0.460	0.506
CN/T	36.37	36.83	34.40	32.67	34.34	33.40	32.43	**37.03**
CN/C	19.38	20.86	19.27	19.94	21.09	20.73	20.60	**23.66**

Apart from country-based mean calculation, this study also investigated group-based SC means, and presented in table 10.

Table 10.

Group-based mean scores

Category	Group 1	Group 2
MLS	**309.3**	280.7
MLT	**268**	253
MLC	152.6	**158.4**
C/S	**20.5**	17.9
VP/T	**24.3**	23.3
C/T	**17.7**	16.1
DC/C	**0.41**	0.37
DC/T	**2.2**	1.1
T/S	**11.5**	11
CT/T	**0.5**	0.4
CP/T	3	**3.3**
CP/C	**0.7**	0.5
CN/T	**35**	34
CN/C	20	**22**

Mann-Whitney tested the categories of MLS, MLT, and MLC which all were not distributed normally according to the result of skewness tests and the result were tabulated in table 11.

Table 11.

Mann-Whitney U test results

Variables		Mean Rank	U	*p*
MLS	Group 1	113.81	3669	**.001***
	Group 2	87.19		
MLT	Group 1	108.89	4161.500	**.040***
	Group 2	92.12		
MLC	Group 1	95.82	4531.500	.252
	Group 2	105.19		

*significant *p* value.

MLS scores of Group 1 (*Mdn*=113.81) were higher than Group 2 (*Mdn*=87.19) and the test indicated that this was statistically significant [*U*=3669, *p*<.05]. Similarly, MLT scores of Group 1 were higher when compared to Group two, which underlines a significant

result, [U=4161.500, p<.05]. However, there was not a statistically significant difference between Group 1 and Group 2 in terms of MLC, [U=4531.500, p>.05].

5. Discussion

This study aimed to investigate MA dissertations of Inner and Expanding-circle countries to find SC levels and to create a datum point (a reference SC level) for prospective MA writers in ELT. Besides, the question of whether there was a statistically significant difference between Inner and Expanding countries in terms of MA writing in English was answered through analyses. The study divided the data into 14 categories and each category was analysed separately.

The analyses for category-based differences between countries showed that 8 out of 14 categories differed from one another (MLS, C/S, VP/T, C/T, DC/C, DC/T, T/S, and CT/T), which is similar to Ai and Lu's study (2013) that compared native and non-native university students, and found that 8 out of 10 categories had statistically significant differences. The only variance with their study was the category of T/S i.e., this study found a statistically significant difference while theirs did not, which may be because this study collected data from MA students whilst theirs collected data from BA students.

USA was the country with the most significant differences from other countries particularly in C/S and C/T categories. It seems that writers in the USA had longer sentences compared to the writers in other countries, which may be because SC plays a crucial role in shaping a good score (Plakans et al., 2019), and this positive correlation between SC and score gaining may incite student writers in USA to endeavour to produce complex sentences compared to their peers in other countries. On the other hand, Davidson (2005), who investigated SC in accounting textbooks over the past 100 years, stated that the complexity of writing in these textbooks decreased through shorter sentences, and we know that readability is of great importance for textbooks (Burton, 2014). Therefore, different from MA dissertations, a textbook with shorter sentences

needs to be taken normal because of its capability (and necessity) to increase readability of textbooks (see Kasule, 2011). Another noteworthy result is concerning the close relation between orthography and writing. This study found that Saudi Arabia is the country that differed more than other countries (see table 5) while Turkey is the country that differed least (Saudi Arabia > Russia > China >Turkey). Saudi Arabia had a variation of six i.e., it showed variance in four different categories with six countries, which may be because of the alphabetical difference that writers used. It is known from one of the earliest studies (Thompson-Panos and Thomas-Ruzic, 1983) that orthography may play an important role in the weaknesses of Arabic writers in writing due to the different alphabet they use. Similarly, Russia and China also have different orthographic structures compared to English while it is only Turkey with a similar alphabet (see Hirshorn et al., 2016 for detail). On the other hand, the categories of MLT, MLC, CP/T, CP/C, CN/T, and CN/C did not show any differences, which means that all countries in the study had similar SC levels in these categories. Regarding CP/T and CP/C, it can be said that the writers either avoided using coordinate phrases for the sake of the readers so that they could be more comprehensible, or just because they were in some way difficult to produce (Temperley, 2005). However, difficulty is not always the only reason for lack of use as in the categories of CN/T and CN/C. Although complex nominals (e.g. wind turbine) are common in English, they are largely used in specialized technical texts (Nakov, 2013) such as Engineering. Taking that the data came from MA dissertations in ELT which is not a technical field into account, the results showed that writers in all countries do not have a tendency to use complex nominals much in their texts, which lowered the number of use, hence caused non-significance in the analyses. Accordingly, this study found lower CN/T and CN/C between 32.43-37.03 and 19.38-23.66, respectively (cf. Cabezas-Garcia and Faber, 2017).

Different from the first tests that emerged country-based differences on the basis of 14 categories, the second tests allowed us to understand whether there was a statistically significant difference between Inner and Expanding circles countries. Following placing the countries into two groups as Group 1 (Inner countries) and Group 2 (Expanding

countries), the analyses yielded that there were statistically significant differences in 9 out of 14 categories (viz., MLS, MLT, C/S, C/T, DC/C, DC/T, T/S, CT/T, and CN/C). It was found that the SC scores were in favour of Group 1 when the average scores were calculated (Table 10). It is understood from MLS and MLT scores that Group 1 writers had longer sentence constructions when compared to Group 2. Accordingly, we know that academic writing is constructed with longer sentences and t-units (Biner and Gray, 2010), and the superiority of native speakers of English in building up more elaborate sentences with higher MLS and MLT rates is not surprising for many writers in academe. Similarly, this was proven with high C/S and C/T scores that the present study found. Group 1 writers had longer clauses per sentence and per t-unit while the scores were lower for Group 2. Although there is a statistically significant difference in favour of Group 1 between two groups, Group 2 scored rather high in C/S and C/T, which is an expected result because persuasive writing as in academic writing necessitates high C/S and C/T SC scores (Beers and Nagy, 2011) when compared to other writing styles. Dependent clauses are another determinant indicator of academic writing particularly when the issue is SC; accordingly, a significant difference in favour of Group 1 means that writers in English-speaking countries are more prone to build their texts with longer dependent clauses when compared to writers in non-English speaking countries. Taguchi et al. (2013) found that high-rated essay has a higher DC/C and DC/T scores when compared to low-rated essays and this does not refer to the quality of the texts. In other words, it is not yet possible to create a possible correlation between writing quality and scores of DC/C and DC/T. About T/S and CT/T, both referring to T-units complexity, it is obvious that Group 1 is more successful to split their sentences into short grammatical units than Group 2. This descriptive result should be taken as a superior skill for Group 1 writers as this common conception among academic writers was challenged by Biber et al. (2011). Finally, to comment on CN/C, some studies on the importance of nominals draw a different perspective and argue that scholarly writing is more nominal than verbal (Biner and Gray, 2010). This view seems to be more valid for non-native writers of English because the statistically significant difference for the category of CN/C is in favour of Group 2, which points out that writers in Expanding

countries use complex nominals more than their peers in Inner countries. Similarly, Mancilla et al. (2017) found that non-native writers of English had higher CN/C scores than native writers of English; however, this study contradicts Ai and Lu's (2013) results indicating the higher scores of native writers compared to non-native writers in terms of CN/C. The reason for the contradictory results may stem from the data which had been collected from offline writing assignments while this study collected its data from MA dissertations which largely necessitate a more carefully planned writing procedure.

The average scores of the groups deserve final remarks: results show that Group 1 that host native writers have superior scores in 11 out of 14 categories while Group 2 have higher averages in 3 categories. Thanks to calculating the average points, we can create a datum point for the expected SC level while writing an MA dissertation as stated in Table 10. One thing to note is that although Group two have higher scores in 3 categories viz., MLC, CP/T, and CN/C, these scores are not much far away from the averages of Group 1; accordingly, a statistically significant result already was not found in the categories of MLC and CN/C. Therefore, it can be summarized that the only category that is in favour of Group 2 is CP/T.

6. Conclusion

The literature does not suggest any reference point of SC for students writing their MA dissertation and this may be a deficiency for some writers caring about the SC level of their writing. This study is just a prelude to creating a standard SC level for MA students writing their dissertation in ELT. Surely, more studies will fill up the gaps that exist in the present study, one of which is the small data that this study worked on. Though a hundred dissertations were analysed, further studies are welcome to increase the data so that they can both validate the results of this study and help to determine a score for each category. Also, this study broke down SC into 14 categories; however, studies in the future can increase the categorization to expand the results or decrease to focalize specifically. Finally, researchers are kindly invited to investigate other disciplines because this study investigated MA theses only in ELT.

Ai and Lu (2013) compared university students' writing in four areas of SC and concluded that they needed pedagogical implications to overcome low SC levels in their writing. One suggestion for the MA students would be that although they do not have to write their dissertations according to a standard SC level, keeping up with the SC scores that this study found may help them gain native-fluency from the aspect of readers (see Fellner and Apple, 2006; Housen and Kuiken, 2009 for the studies establishing a link between SC and writing fluency). MA Students preparing their dissertations are encouraged to pay attention particularly to the SC averages where the difference is significant.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson, *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 249-264). Amsterdam / Philadelphia: John Benjamins.

Barnwell, D. (1988). Some comments on T-unit research. *System*, 16(2), 187-192. https://doi.org/10.1016/0346-251X(88)90032-2.

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. https://doi.org/10.1007/s11145-010-9264-9.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2-20.

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, *24*(2), 183-202.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

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. https://doi.org/10.5054/tq.2011.244483.

Buhr, A., & Zebrowski, P. (2009). Sentence position and syntactic complexity of stuttering in early childhood: A longitudinal study. . *Journal of fluency disorders*, 4(3), 155-172. https://doi.org/10.1016/j.jfludis.2009.08.001.

Bulte, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder, *Dimensions of L2 performance and proficiency –*

Investigating complexity, accuracy and fluency in SLA (pp. 21-46). Amsterdam: John Benjamins.

Burton, R. S. (2014). Readability, Logodiversity, and the Effectiveness of College Science Textbooks. *Bioscene: Journal of College Biology Teaching*, *40*(1), 3-10.

Cabezas-García, M., & Faber, P. (2017, November). A semantic approach to the inclusion of complex nominals in English terminographic resources. In *International Conference on Computational and Corpus-Based Phraseology* (pp. 145-159). Springer, Cham.

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research, 69*, 176–183.

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. http://dx.doi.org/10.1016/j.jslw.2014.09.006.

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28,* 282–311.

Davidson, R. A. (2005). Analysis of the complexity of writing used in accounting textbooks over the past 100 years. *Accounting Education*, *14*(1), 53-74.

Droop, M., & Verhoeven, L. (1998). Background Knowledge, Linguistic Complexity, and Second-Language Reading Comprehension. *Journal of Literacy Research*, 253-271.

Eckstein, G., & Ferris, D. (2018). Comparing L1 and L2 texts and writers in first-year composition. *TESOL Quarterly*, *52*(1), 137-162.

Elcin, D. (2017). Teachers' and Students' Perceptions of English Teachers as Mediators. In S. Sahinkarakas, & J. Inozu, *The Role of the Self in Language Learning* (pp. 135-154). Cambridge Scholars Publishing.

Fellner, T., & Apple, M. (2006). Developing writing fluency and lexical complexity with blogs. *The jalt call Journal*, *2*(1), 15-26.

Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. M. Zwicky, *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 129-189). Cambridge: Cambridge University Press.

Gaies, S. (1980). T-Unit Analysis in second language research: Applications, problems and limitations. *TESOL Quarterly*, 14(1), 53-60. https://doi.org/10.2307/3586808.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 1-76.

Givo´n, T. (2009). *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution.* Amsterdam: John Benjamins.

Givo´n, T., & Shibatan, M. (Eds.). (2009). *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution.*

Gökmen, M. F. (2020). *Post-method Pedagogies for English Language Teaching: Emerging Research and Opportunities*. Nobel

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix:Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202. https://doi.org/10.3758/BF03195564.

Gyllstadt, H., Granfeldt, J., Bernardini, P., & Källkvist, M. (2014). Linguistic correlates to communicative proficiency levels of the CEFR: The case of syntactic complexity in written L2 English, L3 French and L4 Italian. In L. Roberts, I. Vedder & J.H. Hulstijn (Eds.), Eurosla Yearbook, 14 (pp.1–30). Amsterdam: John Benjamins.

Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). London: Arnold

Hirshorn, E. A., Wrencher, A., Durisko, C., Moore, M. W., & Fiez, J. A. (2016). Fusiform gyrus laterality in writing systems with different mapping principles: An artificial orthography training study. *Journal of cognitive neuroscience*, *28*(6), 882-894.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, *30*(4), 461-473.

Hunt, K. (1965). *Grammatical structures written at three grade levels.* Campaign, IL: NCTE Research report No.3.

Hyland, K. (2005). *Metadiscourse: exploring Interaction in Writing.* New York: Continuum.

Kachru, B. B. (1992). A Biographical Sketch. *World Englishes*. **11,** 91–94. doi:10.1111/j.1467-971X.1992.tb00053.x.

Kasule, D. (2011). Textbook readability and ESL learners. *Reading & Writing-Journal of the Reading Association of South Africa*, *2*(1), 63-76.

Klebanov, B. B., & Flor, M. (2018). *Patent No. 10,134,297.* US/ Washington, DC.

Kyle, K. (2014). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication (Doctoral Dissertation).* Georgia, US: Georgia State University. https://scholarworks.gsu.edu/alesl_diss/35.

Kyle, K., & Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. *The Modern Language Journal*, 102(2), 333-349. DOI: 10.1111/modl.12468.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction.* Oxford: Oxford University Press

Lantolf, J. (1988). The Syntactic complexity of written texts in Spanish as a foreign language: A markedness perspective. *Hispania*, 71, 933-940.

Larsen-Freeman, D., & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning*, 27(1), 123-134. https://doi.org/10.1111/j.1467-1770.1977.tb00296.x.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, 15(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45,* 36–62

Mancilla, R. L., Polat, N., & Akcay, A. O. (2017). An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions. *Applied Linguistics*, *38*(1), 112-134.

Miller, J., & Chapman, R. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24:154-161.

Nakov, P. (2013). On the interpretation of noun compounds: Syntax, semantics, and entailment. Natural Language Engineering, 19(03): 291–330.

Neff, J., Dafouz, E., Díez, M., Prieto, R., & Chaudron, C. (2004). Contrastive discourse analysis: Argumentative text in English and Spanish. In *Discourse across languages and cultures* (pp. 267-283). John Benjamins.

Newmeyer, F. J., & Preston, L. B. (2014). *Measuring grammatical complexity.* USA: Oxford University Press.

O'Donnell, R. (1976). A critique of some indices of syntactic maturity. *Research in the Teaching of English*, 10, 31-38.

Ortega, L. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Syndissertation of College-level L2 Writing. *Applied Linguistics*, 24(4), 492-518. https://doi.org/10.1093/applin/24.4.492.

Plakans, L., Gebril, A., & Bilki, Z. (2019). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing*, *36*(2), 161-179.

Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language*, *30*(1), 3-26.

Roll, M., Frid, J., & Horne, M. (2007). Measuring Syntactic Complexity in Spontaneous Spoken Swedish. *Language and Speech*, 50(2), 227-245.

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *Tesol Quarterly*, *47*(2), 420-430.

Temperley, D. (2005). The dependency structure of coordinate phrases: A corpus approach. *Journal of psycholinguistic research*, *34*(6), 577-601.

Thompson-Panos, K., & Thomas-Ruzic, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *Tesol Quarterly*, *17*(4), 609-623.

Walters, J., & Wolf, Y. (1996). Language awareness in non-native writers: Metalinguistic judgments of need for revision. *Language Awareness*, 5(1), 3-25. https://doi.org/10.1080/09658416.1996.9959888.

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

Yngve, V. (1960). A model and a hypodissertation for language structure. *Proceedings of the American Philosophical Society* (pp. 104 (5), 444-466. American Philosophical Society

Appendix

The table representing the results of normality test.

Category	Mean	Mode	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Sum
MLS*	**294.958.07**	**228.409[a]**	**2.704**	**.172**	**13.361**	**.342**	**58.991.613**
MLT*	**260.924.12**	**330.000**	**2.477**	**.172**	**10.689**	**.342**	**52.184.824**
MLC*	**155.372.94**	**125.823[a]**	**2.078**	**.172**	**7.411**	**.342**	**31.074.588**
C/S	19.156.56	20.000	.734	.172	.476	.342	3.831.311
VP/T	23.829.22	23.333	.157	.172	-.533	.342	4.765.844
C/T	16.882.76	18.000[a]	.438	.172	.289	.342	3.376.551
DC/C	3.851.69	3.333[a]	-.354	.172	-.002	.342	770.337
DC/T	6.690.62	5.000[a]	.420	.172	.007	.342	1.338.124
T/S	11.315.97	10.000	.923	.172	.840	.342	2.263.193
CT/T	4.821.58	4.444	.114	.172	-.058	.342	964.316
CP/T	7.907.36	10.000	.745	.172	.678	.342	1.581.472
CP/C	4.767.18	3.333[a]	.670	.172	.443	.342	953.435
CN/T**	**34.753.64**	**25.000**	**.375**	**.172**	**1.264**	**.342**	**6.950.728**
CN/C**	**20.766.61**	**20.000**	**.693**	**.172**	**2.456**	**.342**	**4.153.322**

* Category resulted in skewed distribution in data set.

** Category resulted in skewed distribution but decided to be non-skewed on account of graphic confirmation.