Classification of L2 Thesis Statement Writing Performance Using Syntactic Complexity Indices

Kutay Uzun
Trakya University, Turkey
Department of English Language Teaching
kutayuzun@trakya.edu.tr

Abstract

This study primarily aimed to find out if machine learning classification algorithms could accurately classify L2 thesis statement writing performance as high or low using syntactic complexity indices. Secondarily, the study aimed to reveal how the syntactic complexity indices from which classification algorithms gained the largest amount of information interacted with L2 thesis statement writing performance. The data set of the study consisted of 137 high-performing and 69 low-performing thesis statements written by undergraduate learners of English in a foreign language context. Experiments revealed that the Locally Weighted Learning algorithm could classify L2 thesis statement writing performance with 75.61% accuracy, 20.01% above the baseline. Balancing the data set via Synthetic Minority Oversampling produced the same accuracy percentage with the Stochastic Gradient Descent algorithm, resulting in a slight increase in Kappa Statistic. In both imbalanced and balanced data sets, it was seen that the number of coordinate phrases, coordinate phrase per t-unit, coordinate phrase per clause and verb phrase per t-unit were the variables from which the classification algorithms gained the largest amount of information. Mann-Whitney U tests showed that the high-performing thesis statements had a larger amount of coordinate phrases and higher ratios of coordinate phrase per t-unit and coordinate phrase per clause. The verb phrase per t-unit ratio was seen to be lower in high-performing thesis statements than their low-performing counterparts.

Keywords: L2 Writing Performance, Machine Learning, Syntactic Complexity, Thesis Statement, Performance Classification

1. Introduction

Writing in L2 is one of the difficult skills within the higher education context where most assignments and exams need to be performed and evaluated in written form. This difficulty comes from the fact that writing in L2 requires a variety of motor skills and memory resources for the successful completion of the task (Burdick et al., 2013). Even though there is a large body of research investigating the factors that have effects on L2 writing performance in general, much of the previous literature on the construct seems to fall behind the advances in computational linguistics which offer numerous opportunities to second language acquisition/learning researchers by allowing them to analyse large chunks of learner texts by means of natural language processing and corpus analysis methods (Meurers, 2012).

Many of the texts written in the higher education context are in the form of essays, which typically have a main idea expressed as the thesis statement. Borrowing from Systemic Functional Linguistics the concept of macro-theme (Halliday and Mathiessen, 2004; Martin, 1992), Miller and

Pessoa (2016) define a thesis statement as a generalized main idea, located typically at the end of an introduction paragraph, which serves to predict the overall development of a text by stating the topic and making suggestions regarding how a particular point of view would be supported. Burstein et al. (2001) define the concept in a similar way, indicating that a thesis statement is an explicitly stated sentence that includes the main idea and the purpose of a text. From these definitions, it is seen that a thesis statement is mainly a summary of the core of a text, stating the central claim and the argumentative structure explicitly.

The importance of the thesis statement in writing stems from the fact that it does not only carry the main idea of a text, but it is also a sufficiently powerful part of a text that distinguishes a high-quality text from a low quality one. For instance, Coffin (2006) states that a successful essay in history writing contains a macro-theme which suggests the development of the text. Similarly, Oliveira's (2011) study reveals that history essays written by 11^{th}-grade students were distinguishable in terms of success by having a macro-theme or not. In an English as a Foreign Language (EFL) context, Uzun (2019: 31) discovers that the thesis statement is the strongest rhetorical move in a literary analysis essay in terms of predicting total performance with the following equation for prediction intervals:

$$\text{Essay Score} = 18.377 + (\text{Thesis Statement Score} \times 3.748) \pm (1.96 \times 9.595)$$

Even though the literature indicates that the thesis statement is the most vital part of a text, it is seen that this particular part is yet an underresearched one. For this reason, it is argued in this study that the linguistic variables which contribute to a good thesis statement should be identified using corpus analysis and/or natural language processing methods.

Being an increasingly researched area by means of the mentioned corpus analysis and natural language processing methods, syntactic complexity appears to be an integral part of L2 writing quality. In general, syntactic complexity measures attempt to produce frequency counts of interconnected components within the structures of a language (Pallotti, 2014). Neary-Sundquist (2017) briefly describe those components as the length of certain phrases, their frequency per clause and the frequency of clauses per unit. According to Wolfe-Quintero, Inagaki and Kim (1998), the ratio of dependent clauses to clauses and clauses to t-units as well as the lengths of t-units and clauses are among the measures that can give clues regarding L2 writing performance. In addition, Ai and Lu (2013) suggest that the frequencies of subordination and coordination in addition to the length of production units are also among the syntactic complexity measures. Ortega (2003) suggests that the mean lengths of clause, t-unit and sentence are syntactic complexity measures, too. Casanave (1994) states that the amount of complex t-unit per t-unit is also a measure of the construct. Lu (2011) adds to the others by suggesting coordinate clauses per clause, coordinate phrase per t-unit, complex nominal per clause and complex nominal per t-unit as the measures positively correlated with syntactic complexity and dependent clause per t-unit and per clause as the negatively correlated measures.

Studies of Biber, Gray and Staples (2016), Staples and Reppen (2016), Yang, Lu and Weigle (2015) and Casal and Lee (2019) reveal that syntactic complexity and L2 writing quality are related constructs with higher levels of complexity indicating higher quality and lower levels indicating lower quality in L2 writing according to the findings. The exception to this is Crossley and McNamara's (2014) study, in which they reveal that there is no statistically significant correlation between phrasal syntactic complexity measures L2 writing quality. However, to the researcher's knowledge, none of these studies have a particular focus on the thesis statement, which is the strongest predictor of writing quality as mentioned above.

Considering the significance of both the thesis statement and syntactic complexity in L2 writing performance, it can be said that searching for the syntactic complexity measures that signal L2 thesis statement writing performance seems to be a worthy endeavour. For this reason, this study aims to fill in a gap in the literature by identifying the syntactic complexity measures which can be utilized to identify L2 thesis statement writing performance. In line with the aims of the study, the following research questions have been formulated:

1. Can syntactic complexity measures accurately classify L2 thesis statement writing performance?
2. Can the accuracy of L2 thesis statement writing performance classification using syntactic complexity indices be increased by balancing the data set?

3. How do the syntactic complexity indices from which classification algorithms gain the largest amount of information interact with L2 thesis statement writing performance?

2. Methodology

The study employed a machine learning (ML) approach to solve the classification problem. ML is a subfield of artificial intelligence that is utilized to discover relationships, patterns or rules using statistical methods to solve prediction or classification problems (Hastie et al., 2009; Murphy, 2012; Witten and Frank, 2005). Since this study aimed to classify L2 thesis statement writing performance using syntactic complexity indices, ML was considered suitable for the purposes of the study.

2.1. Context

The research context was a compulsory English Literature course in the English Language Teaching department of a public university in Turkey. Aiming to teach students how to analyze and interpret literary texts written in English, which is their L2, the English Literature course requires an extensive use of essay writing skills. The essays that the students write in this course are literary analysis essays, in which they write their personal interpretations of how a theme or character is presented in a text or how a particular concept is functionally used to form the plot structure.

The literary analysis essays within the context of the study are typically in the expository or argumentative style, 400-600 words in length and include an introduction (stating the background to the work and the thesis of the essay), main body (presenting, supporting/proving and concluding arguments) and conclusion (consolidating the thesis and stating personal opinion).

2.2. The Corpus

For the creation of a corpus relevant to the research aims, 206 literary analysis essays were chosen by the researcher. These were reliably scored in previous studies using the Genre-Based Literary Analysis Essay Scoring Rubric (Uzun, 2019; Uzun, In Press; Uzun, Unpublished Manuscript, Uzun & Zehir Topkaya, 2019), which is an analytical scoring rubric that is used to score each rhetorical move in a literary analysis essay and produce a total score between 0 and 100. The rubric allows for the scoring of the thesis statement separately between 0 and 15 where 15 is suitable for a thesis statement which provides a direct response to the essay question with at least two points that can be developed and justified in the main body, using appropriate grammar and lexis.

Within the research context, the thesis statement of a literary analysis essay is typically located at the end of the introduction paragraph and it can be in the form of a single sentence or a few related sentences (Uzun, 2019). Considering this description, the thesis statements of the essays were manually extracted along with their thesis statement scores by the researcher.

As a result, a corpus of 206 thesis statements with a sum of 3946 words (M = 19.16, SD = 8.93) were obtained. The thesis statements within the corpus had a minimum of 5 and a maximum of 69 words. In accordance with the scoring weights of the rubric, all thesis statements had scores between 0 and 15 (M = 10.97, SD = 3.47).

2.3. The Dataset

Each thesis statement was analysed using the web-based L2 Syntactic Complexity Analyzer (L2SCA) developed by Lu (2010), Lu (2011), Ai and Lu (2013) and Lu and Ai (2015). L2SCA (available for public use on https://aihaiyang.com/software/) is a web-based piece of software which was written in Python and generates syntactic complexity indices by means of Natural Language Processing methods, part-of-speech tagging and morphological analyses. The following variables, all of which were continuous, were obtained in this study as a result of the analyses:

- Word count (W)
- Sentence count (S)
- Verb phrase count (VP)
- Clause count (C)
- Clause per sentence (C/S)
- Verb phrase per t-unit (VP/T)
- Clause per t-unit (C/T)
- Dependent clause per clause (DC/C)

- T-unit count (T)
- Dependent clause count (DC)
- Complex T-unit count (CT)
- Coordinate phrase count (CP)
- Complex nominal count (CN)
- Mean length of sentence (MLS)
- Mean length of t-unit (MLT)
- Mean length of clause (MLC)
- 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)
- Complex nominal per clause (CN/C)

The operational definitions of the key terms related to the variables are presented below in Table 1.

Term	**Definition**	**Source**
Sentence	Group of words ending with a sentence-final punctuation mark	Lu (2011)
Clause	Group of words with a subject, finite verb but no nonfinite verbs	Lu (2011)
Dependent Clause	A finite nominal, adjective or adverbial clause	Lu (2011)
T-unit	A main clause + any subordinate clause or nonclausal structure	Hunt (1970)
Complex T-unit	A t-unit which contains at least one dependent clause	Lu (2011)
Coordinate Phrase	A coordinating verb, noun, adverb or adjective phrase	Lu (2011)
Verb Phrase	Finite or nonfinite verb phrases	Lu (2011)
Complex Nominal	1. Noun + participle, appositive, prepositional, possessive, adjective phrase or clause 2. A nominal clause 3. Gerund or infinitive as subject	Lu (2011)

Table 1. Operational Definitions of Key Terms

Following the computation of the mentioned variables, the thesis statement scores in the corpus were grouped as Low (n = 69, M = 6.74, SD = 2.00) and High (n = 137, M = 13.10, SD = 1.59) by means of a cluster analysis which produced a good fit with two clusters.

As an example of a high-scoring thesis statement, the following thesis statement, written as a response to the question "How is the concept of reputation presented in Beowulf?", can be seen:

> The concept of reputation in Beowulf is represented in through the main character in two aspects: victories of Beowulf and his loyalty. (Essay 28)

As seen above, the thesis statement provides a direct answer to the essay question and includes two arguable points (i.e. victories and loyalty) that can be further explained in the main body paragraphs of the essay. Also having a clear language appropriate for academic writing despite negligible errors, the thesis statement of Essay 28 has a score of 15/15.

The high-scoring thesis statement is a single clause and a single t-unit which has 22 words, one verb phrase, one coordinate phrase and three complex nominals. It does not have any dependent clause or complex t-unit.

An example of a low-scoring thesis statement in the corpus for the same question is given below:

> What given Beowulf by the poet as character are huge power and beautiful, faithful attitude. (Essay 5)

In this example, the thesis cannot be directly linked to the essay question unless the rater makes inferences which may or may not have been considered by the learner-writer. Moreover, erroneous grammar and low-level lexis is visible in the text. Therefore, it has a score of 1/15.

The low-scoring thesis statement in the example is also a single clause and single t-unit with 15 words, two verb phrases, one coordinate phrase and two complex nominals. The statement does not have any dependent clause or complex t-unit.

2.4. Experiment and Data Analysis

Weka 3.8.2 (Eibe, Mark & Witten, 2016) was used for the experiments. The data set of 137 high and 69 low-performing thesis statements was initially divided by 80:20 as training (n_{low} = 51, n_{high} = 114) and test (n_{low} = 18, n_{high} = 23) data to avoid overfitting. To get the baseline classification accuracy, the ZeroR algorithm was run using both sets of data, outputting 56.10% (*KS* = .00) classification accuracy. Following the computation of the baseline accuracy, Naïve-Bayes (NB), Logistic Regression (LR), Sequential Minimal Optimization (SMO), Stochastic Gradient Descent (SGD), KStar (K*), Instance-based Learning with Parameter K (Ibk), J48, Random Forest (RF), Locally Weighted Learning (LWL) and Random Tree (RT) algorithms were tested in terms of their classification accuracy.

Synthetic Minority Oversampling was used to balance the training data set due to its superiority over random resampling methods (Akbani et al., 2004). As a result, a balanced data set of 114 high-performing and 112 low-performing thesis statements was generated. The same algorithms were tested with the balanced data set.

A confusion matrix was produced for the most successful algorithm following the tests with the original and balanced data sets. To find out the variables which provided the largest amount of information to the classifiers, InfoGainAttributeEval algorithm was used.

Along with classification accuracy, the Kappa Statistic was also reported to control for the chance factor in the classification (Ben-David, 2008).

Since none of the variables which provided the largest amount of information to the classifiers was distributed normally, Mann-Whitney U tests were run to see how those variables interacted with high and low L2 thesis statement writing performance groups.

3. Results

The results of the experiments to find the best algorithm that would classify L2 thesis statement writing performance using syntactic complexity indices are presented below.

Algorithm	**Accuracy (%)**	***KS***
LWL	75.61	.50
LR	70.73	.41
SGD	70.73	.40
Ibk	68.29	.34
RF	65.85	.28
K*	65.85	.27
RT	65.85	.29
J48	63.42	.23
NB	60.98	.16
SMO	58.53	.10

Table 2. Classification Performance of Different Algorithms

As seen in Table 2, the best-performing algorithms to classify L2 thesis statement writing performance accurately were LWL, LR and SGD, which outputted classification accuracy percentages of 75.61 (*KS* = .50), 70.73 (*KS* = .41) and 70.73 (*KS* = .40) respectively. The values were seen to be 15-20% above the baseline accuracy. On the other hand, J48 ($\%_{accuracy}$= 63.42, *KS* = .23), NB ($\%_{accuracy}$= 60.98, *KS* = .16) and SMO ($\%_{accuracy}$ = 58.53, *KS* = .10) were seen to be the least successful algorithms, exceeding the baseline accuracy only by a few percents.
The confusion matrix for the LWL algorithm can be seen below in Table 3.

High/Low	**High**	**Low**
High	19	4
Low	6	12

Table 3. LWL Confusion Matrix

According to the matrix, the LWL algorithm classified 19 of 23 (82.61%) high-performing thesis statements and 12 of 18 low-performing thesis statements (66.67%) accurately using syntactic complexity indices. The precision, recall and F-measure values for this classification were .76, .76 and .75 respectively on average. For the high-performing thesis statements, the same values were .76, .83 and .79. They were seen to be slightly lower for the low-performing thesis statements, being .75, .67 and .71 in the same order.

Information gain ranking list for the LWL algorithm, obtained by means of the InfoGainAttributeEval algorithm is tabulated below in Table 4.

Average Merit	Average Rank	Attribute	
0.157 +- 0.013	1.4 +- 0.49	21	CP/C
0.153 +- 0.017	1.9 +- 0.83	8	CP
0.145 +- 0.012	2.7 +- 0.46	20	CP/T
0.089 +- 0.012	4.5 +- 0.81	14	VP/T
0.083 +- 0.014	5.5 +- 1.02	15	CT
0.079 +- 0.015	6.3 +- 1.42	13	C/S
0.069 +- 0.024	8.5 +- 4.01	17	DC/T
0.053 +- 0.028	10.4 +- 2.01	3	VP
0.056 +- 0.021	11.2 +- 3.43	16	DC/C
0.057 +- 0.021	11.2 +- 2.82	19	CT/T
0.046 +- 0.038	11.4 +- 5.39	12	MLC
0.048 +- 0.018	12.3 +- 1.9	7	C/T
0.044 +- 0.023	12.6 +- 1.85	6	DC
0.026 +- 0.033	13.6 +- 2.06	4	C
0.029 +- 0.036	14.5 +- 5.45	10	MLS
0 +- 0	15.2 +- 2.18	2	S
0 +- 0	15.3 +- 2.57	5	T
0 +- 0	17.4 +- 2.06	23	CN/C
0 +- 0	18.1 +- 2.26	9	CN
0 +- 0	19.8 +- 1.17	18	T/S
0.007 +- 0.021	19.9 +- 4.25	11	MLT
0 +- 0	20.8 +- 0.98	22	CN/T
0.009 +- 0.027	21.5 +- 4.5	1	W

Table 4. Information Gain Ranking List for LWL

As seen in the table, coordinate phrases per clause, the number of coordinate phrases, coordinate phrases per t-unit, verb phrases per t-unit and complex t-units were the attributes from which the largest amount of information was gained in the classification of L2 thesis statement writing performance using syntactic complexity indices. On the other hand, the mean length of sentences, the mean length of t-units and the number of words were the attributes from which the smallest amount of information was gained. No information was gained from the number of sentences, the number of t-units, complex nominal per clause, the number of complex nominals, t-units per sentence and complex nominal per t-units.

The results obtained with balanced data by means of SMOTE are presented below in Table 5.

Algorithm	Accuracy (%)	*KS*
SGD	75.61	0.51
LWL	75.61	0.50
LR	73.17	0.45
SMO	73.17	0.45
Ibk	68.29	0.34
RF	68.29	0.34
RT	65.85	0.28
J48	65.85	0.28
K*	65.85	0.27
NB	60.98	0.16

Table 5. Classification Performance of Different Algorithms with Balanced Data

As seen in Table 5, balancing the data set did not cause a significant change in the performance of the algorithms except for SGD, LR and SMO whose performance increased to some extent. In this dataset, SGD, LWL and LR were the most successful classifiers producing 75.61 (*KS* = .51), 75.61 (*KS* = .50) and 73.17 (*KS* = .0.45) percent classification accuracy respectively. The accuracy values obtained were seen to be 18-20% above the baseline accuracy. In this data set, J48, K* and NB were seen to be the least accurate classifiers, producing accuracy values 5-10% above the baseline.
The confusion matrix for the RF algorithm can be seen below in Table 6.

High/Low	High	Low
High	18	5
Low	5	13

Table 6. SGD Confusion Matrix for Balanced Data

The SGD algorithm could classify 18 of 23 high-performing thesis statements (78.26%) and 13 of 18 (72.22%) low-performing thesis statements (90.51%) accurately using syntactic complexity indices. The weighted average precision, recall and F-measure values for this classification were .76, .76 and .76 respectively. For the high-performing thesis statements, the same values were .78, .78 and .78. They were seen to be slightly lower for the low-performing thesis statements, being .72 for each of the values.
Information gain ranking list for the SGD algorithm, obtained by means of the InfoGainAttributeEval algorithm is tabulated below in Table 7.

Average Merit			Average Rank			Attribute	
0.325	+-	0.023	1.1	+-	0.3	8	CP
0.231	+-	0.047	2.5	+-	0.81	20	CP/T
0.227	+-	0.014	2.6	+-	0.66	21	CP/C
0.166	+-	0.026	4.6	+-	0.66	14	VP/T
0.158	+-	0.046	6.2	+-	4.07	23	CN/C
0.119	+-	0.011	7	+-	1.48	17	DC/T
0.119	+-	0.012	7.3	+-	0.9	3	VP
0.114	+-	0.01	7.7	+-	0.78	15	CT
0.105	+-	0.021	11	+-	3.69	12	MLC
0.101	+-	0.009	11.2	+-	1.47	6	DC
0.101	+-	0.009	11.5	+-	0.5	16	DC/C
0.104	+-	0.01	11.5	+-	2.54	13	CS
0.099	+-	0.033	12	+-	5.16	1	W
0.101	+-	0.009	12.5	+-	1.02	19	CT/T
0.101	+-	0.009	12.8	+-	1.25	7	C/T
0.076	+-	0.012	15.3	+-	2	10	MLS
0.071	+-	0.009	16.3	+-	0.9	4	C
0.048	+-	0.024	18.8	+-	1.66	11	MLT
0.036	+-	0.024	19.4	+-	1.43	22	CNT
0	+-	0	20.6	+-	1.2	2	S
0.01	+-	0.019	21.2	+-	1.6	9	CN
0	+-	0	21.3	+-	0.78	18	T/S
0	+-	0	21.6	+-	1.5	5	T

Table 7. Information Gain Ranking List for SGD

According to the results, the SGD algorithm gained the largest amount of information for the classification of L2 thesis statement writing performance using syntactic complexity indices from the number of coordinate phrases, coordinate phrases per t-unit and coordinate phrases per clause. On the other hand, the smallest amount of information was seen to have been gained by the algorithm for the classification task from the mean length of t-units, complex nominal per t-unit and the number of complex nominals. The number of t-units, t-units per sentence and the number of sentences were seen to have had no contribution to the algorithm for the classification task.

Since both LWL and SGD were found out to have gained the largest amount of information from the number of coordinate phrases, coordinate phrase per t-unit, coordinate phrase per clause and verb phrase per t-unit, how they interacted with high and low L2 thesis statement writing performance was tested by means of multiple t-tests. The findings are presented below in Table 8.

Index	Performance	Mean Rank	*U*	*Z*	*p*	*r*
CP	High	115.50	3082.00	4.885	<.001	.34
	Low	79.67				
CP/T	High	117.38	2824.50	5.289	<.001	.37
	Low	75.93				
CP/C	High	121.12	2312.50	6.495	<.001	.45
	Low	68.51				
VP/T	High	92.69	3245.50	3.997	<.001	.28
	Low	124.96				

Table 8. Mann-Whitney U Test Results for High (n = 137) and Low (n = 69) Score Groups

As seen in the table, L2 thesis statement writing performance differed according to the number of coordinate phrases ($Z = 4.89$, $p < .001$, $r = .34$), coordinate phrase per t-unit ($Z = 5.29$, $p < .001$, $r = .37$), coordinate phrase per clause ($Z = 6.50$, $p < .001$, $r = .45$) and verb phrase per t-unit ($Z = 4.00$, $p < .001$, $r = .28$), indicating small effects. The results indicated that the high-performing group had a higher number of coordinate phrases, coordinate phrase per t-unit and coordinate phrase per clause. On the other hand, verb phrase per t-unit ratio was higher in the low-performing group.

4. Discussion and Conclusion

This study mainly aimed to find out if L2 thesis statement writing performance could be successfully classified using syntactic complexity indices. The results showed that an identical classification accuracy percentage of 75.61, which exceeded the baseline accuracy of 55.60% by 20.01% could be obtained using the Locally Weighted Learning algorithm with the original imbalanced data set and the Stochastic Gradient Descent algorithm with the data set balanced by means of Synthetic Minority Oversampling. Even though the classification accuracy percentages were the same in both imbalanced and balanced data, it was seen that the balanced data set produced negligibly more successful results by classifying one more low-performing thesis statement and one fewer high-performing thesis statement accurately with a Kappa Statistic 1% higher than the imbalanced data set.

Being able to classify high and low performance in L2 thesis statement writing, syntactic complexity, indeed, seems to be an integral part of writing quality as suggested by Biber et al. (2016), Staples and Reppen (2016), Yang et al. (2015) and Casal and Lee (2019). Confirming the findings of those studies, the findings of this study revealed that L2 thesis statement writing performance could be classified in a way that exceeded the baseline accuracy to a considerable extent by means of a model solely based on syntactic complexity.

However, it was seen that 75.61% classification accuracy could not be increased in either imbalanced or balanced data. Even though this result exceeded the baseline accuracy percentage to a considerable extent, it appears that other features of L2 writing performance should also be included in classification models for increased classification accuracy. In this respect, a combination of lexical and syntactic complexity indices may result in a higher level of accuracy in the classification of L2 thesis statement writing performance.

An interesting finding was that the number of coordinate phrases, coordinate phrase per t-unit, coordinate phrase per clause and verb phrase per t-unit provided the largest amount of information to the classifiers in both imbalanced and balanced data sets. Further analyses showed that a higher number of coordinate phrases and higher ratios of coordinate phrase per t-unit and coordinate phrase per clause were present in the high-performing thesis statements. On the contrary, a lower ratio of verb phrase per t-unit was present in the high-performing group in comparison to the low-performing one. Apparently, high-performing L2 writers that produced the thesis statements in the data set resorted to coordination more often than their low-performing peers to join multiple concepts and ideas, which may have increased their performance in writing thesis statements in L2 by allowing them to express their textual interpretations from multiple perspectives. In the same vein, a lower ratio of verb phrases per t-unit in those essays may be indicating that high-performing L2 writers made more extensive use of nominalization to express their ideas, avoiding narration through verb phrases, the overuse of which can be indicative of low performance in literary analysis essays (Uzun, 2016).

For further classification studies regarding L2 thesis statement writing performance, both lexical and syntactic complexity indices can be tested in a similar model to see if higher classification accuracy can be obtained. Moreover, which form of coordination, syndetic, asyndetic or polysyndetic, contributes better to L2 thesis statement writing performance was not investigated in this study. For this reason, further studies can be conducted to find out if a particular type of coordination contributes better to the construct. A higher percentage of classification accuracy in terms of L2 thesis statement writing performance can be used to develop automated feedback provision systems to scaffold learners into higher levels of L2 writing performance. Finally, the thesis statements investigated in this study were extracted from essays manually. An algorithm which tokenizes the sentences in an essay and detects the thesis statement automatically may allow for the analysis of larger data in a shorter amount of time, producing more precise findings.

References

Ai, H. and Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Diaz-Negrillo, N. Ballier, P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249-264). Amsterdam: John Benjamins Publishing Company.

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

Akbani R., Kwek S., Japkowicz N. (2004). Applying Support Vector Machines to Imbalanced Datasets. In J. F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi (Eds.), Machine Learning: ECML 2004. Lecture Notes in Computer Science, 3201. Springer, Berlin, Heidelberg.

Ben-David, A. (2008) Comparison of Classification Accuracy Using Cohen's Weighted Kappa. *Expert Systems with Applications*, 34(2), 825-832. DOI: 10.1016/j.eswa.2006.10.022

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5) 639-669.

Briscoe, T. (2006). *An Introduction to Tag Sequence Grammars and the RASP System Parser*. Technical report, University of Cambridge, Computer Laboratory Technical Report.

Burdick, H., Swartz, C., Stenner, J., Fitzgerald, J., Burdick, D., and Hanlon, S. (2013). Measuring students' writing ability on a computer-analytic developmental scale: An exploratory validity study. *Literacy Research & Instruction,* 52:255–280. doi:10.1080/19388071.2013.812162

Burnard, L. (2005). *Developing Linguistic Corpora: a Guide to Good Practice*, chapter Metadata for corpus work. Oxford: Oxbow Books. `http://ota.ahds.ac.uk/documents/creating/dlc/index.htm.`

Burstein, J., D. Marcu, S. Andreyev, and M. Chodorow (2001). Towards automatic classification of discourse elements in essays. In *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, France (pp. 98–105). Association for Computational Linguistics.

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. Journal of Second Language Writing, 44, 51–62. doi:10.1016/j.jslw.2019.03.005

Casanave, C. P. (1994). Language development in students′ journals. *Journal of Second Language Writing*, 3(3): 179–201. http://doi.org/10.1016/1060-3743(94)90016-7

Christ, O. and Schulze, B. M. (1994). *The IMS Corpus Workbench: Corpus Query Processor (CQP) User's Manual*. University of Stuttgart, Germany.

Clear, J. (1992). Corpus Sampling. In Leitner, G., Ed., *New Directions in English Language Corpora*. Berlin: Mouton de Gruyter.

Coffin, C. (2006). *Historical discourse: The language of time, cause and evaluation*. London, England: Continuum.

Crossley, S. A., and McNamara, D. S. (2014). Does writing development equal writing quality?. A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66–79.

de Oliveira, L. C. (2011). Knowing and writing school history: The language of students' expository writing and teachers' expectations. Charlotte, NC: Information Age.

EAGLES. (1996). *EAGLES: Preliminary Recommendations on Corpus Typology. EAGLES Document EAG|TCWG|CTYP/P*. `http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html.`

Fellbaum, C., Ed. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Halliday, M. A. K., and Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar.* London, England: Hodder.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer, New York, NY

Hunt, K. W. (1970). Do sentences in the second language grow like those in the first? *TESOL Quarterly, 4*, 195–202. https://doi.org/10.2307/3585720.

Koeva, S. and Genov, A. (2011). Bulgarian Language Processing *Chain. In Proceedings of Integration of multilingual resources and tools in Web applications. Workshop in conjunction with GSCL 2011, University of Hamburg.*

Koeva, S., Stoyanova, I., Leseva, S., Dimitrova, T., Dekova, R., and Tarpomanova, E. (2012). The Bulgarian National Corpus: Theory and Practice in Corpus Design. *Journal of Language Modelling*, 0(1):65–110.

Lu, X. (2011). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers′ Language Development. *TESOL Quarterly*, 45(1): 36–62.

Lu, Xiaofei & Ai, Haiyang. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. Journal of Second Language Writing, 29, 16-27.

Lu, Xiaofei (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4):474-496.

Lu, Xiaofei (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers's language development. TESOL Quarterly, 45(1):36-62.

Martin, J. R. (1992). *English texts: System and structure*. Amsterdam, the Netherlands: John Benjamins.

Meurers, D. (2012). Natural language processing and language learning. In Carol A. Chapelle (Ed.), Encyclopedia of Applied Linguistics (pp. 4193–4205). Oxford, UK: Blackwell.

Miller, R. T. and Pessoa, S. (2016). Where's your thesis statement and what happened to your topic sentences? Identifying organizational challenges in undergraduate student argumentative writing. *TESOL Quarterly*, 7(4):847-873.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Neary-Sundquist, C. (2017). Syntactic complexity at multiple proficiency levels of L2 German speech. *International Journal of Applied Linguistics*, 27(1):242-262.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4): 492–518.

Palotti, G. (2014). Revisiting the readability of management information systems journals again. *Research in Higher Education Journal, 15*:77-84

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. Journal of Second Language Writing, 32, 17-35. https://doi.org/10.1016/j.jslw.2016.02.002

Uzun, K. (2016). Developing EAP writing skills through genre-based instruction: An action research. International journal of educational researchers, 7(2), 25-38.

Uzun, K. (2019). Using Regression to Reduce L2 Teachers' Scoring Workload: Predicting Essay Quality from Several Rhetorical Moves. i-manager's Journal on English Language Teaching, 9(3), 24-31.

Uzun, K. (in press). Future prediction of L2 writing performance: A machine learning approach. Journal of Educational Technology.

Uzun, K. (Unpublished Manuscript). Using rhetorical writing frames to enhance negotiated independent construction in L2 writing.

Uzun, K., and Zehir Topkaya, E. (2019). The Effects of Genre-Based Instruction and Genre-Focused Feedback on L2 Writing Performance. *Reading & Writing Quarterly: Overcoming Learning Difficulties.* https://doi.org/10.1080/10573569.2019.1661317

Witten, I. H. and Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). Elsevier, San Francisco, CA

Wolfe-Quintero, K., Inagaki, S. and Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity (No. 17)*. Honolulu, HI: University of Hawaii Press.

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