

A Corpus-based Analysis of Syntactic Complexity measures in the Academic Writing of EFL, ESL, and Native English Master's Students

Maryam Nasserri (University of Birmingham, UK)

'Syntactic Complexity' together with 'Lexical Complexity' are frequently regarded as linguistic subdomains or subsystems related to the concept of 'Linguistic Complexity' (Kuiken, et al., 2010; Szmrecsanyi and Kortmann, 2012) and are used to analyse linguistic performance, proficiency, and development in first and second language acquisition research studies. They are oftentimes linked to the studies which describe or analyse Complexity, Accuracy, and Fluency (CAF measures) as initially proposed by Skehan (1989) and continued in the works of Wolfe-Quintero, Inagaki and Kim (1998), Wigglesworth and Storch (2009), Norris and Ortega (2009), and Tavakoli and Rezazadeh (2014) among others.

In second language writing studies, however, syntactic complexity measures are taken as variables to investigate the roles of task complexity, writing ability, genre, and teaching methods across different proficiency levels, age groups, and developmental periods (Foster and Skehan, 1999; Chandler, 2003; Ellis and Yuan, 2004; Lu, 2011; Ahmadian, 2012; Wu and Ortega, 2013) as well as the impact of pedagogical intervention on the development of linguistic features (Ortega, 2003). Syntactic complexity measures are suggested for use in L2 development bands as placement criteria (Larsen-Freeman, 1978) and reported to be good indicators of the overall development in second language (Arthur, 1979; Ferris, 1994; Lu, 2010), proficiency levels (Wolfe-Quintero, Inagaki, and Kim, 1998), and specifically writing ability (Rafoth and Combs, 1983).

A number of measures have been proposed for assessing syntactic complexity, among which Mean Length of T-unit, Mean Length of Clause, Mean Number of Clauses per T-unit, and Mean Number of Dependent Clauses per Clause are reported to be good indicators of proficiency levels in second language writing (Cooper, 1976; Witte and Davis, 1982; Wolfe-Quintero et al., 1998; Ortega, 2003; Norris and Ortega, 2009; Lu, 2010). Some experts developed coordination and subordination indices and found out that coordination values are higher in beginner learners' writing and subordination values are higher in advanced-level learners' production (Monroe, 1975; Bardovi-Harlig & Bofman, 1989; Lu, 2010).

Many of these studies, however, have focused on only one or a few syntactic measures as determiners of students' performance or proficiency levels, and this has often led to mixed or contradictory results. The present study includes an extended list of syntactic complexity measures based on the Lu and Ai (2015) classification, providing more grounds for the actual differences to be spotted among the participants. What follows is an account of an analysis of such measures in the English academic writing of three groups of university graduates with different first languages in order to detect possible differences in the use and the pattern of various syntactic structures as indices of syntactic proficiency.

Research Questions and Methodology

In the second language and academic writing, the comparative studies which analyse the complexity of syntactic structures of texts produced by non-native vs. native English students, have been of special interest to researchers both in terms of the performance and the proficiency level (Ai and Lu, 2013; Qi, 2014; Lu and Ai, 2015). To find out whether the findings of this study support or reject the previous research studies' results in respect of the differences in English academic texts from L1 and L2 writers, and concerning the Abstract section as a distinct sub-genre of a dissertation, the following research questions are formulated:

1. Are there any significant differences between EFL and ESL students in any of the fourteen measures of syntactic complexity produced in the Abstract section of master's dissertations?
2. Are there any significant differences between EFL and English NS students in any of the fourteen measures of syntactic complexity produced in the Abstract section of master's dissertations?
3. Are there any significant differences between ESL and English NS students in any of the fourteen measures of syntactic complexity produced in the Abstract section of master's dissertations?

Method, Data Analysis, and Findings

This study analyses the Abstract section of master's dissertations in Applied Linguistics and any other EFL-related disciplines written by EFL (English as a Foreign Language), ESL (English as a Second Language), and NS (Native Speakers of English) students in terms of syntactic complexity. The EFL students are all Iranian master's students with varying L1s who studied in various universities in Iran with a centralised curriculum and submitted their dissertations within five years prior to the commencement of this study. The ESL and NS students are all master's students who studied and submitted their dissertations in various universities in the UK; the ESL students are from different nationalities and language backgrounds.

The corpus is drawn from a homogeneous age group (20-40 years old) of female and male students; 50 abstracts from each group respectively (a total of 150 abstracts). The corpus was analysed using the Syntactic Complexity Analyzer (L2SCA), a computational system for the automatic analysis of syntactic complexity developed and reliability-tested by Lu (2010, 2011).

The frequencies of nine syntactic structures as production units are reported and, based on that, fourteen measures of syntactic complexity, under five syntactic categories based on Lu (2015) classification are analysed:

- A. Length of Production Unit: MLS (Mean Length of Sentence), MLC (Mean Length of Clause), and MLT (Mean Length of T-unit)
- B. Sentence Complexity: C/S (Sentence Complexity Ratio)
- C. Amount of Subordination: C/T (T-unit Complexity Ratio), CT/T (Complex T-unit Ratio), DC/C (Dependent Clause Ratio), and DC/T (Dependent Clause per T-unit)

- D. Amount of Coordination: CP/C (Coordinate Phrases per Clause), CP/T (Coordinate Phrases per T-unit), and T/S (Sentence Coordination Ratio)
- E. Degree of Phrasal Sophistication: CN/C (Complex Nominals per Clause), CN/T (Complex Nominals per T-unit), and VP/T (Verb Phrases per T-unit)

The respective values were first subject to the Shapiro-Wilk Normality Test and then analysed via 14 one-way ANOVAs, to find any overall differences between the three groups' production of the mentioned syntactic measures. The *Post-hoc* analysis then was followed for the measures which showed between-group differences in the ANOVA tests.

ANOVA tests with the threshold level of $p = 0.05$ for six measures of MLC, C/S, T/S, CP/T, CP/C, and CN/C did not show overall significant differences among the groups. The remaining eight measures of syntactic complexity showed significant differences between at least one pairwise comparison among the three groups. Table 1 demonstrates the results of three sets of *post-hoc* multiple comparison tests for these eight measures. In this study, a difference is considered statistically significant if all three methods confirm the significance of the difference between the pairwise comparisons.

Table 1. P-values of three *post-hoc* multiple comparison tests for eight syntactic complexity measures across EFL, ESL, and NS groups

Measure	EFL vs. NS	ESL vs. NS	EFL vs. ESL
MLS	HSD = 0.02 Scheffé = 0.03 B & H* = 0.02	-	-
MLT	HSD = 0.00 Scheffé = 0.00 B & H* = 0.00	-	-
VP/T	HSD = 0.00 Scheffé = 0.00 B & H = 0.00	-	-
C/T	HSD = 0.00 Scheffé = 0.00 B & H = 0.00	-	-
DC/C	HSD = 0.00 Scheffé = 0.00 B & H = 0.00	-	HSD = 0.04 B & H = 0.04
DC/T	HSD = 0.00 Scheffé = 0.00 B & H = 0.00	-	-
CT/T	HSD = 0.00 Scheffé = 0.00 B & H = 0.00	-	-
CN/T	HSD = 0.00 Scheffé = 0.00	-	-

	B & H = 0.00		
--	--------------	--	--

*B & H refers to Bonferroni and Holm *post-hoc* method

The findings tabulated in table 1 show that the EFL group produced significantly shorter sentences (MLS) and shorter T-units (MLT) than the NS group, supporting the results of Ai and Lu (2013) and Foster and Tavakoli (2009). The amount of subordination as manifested in the four syntactic structures of C/T, DC/C, DC/T, and CT/ was also proved to be significantly lower in the EFL group compared with the NS group. Similar trends were found in the studies of Ai and Lu (2013) and Foster and Tavakoli (2009) among the NNS and NS groups for such measures of syntactic subordination. Tukey and Bonferroni & Holm tests also detected a statistical difference between the EFL and ESL groups regarding dependent clause ratio measure in this category. Finally, concerning the degree of phrasal sophistication, the EFL group produced far fewer verb phrases and complex nominals per T-unit (VP/T and CN/T) than their NS peers which is again consistent with the results of Ai and Lu (2013) where NNS produced a smaller proportion of complex nominals than the NS group.

The p-values obtained from three sets of *post-hoc* tests confirm that for the above eight measures making the three broad categories of length of production unit (MLS, and MLT), amount of subordination (C/T, DC/C, DC/T, and CT/T) as well as degree of phrasal sophistication (VP/T and CN/T), the EFL group's performance was considerably lower than the NS group, suggesting that the Iranian Master's students production of these syntactic structures need to be markedly improved via relevant academic writing courses. The NS group proved to be syntactically more proficient than both NNS groups overall and the EFL group was the least syntactically proficient group. The results could benefit syllabus designers and curriculum decision-makers of EFL-related disciplines in EFL settings to attend to the syntactic proficiency of graduate students to approximate their proficiency level to that of native English speakers, especially in thesis and dissertation writing.

References

- Ahmadian, M. J. (2012). The effects of careful guided online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research*, 16 (1), 1-21.
- Ai, H. & Lu. X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students writing. In Nicolas Ballier, Ana Díaz-Negrillo, and Paul Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*, pp 249-264. Amsterdam: John Benjamins.
- Arthur, B. (1979). Short-term changes in EFL composition skills. Paper presented at the 13th TESOL Convention, Boston.
- Bardovi-Harlig, K. & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *SSLA*, 11, 17-34.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267-296.

- Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, 69 (5), 176-183.
- Ellis, R. & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *SSLA*, 26, 59-84.
- Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL QUARTERLY*, 28 (2), 414-420.
- Foster, P & Skehan, P. (1999). The influence of source planning and focus of planning on task-based performance. *Language Teaching Research*, 3 (3), 215-247.
- Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59 (4), 866-896.
- Kuiken, F., Vedder, I. & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, I. Vedder & G. Pallotti (Eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research* (EUROSLA monographs series, 1) (pp. 81-99). [S.l.]: European Second Language Association.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL QUARTERLY*, 12 (4), 439-448.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15 (4), 474-496.
- 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, X. & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27.
- Monroe, J. H. (1975). Measuring and enhancing syntactic fluency in French. *The French Review*, 48 (6), 1023-1031.
- Norris, J. M & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30 (4), 555-578.
- 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.
- Rafoth, B. A. & Combs, W. (1983). Syntactic complexity and readers' perception of an author's credibility. *Research in the Teaching of English*, 17 (2), 165-169.
- Skehan, P. (1989). *Individual Differences in Second Language Learning*. London: Edward Arnold.
- Szmrecsanyi, B. & Kortmann, B. (2012). Introduction: Linguistics complexity- Second language acquisition, indigenization, contact. In Kortmann Bernd & Benedikt Szmrecsanyi (eds.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin:de Gruyter, 6-34.
- Shapiro, S. S. and Wilk, M. B. (1965). "Analysis of variance test for normality (complete samples)", *Biometrika* 52: 591-611.
- Tavakoli, M. & Rezazadeh, M. (2014). Individual and collaborative planning conditions: Effects on fluency, complexity, and accuracy in L2 argumentative writing. *The Journal of Teaching Language Skills (JTLS)*, 5 (4), 85-110.
- Wigglesworth, G. & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26 (3), 445-466.

- Witte, S. P. & Davis, A. S. (1982). The stability of T-unit length in the written discourse of college freshmen: A second study. *Research in the teaching of English*, 16 (1), 71-84.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu: University of Hawai'i Press.
- Wu, S. & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46 (4), 680-704.