

Analysis of Syntactic Complexity and L2 Proficiency in Efl Writing

Ana Cristina Lahuerta
University of Oviedo, Spain

Abstract

The present study examines differences in syntactic complexity in English writing among advanced and upper intermediate undergraduates by means of quantitative measures of syntactic complexity. Participants were 200 Spanish undergraduates enrolled in a Degree in English Studies. A total of 101 students had an intermediate level (B1 level according to the CEFR) and 99 had an upper intermediate level (B2 level according to the CEFR). Essays were evaluated by quantitative measures gauging different aspects of L2 complexity. Results indicate that the complexity measures chosen can capture significant differences in writing proficiency when comparing different proficiency levels. The scores on the general quality of the writings and on all syntactic complexity measures increased from B1 to B2 and for all complexity measures the increase was statistically significant.

Keywords: writing, higher education, complexity, EFL, proficiency

1. Introduction

Complexity is an intricate construct, made up of several sub-constructs and components, each of which can be independently assessed (Norris and Ortega, 2009). An attempt to capture this multidimensionality is Bulté and Housen (2012, cited in Bulté and Housen 2014). Bulté and Housen (2012, cited in Bulté and Housen 2014: 43) distinguish between three components of L2 complexity: propositional complexity, discourse-interactional complexity, and linguistic complexity. Linguistic complexity, the focus of the present study, has clearly received the most attention in L2 writing research. It has been defined as “the extent to which language produced in performing a task is elaborate and varied” (Ellis 2003:140). As Bulté and Housen (2012, in Bulté and Housen 2014: 43-44) argue, “linguistic complexity can be investigated at the level of the language system as a whole (or of its major subsystems) and at the level of the individual linguistic features (forms, structures, patterns, rules) that make up such (sub-)systems. The complexity of these structures can in turn be studied from a formal and functional dimension. All these different components and subdimensions of complexity can be studied across various domains of language such as the lexicon, syntax, and morphology”.

The present study focuses on syntactic complexity, understood broadly, as “the range and the sophistication of grammatical resources exhibited in language production”

(Ortega 2015:82). The aim of the present paper is to examine the relationship between syntactic complexity and L2 proficiency among intermediate and upper intermediate undergraduates.

There is a wealth of complexity measures available in the L2 acquisition literature and, as Bulté and Housen (2014: 44) point out, most L2 studies use only one or two complexity measures. As a result, L2 complexity research studies suffer from low content validity (Bulté and Housen 2012, cited in Bulté and Housen, 2014).

Thus, from an approach to complexity that regards it as a highly complex construct, consisting of several components, the present study will incorporate measures that capture different aspects of syntactic complexity: at the sentential, the clausal and the phrasal level.

2. A review of the literature

The importance of syntactic complexity in second language (L2) writing research is now widely acknowledged. Some recent studies have examined the relationship between syntactic complexity and L2 writers' global proficiency (e.g. Ai and Lu, 2013; Ji-young Kim, 2014; Lu, 2011; Mazgutova and Kozrmos, 2015).

Lu (2011) analysed a large number of essays written by Chinese learners of English from different university years using 14 syntactic complexity measures and found out that some of these measures were related to proficiency (complex nominals per clause, mean length of clause, complex nominals per T-unit, mean length of sentence, and mean length of T-unit).

Ai and Lu (2013) found discrepancies in length of production unit, amount of subordination and coordination, and degree of phrasal complexity between the writing of non-native students at both low and high proficiency levels and that of native students.

Ji-Young Kim (2014) showed that more proficient L2 writers produced longer texts, used more diverse vocabulary, and showed the ability to write more words per sentence and more complex nominalizations than less proficient learners did.

Lorenzo and Rodríguez (2014) analysed a corpus of historical narratives of subjects from the third year of secondary education to the second year of post-compulsory secondary education (baccalaureate). The study employed complexity measures, among them the mean length of sentence, mean length of clause, clauses per sentence, verb phrases per T-unit, dependent clauses per clause, coordinate phrases per clause, complex nominals per T-unit, and complex nominals per clause. Results (2014) showed that learners in the lowest grades produced texts that lacked dependent clauses, and coordinate phrases. However, their essays improved and writing became more syntactically complex. The data showed significant advances in mean length of clause, sentence subordination, as well as complex nominals per clause and verb diversity and verb tenses (2014: 70).

Mazgutova and Kozrmos (2015) investigated the development of the lexical and syntactic complexity of two groups of upper intermediate students enrolled on a month intensive pre-session English for Academic Purposes programme. The English language proficiency of one group was slightly lower than that of the other. Mazgutova and Kormos (2015) found out different patterns of improvement for these two sample groups and attributed these differences to L2 proficiency. Students in the lower proficiency group demonstrated development in phrasal elaboration measures (noun modification via adjectives and prepositional phrases, complex nominals in subject position, multiple modifiers after the same noun), and in the subordination-related measures (syntactic structure similarity, conditionals, and relative clauses).

3. The present study

What the studies reviewed above seem to reveal is that L2 proficiency tends to be a relevant influence on syntactic complexity. The present study intends to contribute to this issue examining the validity of syntactic complexity measures as indicators of L2 writers' proficiency. This study examines if there a significant difference in syntactic complexity between upper intermediate (B2 according to the Common European Framework of Reference for Languages or CEFR) and intermediate (B1 according to the CEFR) EFL learners' written production.

4. Method

4.1. Participants

Participants were 200 higher education students enrolled in a Degree in English Studies in a Spanish University. Previous studies on writing accuracy (e.g. Thewissen, 2012, 2013) suggest that writing improvement tends to occur frequently between adjacent proficiency levels. To test this finding in complexity, we decided to examine two adjacent levels: intermediate and upper intermediate. The sample was therefore composed of two groups: 101 students had an intermediate level (B1 level according to the CEFR) and 99 had an upper intermediate level (B2 level according to the CEFR). The students' CEFR level was based on the scores obtained in an Oxford Placement Test carried out as part of a previous research on the average skill level in English of these students.

4.2. Procedure

For the present study, the data come from a written composition activity, which was administered to participants in their own classroom. For the written activity, students had to write on the topic 'Do you think success changes people?' All the participants were given 30 minutes for the writing activity. In this way, both time and topic constraints were controlled in order to make results comparable (Wolfe-Quintero, Inagaki, and Kim, 1998).

4.3. Measures

The essays were evaluated by means of both holistic ratings of writing quality as well as by a selection of quantitative measures gauging different aspects of L2 complexity. For the holistic writing, we followed Connor-Linton and Polio (2014) and used five 05 scales: Content, Organization, Language Use, Vocabulary and Mechanics. The holistic composition score was arrived at calculating the mean of the five scales. Some of the excerpts from the rubrics used for the rating scales are, for example “no major error in word order and complex structures”, “no errors that interfere with comprehension”, “excellent sentence variety” (for a detailed description of the scale, see Connor-Linton and Polio, 2014: 8).

For the analysis of the relationship between the holistic ratings and the quantitative complexity measures, we decided to follow Bulté and Housen (2014: 47) and used two rating scores: (i) the mean total score of all five rating scales and the scores of the scale, and (ii) Language Use. Syntactic complexity did not figure in the Language Use scale, so that this could not affect the possible correlations between the syntactic complexity measures and the holistic ratings.

Regarding the syntactic complexity measures, we examined syntactic complexity as a multi-dimensional construct. The syntactic complexity measures were chosen to gauge complexification at the sentential, the clausal, and the phrasal level of syntactic organisation. Three sets of measures on sentential syntactic complexity were chosen, so that “they represented a different but interrelated aspect of sentence complexity” (Bulté & Housen, 2014: 47). The first set gauged sentence complexity in terms of the mean length of sentential unit in words: mean length of sentence. The second set measured sentence composition in terms of clauses and consisted of three measures: the compound sentence ratio, complex sentence ratio, and the compound-complex sentence ratio. The third set of measures measured sentential syntactic complexification in terms of proposition combining and clause integration strategies: the coordinate clause ratio and dependent clause ratio. As Bulté and Housen (2014: 47) point out, “although these sentential syntactic complexity measures are well tried metrics in L1 and L2 research the measurement of syntactic complexity at phrasal level is only a fairly recent development in L1 and L2 complexity research, and the number of available measures is still limited”. I therefore calculated syntactic complexity at the phrasal level using one measure: mean length of noun phrase (noun-phrases per clause).

5. Results

A statistical analysis was carried out with the program R Development Core Team 2012, 2.15 version.

Results showed that upper intermediate students significantly outsourced intermediate students in the general quality of the compositions (intermediate, $M=2.74$; upper intermediate, $M=4.10$).

Upper intermediate students also outperformed intermediate students in all syntactic complexity measures used. With regard to sentential complexity upper intermediate learners wrote longer sentences (intermediate, $M=15.16$; upper intermediate, $M=16.03$). When we look at sentence composition in terms of clauses and the different combinations of propositions under clausal structures, we observe a significant increase in compound sentences (intermediate: $M=0.15$; upper intermediate: $M=0.20$), in complex sentences (intermediate: $M=0.40$; upper intermediate: $M=0.53$), and in compound complex sentences (intermediate: $M=0.18$; upper intermediate: $M=0.35$). At the level of clause linking, the number of coordinated clauses per sentence and the number of dependent clauses per sentence increased significantly from intermediate ($M=0.09$; $M=0.35$) to upper intermediate level ($M=0.23$; $M=0.51$). Finally, a significant increase in the length of noun phrases was observed (intermediate: $M=0.47$; upper intermediate: $M=0.65$), pointing to increased use of determiners and modifiers of the noun phrase head.

Table 1. Intermediate and upper intermediate writings' syntactic complexity

	Intermediate			Upper intermediate			p-value
	Mean	Median	S.D.	Mean	Median	S.D.	
Holistic composition score	2.74	1.00	2.59	4.10	4.50	2.83	<0.001*
Mean length of sentence	15.16	11.00	4.88	16.03	14.50	6.33	<0.001*
Compound sentence ratio	0.15	0.00	0.24	0.20	0.00	0.25	<0.01*
Complex sentence ratio	0.40	0.00	0.47	0.53	0.45	0.58	<0.001*
Compound complex sentence ratio	0.18	0.00	0.69	0.35	0.00	0.49	0.02*
Coordinate clause ratio	0.09	0.00	0.22	0.23	0.00	0.27	<0.001*
Dependent clause ratio	0.35	0.11	0.93	0.51	0.49	0.49	<0.001*
Noun phrase per clause	0.47	0.48	0.48	0.65	0.57	0.33	0.03*

6. Discussion

This study allowed us to identify how syntactic complexity develops at adjacent proficiency levels (intermediate and upper intermediate). We observed that upper intermediate students outperformed intermediate students in the general quality of the compositions and in all complexity measures of writing used.

Results revealed a syntactic complexity development by means of a significant increase in the length of linguistic units at all levels of syntactic organisation examined,

i.e. phrase and sentence, which agrees with previous studies both on secondary-level and college-level EFL learners (e.g., Ai & Lu, 2013; Bulté & Housen, 2014; Ji-Young Kim, 2014; Lorenzo & Rodríguez, 2014; Lu 2011; Yang et al., 2015).

At the level of sentence composition, there was a significant increase in sentence coordination and subordination from the intermediate to the upper intermediate level.

This finding agrees with the result obtained by Bulté and Housen (2014), who found a significant increase in compound sentences at the end of their college-level students' intensive English language programme.

A significant increase was also found in both clause coordination and subordination in upper intermediate writings. This agrees with the results obtained by Lorenzo and Rodríguez (2014) with secondary-level EFL learners. These authors showed that sentence embeddedness presented higher scores in the higher grades with one index (dependent clause per clause) reaching statistical significance in the last year, 12th grade. Our finding is also consistent with the results obtained by Bulté and Housen (2014) with college level learners: they found that the number of coordinated clauses per sentence increased significantly in the students' writings at the end of the programme.

Regarding syntactic phrasal complexity, the data showed higher scores at higher proficiency levels in mean length of noun phrases. The increase in the length of noun phrases points to an increased use of determiners and modifiers of the noun phrase head and as Lorenzo and Rodríguez (2014: 68) state, "it has been seen as a borderline that marks linguistic adulthood". This increase in noun phrase length agrees with the results obtained in the study carried out by Lorenzo and Rodríguez (2014) on secondary-level learners, as this measure presents consistent higher scores at each grade examined (from 9th to 12th grade). This result does tally the results obtained by Bulté and Housen with college-level learners as well: they found that the length of nominal phrases increased significantly.

The findings of the present study show that from the intermediate to the upper intermediate proficiency stage, coordination, subordination as well as clausal and phrasal elaboration index syntactic development and writing progress. These results partially tally with Mazgutova and Kormos' (2015) findings that both phrasal and clausal complexification grow in parallel for their lower proficiency L2 upper intermediate undergraduate writers. They equally coincide in part with the results obtained by Lorenzo and Rodríguez (2014) that show growth in both subordination and phrasal complexity from the intermediate to the upper intermediate level of proficiency among secondary-level writers.

They nevertheless disagree with the results obtained by two studies on college-level writers: Bulté and Housen (2014) and Crossley and McNamara (2014) found that by the end of the study their upper intermediate learners produced longer, more complex phrases (more modifiers per noun phrase), but fewer subordinated clauses.

In sum, our findings regarding the development of syntactic complexity in a university-level context support the view of an increasing growth in multiple dimensions of complexity, they uncover a progress in writing characterized by increasing length, coordination, subordination and phrasal elaboration as proficiency increases. In our view, this has to do with the two levels of proficiency analysed: from the intermediate to the upper intermediate level of proficiency, writing complexity

may be increasing at all levels as students are in a process of writing maturity in which all the constructs tend to grow at the same time. More research is nevertheless called for to confirm this result with an analysis that includes lexical complexity and accuracy at the same two levels of proficiency.

7. Conclusion

The analysis carried out indicate that the complexity measures chosen can capture significant differences in writing proficiency when comparing different proficiency levels. We can also conclude that syntactic complexity and therefore the writing skill, tends to be consolidated in higher levels of English competence.

It is nevertheless important to acknowledge some limitations of the current study, which may be addressed in future research. Since our study focuses on a rather short period within the entire developmental trajectory of L2 learners (one academic year), the trends we observed may not be representative of the learners' overall long-term L2 development. A longitudinal study spanning longer observation periods will be carried out in the future to confirm the findings of the present study.

8. References

- Ai, H. and Lu, A. (2013), "A corpus-based comparison of syntactic complexity in NNS and NS university students writing", Díaz-Negrillo, N. Ballier, and P. Thompson, *Automatic Treatment and Analysis of Learner Corpus Data*, John Benjamins, Amsterdam, 249–264.
- Bulté, B. and Housen, A. (2012), "Defining and operationalising L2 complexity", Housen, A. Kuiken, F. and Vedder, I., *Dimensions of L2 performance and proficiency – Investigating complexity, accuracy and fluency in SLA*, John Benjamins, Amsterdam, 21– 46.
- Bulté, B. and Housen, A. (2014), "Conceptualizing and measuring short-term changes in L2 writing complexity", *Journal of Second Language Writing*, 26, 42–65.
- Connor-Linton, J., and Polio, C. (2014), Comparing perspectives on L2 writing: Multiple analyses of a common corpus, *Journal of Second Language Writing*, 26, 1–9.
- 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, pp. 66–79.
- Ellis, R. (2013), *Task-based Language Learning and Teaching*, Oxford University Press, Oxford.
- Ji-young, K. (2014), "Predicting L2 Writing Proficiency Using Linguistic Complexity Measures: A Corpus-Based Study", *English Teaching*, 69, 4, 27-51.
- Lorenzo, F. and Rodríguez, L. (2014), "Onset and expansion of L2 cognitive academic language proficiency in bilingual settings: CALP in CLIL", *System*, 47, 64-72.
- Lu, X. (2011), "A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development", *TESOL Quarterly*, 445(1), pp. 36– 62.
- Mazgutova, D. and Kormos, K. (2015), "Syntactic and lexical development in an intensive English for Academic Purposes programme", *Journal of Second Language Writing*, 29, 3– 15.

- Norris, J.M. and Ortega, L. (2009), “Measurement for understanding: An organic approach to investigating complexity, accuracy, and fluency in SLA”, *Applied Linguistics*, 30, 555–578.
- Ortega, L. (2015), “Syntactic complexity in L2 writing: Progress and expansion”, *Journal of Second language Writing*, 29, pp. 82-94.
- Thewissen, J. (2012), *Accuracy across proficiency levels: Insights from an error- tagged EFL learner corpus*. Unpublished doctoral thesis. Université catholique de Louvain, LouvainlaNeuve, Belgium.
- Thewissen, J. (2013), “Capturing L2 accuracy developmental patterns”, *The Modern Language Journal*, 97, 77-101.
- Wolfe-Quintero, K. Inagaki, S. and Kim, H.-Y. (1998), *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. University of Hawaii, Second Language Teaching & Curriculum Center.
- Yang, W., Lu, X. and 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, pp. 53-67.