Applied Linguistics 2020: 41/5: 688–711
doi:10.1093/applin/amz034 Advance Access published on 14 June 2019

Longitudinal Growth Trajectories of Written Syntactic Complexity: The Case of Turkish Learners in an Intensive English Program

[1,*]NIHAT POLAT, [1]LAURA MAHALINGAPPA and [2]RAE L. MANCILLA

[1]School of Education, Duquesne University, 106C Canevin Hall, 600 Forbes Avenue, Pittsburgh, PA, 15282, USA and [2]Instructional Designer, University Center for Teaching and Learning, University of Pittsburgh, Pittsburgh, 820 Alumni Hall, 4227 Fifth Avenue, Pittsburgh, PA 15260, USA

[*]E-mail: polatn@duq.edu

1. INTRODUCTION

Recent research on second/foreign language attainment (L2A) has increased its focus on syntactic complexity, employing increasingly stringent and nuanced methodologies that can capture the multivariate nature of attainment. Arguably, however, most of the work on L2A has focused on the end-state, or the attainment of specific aspects of particular language skills at precise developmental points. One area of current research that may help to account for the complexification inherent in L2A includes longitudinal work that can

provide insights into the developmental patterns of different L2 skills over a certain period of time (Ortega and Byrnes 2008; Verspoor *et al.* 2011; Crossley and McNamara 2014). Such research is timely and critical in that it (i) sheds light on the conflicting research findings about growth patterns of syntactic complexity over time, (ii) contributes to our perception of commonly held theoretical assumptions about L2 development (e.g. pace and linearity), (iii) expands our current understanding of the nature of L2A and its dynamic processes of change, (iv) reevaluates existing theories on possible effects of first language (L1) and intensity and form of instruction on L2A, and (v) identifies patterns and establishes benchmarks for L2 development at different proficiency levels that could potentially inform L2 pedagogy in more nuanced ways. An area of research with such promising contributions is the study of the development of L2 written syntactic complexity (Ortega 2015), which refers to how L2 learners' ability to manipulate morphosyntactic elements to generate more complex combinations changes over time as they become more proficient writers.

This study examines the growth trajectories of written syntactic complexity in an English as a foreign language (EFL) setting—Turkey. It builds on current research on the development of syntactic complexity in a number of ways. First, rather than using experimental data, which is common in current research on the subject matter, we analyzed a corpus of 852 samples produced by college-aged EFL students as part of their routine writing assessments in their English learning program. Given research on ecological validity and research-driven pedagogy (van Lier 2000; Polat *et al.* 2019), nonexperimental data are of particular importance in understanding L2 development. Second, the data in this study were longitudinal rather than cross sectional; the samples were produced at three proficiency levels over a period of three semesters (eight weeks each) by the same students as they progressed from elementary (Level 1) to pre-intermediate (Level 2) and intermediate (Level 3) levels in their studies. The longitudinal aspect of this research allows us to examine growth in the same individuals rather than gauge progress by different students at assumed levels of proficiency. Third, we have taken a multidimensional approach to analyzing the data based on a computerized analytic tool (Lu 2010; Lu and Ai 2015). This approach allows for a more comprehensive analysis and understanding of the development of various indices of L2 syntactic complexity instead of basing levels on fewer, discrete indices.

Fourth, unlike most current research on syntactic complexity, this study involves L2 learners in a content-based intensive program in Turkey, with weekly (24 weeks) classroom instruction varying between 20 and 25 hours, allowing us to establish a baseline in terms of instructional intensity and context at three levels (PL1: 8 weeks, 160–200 hours; PL2: 16 weeks, 320–400 hours; PL3: 24 weeks, 480–600 hours). Additionally, since the participants are native speakers of Turkish, another potential contribution of this study is the investigation of an understudied L1 language, one which is structurally very different from the L2 (English). With the notable exception of Finnish (Ringbom 1987, 2007; Jarvis and Odlin 2000), the majority of studies on the

development of syntactic complexity have involved speakers of typologically similar languages (i.e. German speakers learning English), which has theoretical implications for crosslinguistic issues related to cognitive processing, development, and difficulty in L2A (Ringbom 2007; Ortega 2009; Polat 2016).

In the following, we begin by presenting a review of literature, summarizing research on syntactic complexity in L2A, and the development of L2 written syntactic complexity. After introducing the goals of our study, we describe our research methodology, offering brief overviews about the participants and context, as well as descriptions of the corpus data, syntactic complexity measures, and data analysis procedures. Next, we present descriptive and inferential statistical results for each research question under study. Before summarizing the major conclusions and directions for future research, each result is cautiously interpreted and discussed in terms of its potential contributions to the field.

2. BACKGROUND

2.1 Syntactic complexity in L2A

Syntactic complexity, which refers to the level of variation and sophistication of syntactic forms in L2 output (Lu 2011; Ortega 2015), has gained a great deal of attention in L2A research in the last two decades (Bulté and Housen 2014; Polio and Yoon 2018). Because syntactic complexity is considered a strong indicator of L2 competence, studying it can improve our understanding of a broad range of issues related to L2 learning, use, development, and assessment. According to Ortega (2015), such work helps researchers assess L2 proficiency, frame performance indicators, and identify developmental benchmarks. Nevertheless, as is the case with many other constructs in the field, operationally defining and measuring syntactic complexity in valid and reliable ways are still unsettled for a number of reasons (Pallotti 2009; Bulté and Housen 2012; Polio and Yoon 2018).

The most visible challenge is that the construct of syntactic complexity is based on the combination of two separate and elusive concepts—*syntax* and *complexity*. First, it is difficult to reduce the study of syntax to a few indices and measures. For example, in their taxonomy of L2 complexity where they described the notion in terms of *propositional*, *discourse-interactional* related, and *linguistic* complexity, Bulté and Housen (2012, 2014) argued that 'linguistic complexity can be investigated both at the level of the language system as a whole (or of its major subsystems and layers) as well as at the level of the individual linguistic features (forms, items, structures, patterns, rules) that make up such (sub-)systems' (pp. 43–44). Just as one can study a learner's use of structural rules at the word (morphology) or phrase (syntax) level, a study can also only focus on how 'complex nominals' develop in an L2 over time (Ravid and Zilberbuch 2003).

Second, the use of the term 'complexity' in the study of morphosyntax now requires a more cautious and nuanced approach considering the applications

of the dynamic systems theory (Verspoor *et al.* 2011) (e.g. nonlinearity and unpredictability) to L2 development research. For instance, we should not dismiss the constitutive elements of the parts that make the whole; as such we should not reduce the whole only to the parts by which it is constituted, as 'at every level of development properties emerge that cannot be reduced to those of prior levels' (van Lier 2000: 246). Further, in the context of research on the use, development, and attainment of L2 input, syntactic complexity extends beyond just the linguistic forms, involving other factors, such as learner psychology in input processing (e.g. cognitive load), the environment, and interactional factors (Polat 2016). Thus, an enduring challenge for researchers will be whether to study certain dimensions of syntactic complexity discretely for more in-depth analyses and nuanced pedagogical implications, or to take a multidimensional approach, where syntactic aspects are seen as inseparably intertwined. To ensure a more comprehensive and multidimensional exploration of written syntactic complexity, in this study, we utilize the L2 Syntactic Complexity Analyzer (L2SCA) developed by Lu and his colleagues (Lu 2010, 2011, 2017; Ai and Lu 2013; Lu and Ai 2015).

Finally, researchers have also called into question viewing syntactic complexity in terms of grammatical forms. Indeed, most research on this topic has been based on 'range and complexity' in L2 learner's production of grammatical forms, including studies on effects of tasks or genre on syntactic complexity (Kormos 2011; Johnson 2017; Yoon and Polio 2017). For example, in a special issue of *Journal of Second Language Writing* (Volume 29) devoted to research on L2 writing complexity, Ryshina-Pankova (2015) challenged common 'form-based' definitions of syntactic complexity. Building on the assumptions of systemic functional linguistics (Halliday and Matthiessen 2013), she argued that the level of complexity of L2 output depends largely on the task to which the learner aims to respond (communication purpose). By using only one genre (descriptive) and thereby controlling for possible genre effect on syntactic complexity (for a review, see Yoon and Polio 2017), this study lays a foundation for research on the development of specific syntactic complexity measures over a more lengthy period of time—three semesters of intensive English instruction.

2.2 Development of syntactic complexity in L2 writing

As a flourishing area of interest in the field, there is a plethora of research on the attainment of L2 written syntactic complexity (Johnson 2017) as well as its development (Ortega 2003; Bulté and Housen 2014). Some of this research has examined the attainment and/or use of its specific dimensions in cyber ecologies (Mancilla *et al.* 2017) in different foreign or second language settings (Byrnes *et al.* 2010; Lu 2011), while others have explored variability in certain indices in L2 assessment (Ortega 2003) or development (Verspoor *et al.* 2011; Serrano *et al.* 2012). The main rationale behind the use of writing-based metrics to study L2 development instead of oral output is that, presumably, learners have extended time (heightened attention) and more control over

processing new input, which also enhances their long-term retention (Williams 2012).

Overall, research on L2 written syntactic complexity has concentrated on a variety of aspects and measures. The focus of early research in the late 1970s (Larsen-Freeman 1978) was to identify lexical and grammatical indices to study overall L2 development. Since then, as argued by Bulté and Housen (2014), some research on written syntactic complexity has explored quantifiable 'changes' in selected numbers of indices and measures, concluding that the observed developmental patterns in writing output constitute evidence for a linear growth in a learner's interlanguage (Polat *et al.* 2019). Some scholars have recommended at least a year of instruction to increase the level of complexification (Ortega 2003), although no consensus in the research exists about length of instruction needed for the improvement of L2 syntax. This is especially true for learners with different L1 backgrounds and/or at different proficiency levels due to limited research, a gap this study aims to fill.

As far as the longitudinal development of L2 written syntactic complexity is concerned, there is contradictory evidence in current research. Some of this research has focused on the instructional effectiveness in various L2 programs over the course of imprecise lengths of time, ranging from less than a year to a year and longer. For example, in an early study on the learning of German, Cooper (1976) examined the development of written syntactic complexity of learners at five different proficiency levels, reporting significant increases in the complexification of some comparable indices, such as length of clauses, sentences, and T-units; however, such significant increases were only between levels beyond one year of instruction. Although they used different indices of complexity (e.g. fluency and lexical sophistication) than the framework used in this study (Lu 2010), several other studies have reported linear improvements in the complexity of L2 writing in less than a year (Larsen-Freeman 2006) or a yearlong instructional period (Serrano *et al.* 2012). Note that both the length and intensity of classroom instruction in most previous research on L2 writing (for a meta-analysis, see Sasaki 2007) were shorter than those of the current study (20–25 weekly hours, over 24 weeks; total of 480 and 600 hours of instruction).

Using the same measurement model as the present study, Lu (2011) identified 10 T-unit indices capable of discriminating between grade levels for university ESL students. Nonetheless, only seven developed linearly across proficiency levels, including length of production, coordination, and complex nominalization. In contrast, some previous research has reported a lack of evidence for linear development for different indices of L2 syntactic complexity (Wolfe-Quintero *et al.* 1998; Ortega 2003). For example, Knoch *et al.* (2014) reported no significant increases in levels of written syntactic complexity of ESL learners after a year of study. Indeed, as rate and nature of improvement in syntactic complexity have been linked to the density and quality of

instruction, learning tasks, and so forth (Pallotti 2009; Williams 2012), research on gains in different aspects of syntactic complexity over time is not straightforward.

Since researchers in this area have used a variety of different measures and indices, and there are no replication studies (to our knowledge), it is hard to draw meaningful comparisons here. Despite the fact that some researchers, who utilized longitudinal writing data to explore the dynamic mechanisms of L2 development (Verspoor *et al.* 2008; Spoelman and Verspoor 2010), have suggested that learners do produce more varied and sophisticated written output as they become more proficient over time, each study reported different levels and patterns of development and only for a group of selected indices, due to reasons related to methodological design, study setting, and program of study, among other possible factors. For example, Bulté and Housen (2014) reported significant increases over time in the length of syntactic units at phrase, clause, sentence, and T-units, as well as clause coordination in simple sentences of adult ESL learners. Byrnes *et al.* (2010), who explored the development of certain syntactic complexity indices, reported incremental increases in overall complexity (number of words per T-unit) over time; however, the development of specific indices demonstrated unclear patterns. Likewise, results of Crossley and McNamara (2014) also suggested that even in one semester, some aspects of L2 written syntactic structures became more complex. Their participants produced fewer matrix and embedded clauses and longer noun phrases, as well as a broader range of syntactic structures.

3. CURRENT STUDY

This study investigates the growth patterns of L2 written syntactic complexity. It utilizes a nonexperimental corpus produced by Turkish learners of English over the course of 24 weeks as they progressed from elementary to pre-intermediate and intermediate proficiency levels. Conducting several statistical analyses on a nonexperimental longitudinal data set from the Turkish setting, this study addresses the following research questions:

1 Which indices of syntactic complexity (listed under *length of production, amount of subordination and coordination, degree of phrasal sophistication*, and *overall sentence complexity*) characterize the writing of Turkish learners of English most saliently at the elementary, pre-intermediate, and intermediate levels?
2 How do levels of these indices of syntactic complexity change over time as these learners progress from the elementary to the pre-intermediate and intermediate levels?

4. METHODS

4.1 Participants and context

Participants included 284 students who were learning EFL in an intensive preparatory program that typically required one to one and a half years of instruction. Of the total participants, 146 were female and 138 were male, with an age range of 18–22. In Turkey, students are placed in their undergraduate studies (e.g. business and law) based on their scores on a state-mandated university entrance exam, which is only in Turkish and covers knowledge and skills in content areas, such as math, science, and social studies. Some Turkish universities, like in our setting, offer certain undergraduate programs entirely in English. Once admitted into an English-medium program, students take an English placement test to determine their English proficiency levels. All participants in this study took this test and began their English study in the program at the elementary level (PL1) and progressed into pre-intermediate (PL2) and then upper-intermediate (PL3) levels. They moved into their first year of college after completing the advanced level (PL4). To put into perspective our participants' levels with a familiar scale, we used the Common European Framework of Reference for Languages (CEFR). Based on our understanding of the students' scores on the L2SCA indices (Lu 2010) and CEFR descriptors, our participants were (approximately) at A2 at the end of PL1 (which is when they wrote their first sample that we used for this level), B1 at PL2, and B2 at PL3.

The goal of this preparatory program was to help students build a solid background in basic communication skills and academic language proficiency to become ready to receive their undergraduate studies through the medium of English. This was an intensive program because (i) the weekly amount of classroom instruction varied between 20 and 25 hours, (ii) which continued for three semesters, (iii) and allowed the students to receive uninterrupted instruction on the same curriculum over the course of a year and a half period. The program took a content-based approach that involved all receptive (e.g. reading and listening), productive (e.g. writing and speaking), and complementary (e.g. grammar) skills. At each proficiency level (PL1: 8 weeks, 160–200 hours; PL2: 16 weeks, 320–400 hours; PL3: 24 weeks, 480–600 hours), students received 8 weeks of classroom instruction, during a total of 24 weeks. The teaching of writing skills constituted around 10–15 per cent of the overall classroom instructional time at the beginning of the program, with that time gradually increasing as the students' overall proficiency level increased. In addition, the focus of writing instruction increasingly involved more academic skills and tasks as the students moved from elementary to higher levels.

The institution was a private university in a large city in Turkey. The English preparatory program had written policy about class size, number of credit hours, instructor qualifications, and rotation of instructors across classrooms.

All students used the same textbooks and instructional resources and completed each grade level as a cohort. At the completion of each proficiency level, a weighted average score was calculated based on weekly quizzes and final achievement tests (e.g. level exit test) the students had taken in the program; afterward a decision was made whether the student should move onto the next level or repeat the level with the next cohort. Rather than allowing teachers to prepare their own tests based on their classroom instruction, a testing office standardized, prepared, and administered all exams to make sure that curriculum and assessment goals were aligned and that all the students met the intended goals. Therefore, each writing sample used comes from the same standardized task, with specific test administration requirements (e.g. testing time).

4.2 Description of corpus

The corpus used in this study was nonexperimental in that the samples collected were part of students' routine educational experiences, not produced under experimental conditions. Data included three paragraphs that were collected from each student at different points in the cohort: one from elementary (Level 1), one from pre-intermediate (Level 2), and one from intermediate level (Level 3). Thus, the corpus used in this study included 852 (284 × 3) writing samples. The total number of words was 32,375 (M = 114; SD = 41.3) for PL1, 38,623 (M = 136; SD = 53.1) for PL2, and 44,871 (M = 158; SD = 68.6) for PL3 essays. After receiving approval from the university, students' writing samples were obtained from the university's testing office. Although the office stored all students' graded tests, researchers were only given access to writing samples from three cohorts of students. These samples were then de-identified for anonymity purposes. To maximize homogeneity of the corpus in levels of syntactic complexity (Ortega 2003; Ortega and Byrnes 2008; Lu 2011), we only used writing samples that were produced by the same participants at different time points (repeated measure), rather than using cross-sectional samples with different tasks written by different participants at different proficiency levels.

Each writing assignment was part of a test that assessed multiple other language skills (e.g. reading and listening) in line with the goals of the curriculum during the semester. All writing samples were written under testing conditions, where the learners had a total amount of time for the test as a whole. Specifically, these samples included paragraph-level responses from the last tests the students took at each proficiency level to ensure that the samples represented students' highest writing ability at a level before transitioning to the next. The length of the paragraph was not restricted by the number of words or sentences for any of the three samples. To control for possible effects of task/genre on complexity (Kormos 2011; Tavakoli and Foster 2011; Yoon and Polio 2017), we selected samples from only one genre: descriptive. Another reason for this was availability; students had written descriptive

samples at all levels, but did not write, for example, argumentative ones at the elementary level. As regulated by the testing office, the task in all three samples was stated as: 'Write a paragraph about . . .' your 'best friend' (PL1: elementary), 'hometown' (PL2: pre-intermediate), and 'favorite movie' (PL3: intermediate).

Although participants also underwent advanced-level studies before progressing on to their first year in college, writing samples at this level were not used in this study for two reasons. First, the advanced level corpus was inconsistent with the other data in terms of the writing tasks, and second, they involved specialty-based foci and differentiation to prepare the students for academic writing in their respective undergraduate disciplines (e.g. engineering and psychology).

4.3 Syntactic complexity measures

Previous research on the measurement of L2 syntactic complexity has thus far focused on a broad set of indices (Norris and Ortega 2009; Lu 2011; Bulté and Housen 2012, 2014; Lu 2017), using various tools (e.g. Coh-Metrix: McNamara *et al.* 2014). In this study, we have utilized the L2SCA by Lu (2010). This computerized tool automatically calculates numeric values for the corpora inserted into the system for 14 complexity indices categorized around five measures of *length of production unit*, *amount of subordination*, *amount of coordination*, *degree of phrasal sophistication*, and *overall sentence complexity* (Table 1). As demonstrated in Table 1, *length of production* is constituted by indices that measure mean length of (i) clause, (ii) sentence, and (iii) T-unit. *Amount of subordination* is measured by the number of (iv) clauses per T-unit (CT), (v) complex T-units per T-unit (CTT), (vi) dependent clauses per clause (DCC), and (vii) dependent CT (DCT). *Amount of coordination* measures include number of coordinate phrases per (viii) clause, (ix) T-unit, and (x) sentence. While *phrasal sophistication* is measured by the number of complex nominals per (xi) clause, (xii) T-unit, and (xiii) verb phrases per T-unit (VPT), *overall sentence complexity* is measured by (xiv) the number of clauses per sentence (CS). For more detailed information about L2SCA's computation procedures (e.g. frequency of words, phrases, clauses, sentences, and T-units), see Lu and his colleagues' work (Lu 2010, Lu and Ai 2015; Ai and Lu 2013). While, like other research tools, the use of L2SCA is not without controversy, it has been tested and reported to be highly reliable. Please see Lu (2010) and Polio and Yoon (2018) for more information about its reliability.

We have used this framework for three main reasons. First, it takes a multidimensional approach. From a measurement perspective, taking such an approach in studying syntactic complexity is a critical mandate related to the issue of content validity. The fact that syntactic complexity is a multidimensional phenomenon and each of these dimensions is constituted by a set of underlying indices is a commonly acknowledged stance in the field (Bulté and Housen 2012; Lu and Ai 2015; Polio and Yoon 2018). Note that there is a lack

Table 1: Descriptive statistics for syntactic complexity measures by Level 1 (Elementary), Level 2 (Pre-intermediate), and Level 3 (Intermediate)

Indices	*N*	Level 1 M (SD)	Level 2 M (SD)	Level 3 M (SD)
Length of production unit				
MLC	284	4.60 (1.85)	4.91 (1.68)	5.49 (1.84)
MLS	284	7.13 (1.75)	7.61 (1.92)	11.4 (3.19)
MLT	284	6.90 (1.64)	7.10 (2.21)	8.21 (2.11)
Amount of subordination				
CT	284	0.85 (0.22)	0.87 (0.21)	1.31 (0.24)
CTT	284	0.29 (0.07)	0.30 (0.08)	0.32 (0.08)
DCC	284	0.20 (0.04)	0.23 (0.06)	0.26 (0.10)
DCT	284	0.21 (0.06)	0.23 (0.08)	0.48 (0.14)
Amount of coordination				
CPC	284	0.18 (0.04)	0.20 (0.07)	0.12 (0.03)
CPT	284	0.20 (0.09)	0.18 (0.07)	0.11 (0.04)
TS	284	1.03 (0.23)	1.04 (0.16)	1.05 (0.09)
Degree of phrasal sophistication				
CNC	284	0.75 (0.19)	.78 (0.22)	0.91 (0.18)
CNT	284	0.94 (0.34)	.95 (0.39)	0.97 (49)
VPT	284	1.43 (0.46)	1.46 (0.45)	2 (0.50)
Overall sentence complexity				
CS	284	1.23 (0.44)	1.27 (0.42)	1.44 (0.56)

of consensus about the constitutive elements of syntactic complexity, with previous studies utilizing a varying number of different measures (Housen and Kuiken 2009; Ortega 2015; Lu 2017). Since this study does not propose directional hypotheses that certain indices predict the development of L2 written syntactic complexity better than others, utilizing a broad range of measures allows for a more comprehensive exploration. A growing body of studies (Ai and Lu 2013; Lu and Ai 2015; Polio and Yoon 2018) has addressed the accuracy of the L2SCA as a means of capturing the full trajectory of complexity for non-native English-speaker writers. Second, there have been several recent studies that have utilized this model in L2 writing (Ai and Lu 2013; Lu and Ai 2015; Mancilla *et al.* 2017), which allows for more direct comparisons and interpretations of our results through a publically available, reliable model. Finally, given the size of our corpus (852 writing samples), the use of a computerized tool such as this allowed us to conduct numerous multivariate analyses objectively (without involving human raters) to document change over time across three different proficiency levels.

4.4 Analytic procedures

Both descriptive and inferential statistics were computed for each question under study using the statistical software *R*. Mean scores and standard deviations were calculated for all 14 syntactic complexity indices across the three proficiency levels. To identify the syntactic complexity indices that uniquely characterize the elementary, pre-intermediate, and intermediate levels (question 1), we generated profile plots and curves (Figures 1 and 2). Based on participants' mean scores on these 14 indices, the profile plots help visualize patterns of syntactic structures at the measure level (e.g. degree of phrasal sophistication), whereas the profile curves visually demonstrate participants' means for each complexity index (e.g. number of complex nominals per clause (CNC)) in log scale across the three proficiency levels. To determine if any change occurred over time in levels of indices of syntactic complexity (question 2), we employed the nonlinear mixed effect with the function generalized least squares package because it allows for flexibility in modeling the covariance structure with repeated measure data. In this procedure, we modeled each syntactic complexity measure as a function of Level to determine how they change over time (e.g. from PL1 to PL2 and then PL3). In other words, for each measure, we performed pairwise comparisons between levels: L1 versus L2, L1 versus L3, and L2 versus L3. For example, for mean length of sentence (MLS), we used the model MLS $\sim$ Level and made comparisons between L1 versus L2, L1 versus L3, and L2 versus L3 values.

To determine a better fit, we considered all four var-cov (variance-covariance) structures in our mixed effects modeling and labeled the models as *fit.cs* for compound symmetry, *fit.un* for unstructured *var-cov* structure, *fit.ar1* for autoregressive first order, and *fit.arh1* for autoregressive heterogeneous variances. Based on initial analyses, the most promising structure was the unstructured var-cov because the corresponding model had the lowest AIC and BIC and (at each comparison) it had a significant improvement over the other test with a significant LRT *p*-value (Supplementary Material). Moreover, unstructured var-cov assumes that each variance and covariance is unique, with each trial and pair of trials having their own (co)variance structured as σ_1^2, σ_{21} σ_2^2, σ_{31} σ_{32} σ_3^2.

5. FINDINGS

5.1 Indices of syntactic complexity characterizing L2 writing at PL1, PL2, and PL3

To identify indices of syntactic complexity that most dominantly characterize L2 use at each proficiency level, we report descriptive statistics (Table 1) and profile plots. Here, we describe our findings for the complexity measures, which are based on the data that are not observer-dependent (computerized), and hence objective in nature. The Everitt's Chi-square quantile plot

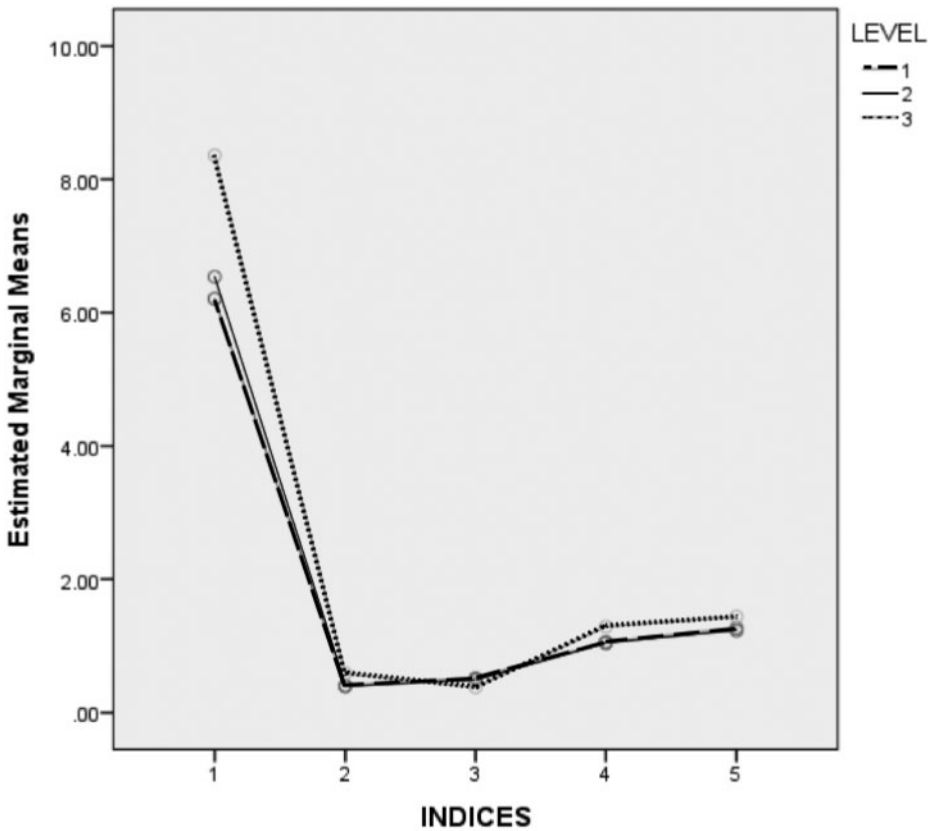

Figure 1: Profile plot indicating combined means for the five syntactic complexity dimensions across the three proficiency levels.
Notes*: 1 = length of production; 2 = amount of subordination; 3 = amount of coordination; 4 = degree of phrasal sophistication; and 5 = overall sentence complexity*

(Supplementary material) indicated that the data, in all three levels, conform to multivariate normality, with a concave down trend at the upper end.

To identify patterns at the measure level, we calculated means (from indices) across the three levels. The profile plot (Figure 1) demonstrates that *length of production unit* has the highest values at all three levels, while the other four measures formed a flat line constituting a pattern similar to an L-shape. With the exception of *coordination*, at all three levels, *subordination, phrasal sophistication*, and *overall sentence* complexity values are higher for upper levels, revealing a PL3 < PL2 < PL1-pattern. Table 1 shows that the lower bound writing samples shared many similarities in levels of syntactic complexity across different measures. At PL1 and PL2, EFL learners produced writing

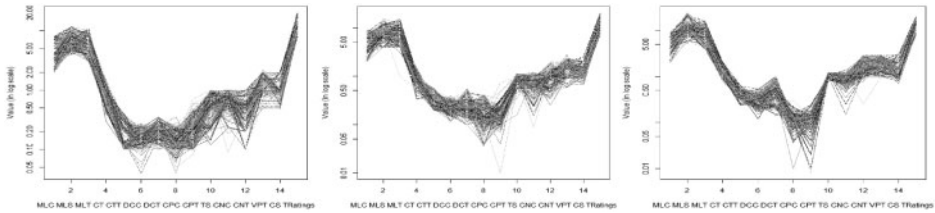

Figure 2: Profile curves indicating participant data for the 14 complexity indices in log scale across L1, L2, and L3

output at the highest level of syntactic sophistication in the area of *coordination*. On the other hand, more *subordination*, higher *degrees of phrasal sophistication*, and longer *production units* seemed to characterize PL3 most visibly.

As for SC patterns of the 14 indices, descriptive data in all three levels show a clear scale difference between the variables without skewness in measures. As presented in the profile curves (Figure 2) in log scale, there is a U-shaped trend (for the given order of the variables) in PL1 and PL2, whereas the pattern resembles more a V-shape trend in PL3.

Specifically, index-level analyses revealed that at PL1, participants produced the highest number of coordinate phrases per T-unit (CPT) and the lowest MLS, whereas at PL2, they produced the highest number of coordinate phrases per clause (CPC) and the lowest number of CPT. In contrast, the highest number of VPT and DCT and the lowest number of CPT and CPC most saliently characterized participants' writing at PL3.

5.2 Changes in indices of syntactic complexity in L2 writing over time

Using pairwise comparisons between levels (from PL1 to PL2 and PL3), we explored possible changes in complexification of SC measures over time. Again, we used the unstructured var-cov (variance-covariance) structure, but modeled each SC measure as a function of Level. For example, for mean length of clause (MLC), we constructed an MLC ~ PLevel model and compared MLC values between PL1 and PL2, PL1 and PL3, and PL2 and PL3. Overall, significant changes were observed in all measures; however, the trends were different. We observed that variables MLS ($F = 42.8$, $p < 0.01$), MLC ($F = 20.2$, $p < 0.01$), mean length of T-unit (MLT) ($F = 29.6$, $p < 0.01$), CTT ($F = 84.9$, $p < 0.01$), DCC ($F = 78.1$, $p < 0.01$), DCT ($F = 42.3$, $p < 0.01$), T-units per sentence (TS) ($F = 96.5$, $p < 0.01$), CNC ($F = 26.6$, $p < 0.01$), complex nominals per T-units (CNT) ($F = 98.3$, $p < 0.01$), and CS ($F = 20.1$, $p < 0.01$) followed stable upward trends in complexification as proficiency levels increased (PL3 > PL2 > PL1) from PL1 to PL3. Below, we present these results for each measure in the same order they are presented on the L2SCA (Lu 2010).

Specifically, all three *length of production unit* indices followed the same pattern. PL3MLC values were significantly larger than both PL1MLC ($z = 6.42$, $p < 0.01$) and PL2MLC values ($z = 4.66$, $p < 0.01$), and PL2MLC values were significantly larger than PL1MLC values ($z = 3.11$). Similarly, PL3MLS values were significantly larger than both PL1MLS ($z = 20.5$, $p < 0.01$) and PL2MLS values ($z = 16.25$, $p < 0.01$), and PL2MLS values were significantly larger than PL1MLS values ($z = 4.31$, $p < 0.01$). Likewise, PL3MLT values were significantly larger than both PL1MLT ($z = 13.44$, $p < 0.01$) and PL2MLT values ($z = 8.65$, $p < 0.01$), and PL2MLT values were significantly larger than PL1MLT values ($z = 7.34$, $p < 0.01$).

Except for CT, all other *subordination* indices also demonstrated a PL3 > PL2 > PL1-pattern. PL2CTT values were significantly larger than PL1CTT values ($z = 16.8$, $p < 0.01$), PL3CTT values were significantly larger than PL1CTT values ($z = 29.7$, $p < 0.01$), and PL3CTT values were significantly larger than PL2CTT values ($z = 14.8$, $p < 0.01$). Likewise, PL2DCC values were significantly larger than PL1DCC values ($z = 15.7$, $p < 0.01$), PL3DCC values were significantly larger than PL1DCC values ($z = 27.9$, $p < 0.01$), and PL3DCC values were significantly larger than L2DCC values ($z = 13.4$, $p < 0.01$). Similarly, PL2DCT values were significantly larger than PL1DCT values ($z = 3.93$, $p < 0.01$), PL3DCT values were significantly larger than PL1DCT values ($z = 21.0$, $p < 0.01$), and PL3DCT values were significantly larger than PL2DCT values ($z = 19.0$, $p < 0.01$). Of the three *coordination* indices, only one followed this development trend. Specifically, PL2TS values were significantly larger than PL1TS values ($z = 18.2$, $p < 0.01$), PL3TS values were significantly larger than PL1TS values ($z = 30.7$, $p < 0.01$), and PL3TS values were significantly larger than PL2TS values ($z = 11.9$, $p < 0.01$).

In the area of *degree of phrasal sophistication*, two indices demonstrated a PL3 > PL2 > PL1-pattern, while one (VPT) did not. In particular, PL2CNC values were significantly larger than PL1CNC values ($z = 4.84$, $p < 0.01$), PL3CNC values were significantly larger than PL1CNC values ($z = 12.6$, $p < 0.01$), and PL3CNC values were significantly larger than PL2CNC values ($z = 8.39$, $p < 0.01$). Similarly, PL2CNT values were significantly larger than PL1CNT values ($z = 20.1$, $p < 0.01$), PL3CNT values were significantly larger than PL1CNT values ($z = 30.3$, $p < 0.01$), and PL3CNT values were significantly larger than PL2CNT values ($z = 15.5$, $p < 0.01$). As for the *overall sentence complexity* measure, CS, PL2 values were also significantly larger than PL1 values ($z = 2.61$, $p < 0.05$), PL3 values were significantly larger than PL1 values ($z = 6.41$, $p < 0.01$), and PL3 values were significantly larger than PL2 values ($z = 4.67$, $p < 0.01$).

Nonetheless, the developmental patterns of the remaining four indices seemed to be rather different. For example, both the CT (subordination index) and VPT (phrasal sophistication index) values showed significant difference over time (i.e. Level), but neither PL2CT/PL2VPT nor PL1CT/PL1VPT values (lower bounds) were significantly different from each other. Yet, PL3CT/PL3VPT values were significantly larger than PL1CT/PL1VPT values

($z = 22.4$, $p < 0.01$/$z = 12.7$, $p < 0.01$), and PL3CT/PL3VPT values were significantly larger than PL2CT/PL2VPT values ($z = 21.5$, $p < 0.01$/$z = 11.3$, $p < 0.01$), which revealed a PL3CT/PL3VPT > PL2CT/PL2VPT ≈ PL1CT//PL1VPT-pattern.

For CPC and CPT indices (*coordination* indices), developmental patterns were mixed. For CPC, we observed a significant increase from PL1 to PL2 ($z = 5.82$, $p < 0.01$); however, there was a significant decrease from PL2 to PL3 ($z = -10.8$, $p < 0.01$). In fact, PL3CPC values were even significantly smaller than PL1CPC values ($z = -8.48$, $p < 0.01$), which yielded this pattern: PL3CPC < PL1CPC < PL2CPT. As for CPT, we observed that the values decreased as proficiency levels increased, which is a mirror image to the first 10 indices. Specifically, PL2CPT values were significantly lower than PL1CPT values ($z = -8.53$, $p < 0.01$), PL3CPT values were significantly lower than PL1CPT values ($z = -16.2$, $p < 0.01$), and PL3CPT values were significantly larger than PL2CPT values ($z = -10.7$, $p < 0.01$), resulting in the following pattern PL3CPT < PL2CPT < PL1CPT.

6. DISCUSSION

6.1 Indices of syntactic complexity characterizing different proficiency levels

As helpful visual representations of how these indices characterize each proficiency level, we created profile curves (Figure 2) that revealed a U-shaped trend (for the given order of the variables) in PL1 and PL2 and a V-shape trend in PL3. Overall, our analyses of EFL writers at three stages of language proficiency have shown that learners generally progress in their syntactic sophistication following a pattern similar to Norris and Ortega's (2009) three-phase model; however, the pace of development seemed rather slow despite the intensive nature of the program. Specifically, coordination emerged as the most salient syntactic feature at both the elementary and pre-intermediate levels. This finding agrees with a number of studies that have documented the importance of clausal coordination for beginning L2 writers (Bardovi-Harlig 1992; Wolfe-Quintero *et al.* 1998; Lu 2011; Bulté and Housen 2014).

Unlike these studies however, our analyses provide insight into the development of L2 syntax at intermediary stages (between elementary and intermediate proficiency). For example, here, we observed that PL1 and PL2 writers were markedly similar and practically indistinguishable as they continued to rely on phrasal coordination strategies, such as the increased use of CPT and CPC after a year, underscoring the slow growth rate of complexity. This finding supports the idea that short-term development studies, in our case, approximately 160–200 hours (PL1 = 8 weeks), or for most measures even 320–400 hours (PL2 = 16 weeks) of classroom instruction, cannot fully capture L2 writers' syntactic growth (Cooper 1976). It also helps to explain why some previous studies reported no meaningful increases in complexity levels (Knoch *et al.* 2014).

Nonetheless, given the amount of classroom instruction (20–25 hours weekly) the participants received during the two semesters (16 weeks, 320–400 hours), we expected more improvement from PL1 to PL2. Among other possible reasons, the slow pace of development in these indices could be attributed to the participants' L1 background, particularly at PL1 and PL2, since crosslinguistic effects may be stronger at lower proficiency levels (Ortega 2009). Based on previous research on crosslinguistic influences, morphosyntactic similarities between L1 and L2 may have an impact on L2A (Ringbom 1987, 2007). Turkish, with an agglutinative morphosyntactic system allowing for the generation of remarkably long, morphologically complex words, shares few similarities with English, which relies on syntactic structures to achieve the same meanings. Although there is no previous research on Turkish learners of English on this topic, the research on Finnish (Ringbom 1987, 2007; Jarvis and Odlin 2000) versus Swedish learners of English found that morphosyntactic similarities between Swedish and English allowed Swedish speakers to outperform Finnish speakers in English writing tasks. This implies that lack of commonalities between the L1 and L2 can lead to a slower pace of the acquisition of particular morphosyntactic features.

In our study, participants continued to use more coordination at PL2 rather than proceeding to subordination. In Turkish, subordinate clauses are basically nominalizations and sometimes subordinate connecters are not overtly marked. In addition, Turkish derivations and inflections do not allow for single units as they do in English (single syntactic structures). For example, the Turkish word *geleceklerdide*, translates into English as a subordinate clause, meaning 'if they were going to come'. In this example, the morpheme /-de/ is the marker of subordination, which is affixed at the end of the word. Moreover, in their study on syntax-to-morphology mapping (for machine translation: Turkish-English), Yeniterzi and Oflazer (2010) suggest that the preprocessing of some English morphosyntactic structures requires 'rather long distance relations in the source side (Turkish) syntactic structure' (p. 462). Thus, potential influences of such differences may not be due to language knowledge only, but rather includes cognitive functions (in constructing meaning) since subordination involves complex underlying connections between tasks and events.

Intermediate writers (PL3: 24 weeks, 480–600 hours) on the other hand, produced a greater variety of syntactic constructions at multiple levels of syntactic organization: sentential, phrasal, clausal, and T-unit. At this level, the learners increased the length of their T-units (MLT) and improved the overall level of sentential complexity (CS), which can be expected, since length of production has been associated with more proficient writing at around B2 or higher levels on the CEFR scale, at least at nonadjacent levels (Ortega 2003) and incremental increases across grade levels (Lu 2011; Ai and Lu 2013). Consistent with Norris and Ortega's (2009) claim that subordination becomes the dominant means of complexification at the intermediate level, PL3 writers used more CT and DCT. Unlike when they were at PL1 and PL2, the

participants at PL3 traded off between coordination and subordination upon reaching the intermediate and upper-intermediate proficiency levels. These results are similar to those of Bulté and Housen's (2014) whose participants were also at B1 or B2 on the CEFR scale. Finally, compared to their writing at PL1 and PL2, at PL3 learners showed some signs of enhanced phrasal sophistication, producing significantly more CNC and VPT, despite Norris and Ortega's (2009) use of phrasal sophistication as an indicator of advanced L2 proficiency—this could be attributed to the intensive amount of classroom instruction the participants had received at this level (24 weeks, 480–600 hours). This early tendency toward phrasal complexification has been attributed to the importance of nominalization in academic writing in some previous research (Crossley and McNamara 2014; Mancilla *et al.* 2017).

6.2 Developmental patterns of indices of syntactic complexity over time

Our findings revealed that over time, the EFL learners experienced significant syntactic changes in multiple aspects of complexity; however, these changes were not necessarily straightforward or incrementally distributed across the three levels. Our results can be summarized in three change-trends over time: (i) 10 indices showed significant growth trends in complexification as proficiency levels increased (PL3 > PL2 > PL1) from PL1 to PL3; (ii) two indices (CT and VPT) demonstrated a PL3 > PL2 ≈ PL1-pattern, with no difference at lower bounds (PL1 and PL2), but yielded significant growth at the upper bounds (PL2 and PL3) and between the lowest and highest levels (PL1 and PL3); and (iii) two indices showed mixed patterns, with CPC values increasing from PL1 to PL2 but decreasing from PL2 to PL3 (PL3CPC < PL1CPC < PL2CPC) and CPT values decreasing as proficiency levels increased (PL3CPT < PL2CPT < PL1CPT).

Regarding *length of production*, pairwise comparisons indicated that levels of sophistication in students' writing regarding length of sentences (MLS), T-units (MLT), and clauses (MLC), as well as overall sentence complexity (CS), steadily grew from PL1 to PL2 and to PL3. This finding seems to confirm research showing consistent incremental increases in these three indices of elongation across consecutive instructional levels (Lu 2011), but does contradict Cooper's (1976) findings that complexification in the area of length may require at least one year to achieve visibility. This could be due to the fact that our participants received more intense classroom instruction than what has been reported in most previous research (Crossley and McNamara 2014), which suggests that the use of 'a semester', 'a year', or other vague descriptions may not be helpful in estimating L2 development; rather, researchers should specify the number of instructional hours. In our case, in their intensive English program, even by the end of PL2, our participants received between 320 and 400 hours of uninterrupted classroom instruction over the course of 16 weeks, which is more than any other study of which we know. Moreover, the significant differences

we observed between levels PL1 and PL3 regarding MLS, MLT, and MLS are consistent with Ai and Lu's (2013) corpus comparison of high- and low-level EFL learners, in which the values of MLS, MLT, and MLC all increased as writers grew in proficiency. Like Ai and Lu (2013), who reported that high-level EFL learners approximated native speakers in terms of length of T-units and sentences, at the highest level (PL3), our participants produced longer T-units and sentences than previous levels. This result echoes Byrnes *et al.*'s (2010) work with German L2 learners who produced incremental growths in overall complexity (number of words per T-unit) over time.

Three of the four s*ubordination* indices also demonstrated stable growth over time. The number of CTT, DCC, and DCT increased incrementally. These results are somewhat different from those of Ai and Lu (2013), who noted slight descriptive increases in DCC and DCT among EFL learners of high and low proficiency with no statistical significance. They also differ from Crossley and McNamara's (2014) findings that writers used fewer subclauses in descriptive essays after a single academic term, but complemented Lu (2011), who reported significant increases in the amount of subordination employed by nonnative speaker writers in their argumentative essays from their first to second years of study. Overall, this means that the semester long 20–25 weekly hours of intensive classroom instruction, and possibly the content-focused form of instruction that aimed to prepare these students for English-medium college education in different academic fields (e.g. engineering and social sciences), contributed substantially to the development of subordination indices, a finding that seems at odds with Bulté and Housen (2014), who found no differences in subordination. Nonetheless, these findings are challenging to interpret given the mixed literature on the relationship between subordination and L2 proficiency, partially due to differences between existing studies concerning the nature of programs, heterogeneity in participant samples, and intensity of instruction.

As for the number of CT, while the values also increased from PL2 to PL3 and PL1 to PL3, no such increase occurred between PL1 and PL2. This result supports Norris and Ortega's (2009: 563) argument that subordination may be 'a useful index of complexification at intermediate levels'. These differences between levels (PL1–PL3) also parallel trends alluded to by both Byrnes *et al.* (2010) and Vyatkina (2012), where the number of CTs increased across adjacent instructional levels. Taken together it seems that the complexification of subordination indices also occurs rather slowly and takes a period beyond one year of study, depending on length and intensity of the instruction.

In the area of *coordination*, growth patterns were rather mixed. While number of TS incrementally increased from PL1 to PL3, learners' use of CPC increased from PL1 to PL2, but decreased from PL2 to PL3 and from PL3 to PL1. The number of CPT, on the other hand, steadily decreased from PL3 to PL1. Some of these findings are congruent with Lu's (2011) conclusion about these indices discriminating nonadjacent, versus adjacent, instructional levels and suggest a certain degree of developmental latency at the lower levels of

proficiency. This seems to be the case for coordination, as similar findings have been reported for the mean values of CPT and CPC between NNS groups of low and high levels of proficiency and native speakers, but not within NNS groups (Ai and Lu 2013).

In reference to *phrasal sophistication*, participants demonstrated consistent growth in their use of CNC and use of CNT from PL1 to PL3. Like Crossley and McNamara (2014), we observed changes at the clause level in the use of CNC between levels PL1 and PL3, which also agrees with several studies that have found that L2 writers rely more on nominalization as they increase in proficiency (Ai and Lu 2013; Bulté and Housen 2014; Mancilla *et al.* 2017). While VPT levels of complexity increased from PL2 to PL3 and PL1 to PL3, our results revealed no difference between the numbers of VPT at PL1 versus PL2, possibly due to L1 influence at these levels. This finding confirms previous research that the use of VPT—as a developmental marker of complexity—is controversial because of its weak correlation with L2 proficiency (Cooper 1976; Lu 2011).

Moreover, our result that VPT grew between two adjacent levels—pre-intermediate and intermediate—differs from some recent findings by Crossley and McNamara (2014), who noted that learners used fewer verb phrases over time. We believe that this could be due to two possible reasons, among others. First, Crossley and McNamara's (2014) results were based on a sample of 57 participants who received approximately 45 hours of total instruction through 'a semester-long writing course' (p. 69), whereas our participants ($N = 284$) had received substantially more uninterrupted classroom instruction over the course of three semesters (PL1: 8 weeks, 160–200 hours; PL2: 16 weeks, 320–400 hours; PL3: 24 weeks, 480–600 hours). Second, the length of time between when our participants wrote their first, second, and third samples is much longer than those of the other study. Therefore, these differences may have been due to the amount of instruction and length of time between when the writing samples were written in these two studies.

6.3 Theoretical implications of the results

Finally, an examination of the theoretical implications of our findings is also warranted. Specifically, what do these results imply concerning certain assumptions of relevant L2A theories, namely crosslinguistic influence, complexity, and linearity? First, as far as the use of different syntactic complexity indices at different proficiency levels are concerned, neither the number of differences nor the sizes of effects between PL1 and PL2 (lower bound) or PL2 and PL3 (upper bound) were as significant as the differences between PL1 and PL3. These results imply that a micro-level linear development, where learners incrementally build more and more sophisticated syntactic elements as they progress from elementary to pre-intermediate and to intermediate levels, is not entirely the case for the duration of a semester or even a year of intensive L2 instruction. Further, it seems that while some uses of more

complex syntactic structures are volatile, others manifest more stable trends at higher proficiency levels (PL3 versus PL2), as learners in this study improved their writing skills after more than a whole year of intensive education that varied between 20 and 25 hours of weekly classroom instruction.

Taken together, these findings may also underscore issues of L1 effect (Ringbom 2007; Ortega 2009) and the role of L2 use and instruction (Robinson and Ellis 2008) in L2A, while also cautioning us about assumptions pertaining to unpredictability and incremental linearity in L2 development (Ellis and Larsen-Freeman 2006; Norris and Ortega 2009; Larsen-Freeman 2011) based on 'years' of study instead of well-defined 'intensity of classroom instruction'. First, in line with assumptions of complex and dynamic systems views (Larsen-Freeman 1997, 2011; Verspoor *et al.* 2011), L2 systems and their subsystems are constituted by adaptive properties (including L1 system) that function in a state of constant flux, resulting in nonlinear and self-organizing changes that lead to subsequent adaptive changes, and so forth (Polat 2016). Therefore, it is not possible to pinpoint exactly why some of these complexity measures developed the way they did at a given level.

Second, without conflating theories about input (Polat 2016) and L2 usage (Robinson and Ellis 2008), we simply assume that the incrementally increasing amount of instruction (PL1 = 160–200 hours; PL2 = 320–400 hours; PL3 = 480–600 hours) over three eight-week semesters allowed for more exposure to input through usage from which the learners came to build more complex syntactic structures at the end of PL3. This aligns with L2A views about exposure to input and usage. Basically, regardless of disagreements about its nature and adequacy, the role of exposure to varying amounts of comprehensible input ('*the sine qua non*' of L2A) in L2 development is well acknowledged in the field (Gass and Mackey 2012; Polat 2016). Although it is hard to determine possible effects of the intensity of instruction on these students' written syntactic complexity development, by offering specifics about the amount of classroom instruction at each proficiency level, this study lays a foundation for comparative work in future research in this area.

Third, the constituent order, agglutination, and morphosyntactic transformation mechanisms of the Turkish language (Oflazer 2014) may have influenced the development of some measures, such as subordination and phrasal sophistication. For example, the relatively free word order nature of Turkish morphosyntax may have implications for possible crosslinguistic effects on our participants' learning of English word order (Papadopoulou *et al.* 2011). Therefore, whether due to structural differences or cognitive functions (Ortega 2009), crosslinguistic factors (Turkish) may have influenced the development of particular complexity measures under study. Lastly, some previous research (Ellis and Yuan 2004; Lu 2011) suggested that compared to writing produced under timed conditions, L2 learners write more syntactically complex structures when they have more time to plan and revise their writing. The corpus used in this study was produced under timed conditions, which

implies that the reported levels of syntactic sophistication at each proficiency level represent the lower bounds of students' levels.

7. CONCLUSIONS

We offer the following conclusions. First, from a theory perspective, our results revealed an overall growth in syntactic sophistication over time; however, a micro-level linear development where learners incrementally build more and more sophisticated L2 syntax as they progress from elementary to pre-intermediate and to intermediate level is not exactly the case. Our findings offer more evidence for the slow-paced, unpredictable, and nonlinear nature of L2 development (Ellis and Larsen-Freeman 2006), underscoring the importance of the role of context particularities in L2A, including the kind of program, intensity of instruction, and learner's L1, among other factors (Ortega 2009). In particular, results of this study are based on a specific number of hours (PL1: 8 weeks, 160–200 hours; PL2: 16 weeks, 320–400 hours; PL3: 24 weeks, 480–600 hours) of classroom instruction. Therefore, the reported patterns of development of particular syntactic complexity measures at each proficiency level constitute developmental benchmarks as to how much growth and in which measures future studies, particularly those in similar contexts, could anticipate.

Second, as for growth patterns, we offer three conclusions that underscore the complex nature of variability in these SC indices over time: (i) 10 indices showed significant growth trends in complexification as proficiency levels increased (PL3 > PL2 > PL1) from PL1 to PL3; (ii) two indices (CT and VPT) demonstrated a PL3 > PL2 ≈ PL1-pattern, with no difference at lower bounds (PL1 and PL2), but yielded significant growth at the upper bounds (PL2 and PL3) and between the lowest and highest levels (PL1 and PL3); and (iii) two indices showed mixed patterns, with CPC values increasing from PL1 to PL2 but decreasing from PL2 to PL3 (PL3CPC < PL1CPC < PL2CPC), and CPT values decreasing as proficiency levels increased (PL3CPT < PL2CPT < PL1CPT). Needless to state, as per our research goals, we offer no inferences about possible changes in 'writing quality' over time based on these results (Bulté and Housen 2014; Crossley and McNamara 2014).

Third, like earlier research, this study is bound by its limitations, which may restrict the generalizability of some of its findings. In light of the research and theory on crosslinguistic influences (Ringbom 2007; Ortega 2009), and other recent research that has reported significant differences in levels and indices of written syntactic sophistication among L2 learners with different L1 backgrounds (Lu and Ai 2015), some of our results may be bound by our participants' L1 background. Thus, future studies can replicate our research design on other L1 backgrounds to test the generalizability of our findings.

Fourth, some of our findings can be interpreted as differences in terms of lower versus upper proficiency levels. Thus, more nuanced results about the use of complexity indices by learners at different levels might be obtained if a

wider range (e.g. five levels) of proficiency were included. Finally, although the LASCA is a commonly used tool in analyses of L2 written input and has been validated (Lu 2010), the computations of 14 different indices from the same unit of input raises concerns about scope and consistency (Polio and Yoon 2018), thereby necessitating caution in interpreting correlations among certain indices.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

REFERENCES

Ai, H. and X. Lu. 2013. 'A corpus-based comparison of syntactic complexity in NNS and NS university students' writing' in A. Diaz-Negrillo, N. Ballier, and P. Thompson (eds): *Automatic Treatment and Analysis of Learner Corpus Data*. John Benjamins, pp. 249–64.

Bardovi-Harlig, K. 1992. 'A second look at T-unit analysis: Reconsidering the sentence,' *TESOL Quarterly* 26: 390–5.

Bulté, B. and A. Housen. 2012. 'Defining and operationalizing L2 complexity' in A. Housen, F. Kuiken, and I. Vedder (eds): *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. John Benjamins, pp. 21–46.

Bulté, B. and A. Housen. 2014. 'Conceptualizing and measuring short-term changes in L2 writing complexity,' *Journal of Second Language Writing* 26: 42–65.

Byrnes, H., H. H. Maxim, and J. M. Norris. 2010. 'Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment,' *The Modern Language Journal* 94: 1–235.

Cooper, T. C. 1976. 'Measuring written syntactic patterns of second language learners of German,' *Journal of Educational Research* 69: 176–83.

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

Ellis, N. C. and D. Larsen-Freeman. 2006. 'Language emergence: Implications for applied linguistics—Introduction to the special issue,' *Applied Linguistics* 27: 558–89.

Ellis, R. and F. Yuan. 2004. 'The effects of planning on fluency, complexity, and accuracy in second language narrative writing,' *Studies in Second Language Acquisition* 26: 59–84.

Gass, S. and A. Mackey (eds). 2012. *Handbook of Second Language Acquisition*. Routledge.

Halliday, M. A. K. and C. M. Matthiessen. 2013. *Halliday's Introduction to Functional Grammar*. Routledge.

Housen, A. and F. Kuiken. 2009. 'Complexity, accuracy, and fluency in second language acquisition,' *Applied Linguistics* 30: 461–73.

Jarvis, S. and T. Odlin. 2000. 'Morphological type, spatial reference, and language transfer,' *Studies in Second Language Acquisition* 22: 535–56.

Johnson, M. D. 2017. 'Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis,' *Journal of Second Language Writing* 37: 13–38.

Knoch, U., A. Rouhshad, and N. Storch. 2014. 'Does the writing of undergraduate ESL students develop after one year of study in an English-medium university?,' *Assessing Writing* 2: 1–17.

Kormos, J. 2011. 'Task complexity and linguistic and discourse features of narrative writing performance,' *Journal of Second Language Writing* 20: 148–61.

Larsen-Freeman, D. 1978. 'An ESL index of development,' *TESOL Quarterly* 12: 439–48.

Larsen-Freeman, D. 1997. 'Chaos/complexity science and second language acquisition,' *Applied Linguistics* 18: 141–65.

Larsen-Freeman, D. 2006. 'The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese

learners of English,' *Applied Linguistics* 27: 590–619.

Larsen-Freeman, D. 2011. 'A complexity theory approach to second language development/acquisition' in D. Atkinson (ed.): *Alternative Approaches to Second Language Acquisition*. Routledge, pp. 48–72.

Lu, X. 2010. 'Automatic analysis of syntactic complexity in second language writing,' *International Journal of Corpus Linguistics* 15: 474–96.

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.

Lu, X. 2017. 'Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment,' *Language Testing* 34: 493–511.

Lu, X. and H. Ai. 2015. 'Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds,' *Journal of Second Language Writing* 29: 16–27.

Mancilla, R., N. Polat, and A. Akcay. 2017. 'An investigation of native and non-native English speakers' levels of written syntactic complexity in asynchronous online discussions,' *Applied Linguistics* 38: 112–34.

McNamara, D. S., A. C. Graesser, P. M. McCarthy, and Z. Cai. 2014. *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.

Norris, J. M. and L. Ortega. 2009. 'Toward on organic approach to investigating CAF in instructed SLA: The case of complexity,' *Applied Linguistics* 30: 555–78.

Oflazer, K. 2014. 'Turkish and its challenges for language processing,' *Language Resources & Evaluation* 48: 639–53.

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

Ortega, L. 2009. *Understanding Second Language Acquisition*. Hodder Arnold.

Ortega, L. 2015. 'Syntactic complexity in L2 writing: Progress and expansion,' *Journal of Second Language Writing* 29: 82–94.

Ortega, L. and H. Byrnes. 2008. *The Longitudinal Study of Advanced L2 Capacities*. Routledge.

Pallotti, G. 2009. 'CAF: Defining, refining and differentiating constructs,' *Applied Linguistics* 30: 590–601.

Papadopoulou, D., S. Varlokosta, V. Spyropoulos, H. Kaili, S. Prokou, and A. Revithiadou. 2011. 'Case morphology and word order in second language Turkish: Evidence from Greek learners,' *Second Language Research* 27: 173–204.

Polat, N., T. Gregersen, and P. MacIntyre (eds). 2019. *Research-Driven Pedagogy: Implications of L2A Theory and Research for the Teaching of Language Skills*. Routledge.

Polat, N. 2016. *L2 Learning, Teaching, and Assessment: A Comprehensible Input Perspective*. Multilingual Matters.

Polio, C. and H. Yoon. 2018. 'The reliability and validity of automated tools for examining variation in syntactic complexity across genres,' *International Journal of Applied Linguistics* 28: 1–24.

Ravid, D. and S. Zilberbuch. 2003. 'The development of complex nominals in expert and non-expert writing: A comparative study,' *Pragmatics and Cognition* 11: 267–96.

Ringbom, H. 1987. *The Role of the First Language in Foreign Language Learning*. Multilingual Matters.

Ringbom, H. 2007. *Cross-Linguistic Similarity in Foreign Language Learning*. Multilingual Matters.

Robinson, P. and N. Ellis (eds). 2008. *Handbook of Cognitive Linguistics and Second Language Acquisition*. Routledge.

Ryshina-Pankova, M. 2015. 'A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor,' *Journal of Second Language Writing* 29: 51–63.

Sasaki, M.2007. 'Effects of study-abroad experiences on EFL writers: A multiple-data analysis,' *The Modern Language Journal*. 91: 602–20. doi: 10.1111/j.1540-4781.2007.00625.x

Serrano, R., E. Tragant, and À. Llanes. 2012. 'A longitudinal analysis of the effects of one year abroad,' *Canadian Modern Language Review* 68: 138–63.

Spoelman, M. and M. Verspoor. 2010. 'Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish,' *Applied Linguistics* 31: 532–53.

Tavakoli, P. and P. Foster. 2011. 'Task design and second language performance: The effect of narrative type on learner output,' *Language Learning* 61: 37–72.

van Lier, L. 2000. 'From input to affordance: Social-interactive learning from an ecological perspective' in J. P. Lantolf (ed.): *Sociocultural Theory and Second Language Learning*. Oxford University Press, pp. 245–60.

Verspoor, M., K. de Bot, and W. Lowie (eds). 2011. *A Dynamic Approach to Second Language Development*. Philadelphia, PA: John Benjamins.

Verspoor, M., W. Lowie, and M. van Dijk. 2008. 'Variability in second language development from a dynamic systems perspective,' *The Modern Language Journal* 92: 214–31.

Vyatkina, N. 2012. 'The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study,' *The Modern Language Journal* 96: 576–98.

Williams, J. 2012. 'The potential role(s) of writing in second language development,' *Journal of Second Language Writing* 21: 321–31.

Wolfe-Quintero, K., S. Inagaki, and H. Kim. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity*. National Foreign Language Resource Center.

Yeniterzi, R. and K. Oflazer. 2010. 'Syntax-to-morphology mapping in factored phrase-based statistical machine translation from English to Turkish'. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 454–64.

Yoon, H. and C. Polio. 2017. 'The linguistic development of students of English as a second language in two written genres,' *TESOL Quarterly* 51: 275–301.