

A Multidimensional Analysis of L2 Written Complexity: The Case of Sojourners

Zeynep Köylü, University of Basel

Abstract

This study investigated the developmental trajectories and the changing nature of the relationship between holistic proficiency, syntactic and lexical complexity in written L2 English. The data analyzed come from the diary subcorpus of the Study Abroad and Language Acquisition Project (Pérez-Vidal, 2014). The Catalan/Spanish bilingual sojourner-participants ($n = 26$) who were advanced L2 learners of English produced a weekly diary entry ($n = 383$) about their sojourn experiences in terms of language use and development, culture, and their daily life in the host country. Drawing on a dynamic systems perspective, the diary entries were analyzed through visual inspection, correlations, and finally lagged cross correlation (LCCA) to confirm any supportive, competitive, and precursor relationships. The results of the analysis confirmed much variation and a dynamic pattern for the indices investigated. There was a competitive relationship between most syntactic and lexical measures, except for lexical sophistication. The LCCAs confirmed these patterns but also indicated a precursor relationship between some indices at certain times during sojourn, such as lower holistic proficiency scores during first weeks predicting higher amounts of subordination in mid-sojourn. The study concludes that the participants paid attention to different dimensions of their written performance at different times during their studies abroad.

Keywords: L2 written English; study abroad; dynamic systems theory; syntactic complexity; lexical complexity

Introduction

Discussions have been ongoing around how to best measure L2 performance and evaluate second language development (SLD) (Ortega, 2012). Productive tasks, such as writing tests given that they might provide a bigger picture of learner language (Verspoor et al., 2012), as opposed to standardized tests measuring passive receptive knowledge, are often preferred. Traditionally, SLD researchers explore measurable subdimensions of L2 performance such as morphological, lexical and syntactic properties; fluency; and accuracy through a written sample (see Housen et al., 2012) to determine proficiency. For a holistic assessment, on the other hand, we expect different characteristics at each proficiency level, mostly following a descriptive assessment scheme, such as the Common European Framework of Reference (Verspoor et al., 2012). Such references have been contentious given the huge individual

variation empirically confirmed by SLD research from a dynamic perspective (Verspoor et al., 2021). Thus, it is unrealistic to expect standardized performances from all as we cannot rule out the influence of task, L1 differences, type of exposure and a whole range of individual differences (Verspoor et al., 2012). Such variation urges alternative assessment criteria that examine multiple subdimensions of L2 performance while accounting for all potential sources of variation. Inspecting certain subdimensions of L2 performance can also inform us about the current state and developmental trajectories of learners' L2 systems at varying proficiency levels (Bulté & Housen, 2020). These have been the motive behind research exploring complexity, accuracy, and fluency dimensions of L2 performance from a dynamic perspective (Verspoor et al., 2021).

Thus, this study aims to dynamically investigate the L2 developmental trajectories of a group of sojourners as advanced writers of L2 English. It draws on the Dynamic Systems Theory (DST) to explore the nature of the relationship between holistic proficiency and syntactic and lexical complexity through the inspection of dense data over the course of a semester abroad. It also aims to broaden our understanding of the interplay between fixed (time) and random effects (participant behavior) in a nonlinear fashion as the sojourner participants developed their written L2.

Literature Review

L2 Complexity

Research on linguistic complexity in syntactic and lexical layers as indices of L2 development, performance and proficiency (Bulté & Housen, 2018). has been well-documented in literature and in most SLD studies, it is measured as a dependent *performative* variable. We know from previous research that learners' overall performance and their syntactic complexity profiles are significantly linked (Bulté & Housen, 2018; Ortega, 2003;

2012). Thus, it makes sense to analyze complexity through both layers (syntactic and lexical) and levels (e.g., clause, phrase, sentence-level) to get a broader picture of L2 performance. L2 complexity may develop as a function of time as learners become more proficient, but it is wrong to claim that holistic proficiency and complexity always have a synchronous and linear developmental trajectory (Yu & Lowie, 2020). We have ample evidence that different layers and subdimensions of complexity may have differentiated developmental routes and interrelationships representing how dynamic the process of SLD is (Verspoor et al., 2021). In terms of a construct definition, the current study operationalizes L2 complexity as “complex, multilayered, multidimensional” (Bulté & Housen, 2020, p. 52) that is compiled of a collection of linguistic forms and structures in interlanguage; and their elaboration, density, and sophistication (Ortega, 2012). It follows the well-established quantitative measurement concepts for L2 syntactic and lexical complexity on clausal, sentential, and phrasal levels for the former; and variation, sophistication, and diversity levels for the latter.

A Dynamic Approach to L2 Development

SLA scholars increasingly draw on the tenets of DST to better understand SLD as they are motivated by the idea that dimensions of L2 performance, such as complexity, accuracy, and fluency (CAF) are multivariate and non-collinear in nature (Norris & Ortega, 2009). Thus, we should adopt an integrative approach considering the whole developmental trajectory to see the changing dynamics of L2 performance (Spoelman & Verspoor, 2010). DST is an efficient theoretical framework with its emphasis on individual variation and nonlinear patterns of development following an iterative progression (Penris & Verspoor, 2017). Accordingly, L2 (or interlanguage) as a system is sensitive to initial conditions (e.g., pre-departure proficiency influences the amount of L2 gains after studying abroad) and constantly changes over time due to the interaction between internal (e.g., language learning

motivation) and external (e.g., the learning context) factors. Such interaction and change may result in significant development of the L2 linguistic system over time, but trade-offs are observed as some subsystems develop at the expense of others (e.g., fluency over accuracy). This study draws on the tenets of DST to understand SLD phenomena in sojourners' written L2 development. Predominantly, it focuses on syntactic and lexical complexity as subsystems of the larger system: the interlanguage. It likewise assumes that these subsystems are interconnected and iteratively influence the development of one another as "each new step emerges from all previous steps" (Verspoor et al., 2012, p. 241). The following question guides the study:

What is the nature of relationship between syntactic and lexical complexity and holistic proficiency (supportive, competitive, precursor) in sojourners' written L2 production over a semester abroad?

Methodology

The study analyzes dense data to track the written development of sojourners during a semester in an Anglophone country. It examines supportive, competitive, precursor relationships among holistic proficiency, syntactic and lexical complexity over 12-17 weeks abroad. It employs a set of multidimensional measures to tap the dynamic nature of L2 written complexity.

Participants

The written samples analyzed come from the Study Abroad and Language Acquisition Project (SALA) (Pérez-Vidal, 2014). A subgroup of SALA participants (n = 26) volunteered to provide weekly diaries during their sojourn. These participants were all Catalan/Spanish

bilinguals at an advanced (B2.2-C1.2) proficiency. They were aged between 17-27 (mean age = 18.2) during the study.

The SALA Diary Corpus

The 26 participants of the study provided weekly diary entries during their sojourn, amounting to 388 entries (12 to 17 weekly entries per participant, a total of 274041 words). These entries were compiled in the SALA Diary Corpus by the researcher.

The participants were given a prompt to guide their diary writing which included their experiences of language learning, cultural and linguistic remarks, details about their social networks and if they changed as a person and developed their language skills during their sojourn. Each participant wrote weekly on these prompts in an electronic diary (MS Word). The entries were not typical *dear diary* entries per se. Instead, they adopted an academic register but kept the typical storytelling component of diary writing through personal anecdotes. The following entry illustrates the type of production in the diaries (all errors are original).

They told us at King's College that we had to choose one or two in-session classes to improve the aspects of the English language in which we are less strong. They offer us the possibility to work on seminar skills, critical analysis skills, pronunciation, grammar and vocabulary. I have been thinking throughout the week about this decision I have to make and I have finally seen that I am weakest at pronunciation and vocabulary. Not that people do not understand me (thanks God!), but I think I can improve those aspects. (P06W01)

Measures of L2 Complexity and Proficiency

First, the data were transformed into plain text and CHAT format (MacWhinney, 2000) for initial analyses using L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010), Lexical Complexity Analyzer (LCA) (Lu, 2012) and CLAN (MacWhinney, 2000).

Syntactic Complexity

The data from the diaries were analyzed for clausal, sentence-level, and phrasal syntactic complexity; and for lexical variation, sophistication, and diversity. The current study included six indices from L2SCA. As this software does not calculate non-finite clauses (e.g., participle reductions) when determining mean length of clauses (MLC), CLAN was used to code and calculate both finite and non-finite MLC. Additional measures were utilized to determine the amount of subordination, coordination, and phrasal sophistication. Table 1 gives a short description of the syntactic complexity indices used.

Table 1

L2SCA Indices (Lu, 2010, p. 479)

Measure abbreviation	Full name	Description
MLT	mean length of T-unit	# of words per # of T-unit
DC/T	dependent clauses per T-unit	# of dependent clauses per T-unit
VP/T	verb phrases per T-unit	# of verb phrases per T-unit

= number

Lexical Complexity

Several tools were utilized to determine lexical complexity: lexical density (LD), sophistication (AWL), and diversity (MTLD). Table 2 illustrates lexical complexity indices used in the analyses with their descriptions (Taken from Köylü et al., 2023)

Table 2

Lexical Complexity Indices

Dimension	Abbreviation	Full name	Description	Software
Density	LD	Lexical density	# of content words per # total words	Vocabprofile
Sophistication	AWL	Academic Word List	# of words from the 570 frequently used words in an academic context (Coxhead, 2000)	Vocabprofile
Diversity	MTLD*	Measure of textual lexical diversity	Average number of tokens in a text required to reach a given type/token value	TAALED

= number

* see McCarthy & Jarvis (2010)

Holistic Scores

Four English as an L2 instructors (native or near-native) rated the diary entries for holistic proficiency on a scale of 0-5. The texts were presented in random order to avoid time-related bias, using a rubric from Verspoor et al. (2017, pp. 4-5) and Köylü et al. (2023).

Data Analysis

A series of time-lagged cross-correlations, controlling for the variable time, was used to determine the strength and nature of the relationship between syntactic and lexical complexity and developmental systematicity following Spoelman and Verspoor (2010). This analysis controlled for a shared linear trend that could inflate correlations as participants' overall proficiency improved (Bulté & Housen, 2020). To trace if a longitudinal developmental pattern exists, all changes in between the two measures analyzed were determined via a series of time-lagged cross-correlations. Lags are certain time lapses, explaining the dependence of a variable on another one over time. In other words, to see the relationship between a sojourner's pre-departure syntactic complexity on in-sojourn or post-sojourn lexical development (around week 6, for instance), a lagged cross-correlation might be used. This lagged solution also explains how the relationship between two variables changes over time. As the data analyzed are weekly diary entries, a lag of seven time units (i.e., seven days of a week) were utilized in the analysis. The two participants with 12-week entries were omitted from the dataset given the 7-lagged solution would not work for their performances. A lagged cross-correlation analysis (LCCA) captures the changing nature of the trend between the measures analyzed given the principle of *dynamic* interconnectedness of subsystems. Additionally, correlograms for statistically correlated variables were plotted to visualize the developmental trajectories of individuals over the course of the sojourn. Before running the LCCAs, the dataset was visually inspected to determine the changing

developmental trajectories over time. To remove noise in the dataset and to detrend the general trend, a moving window of four was selected given the amount of weekly data available (12 to 17 weeks) following Spoelman & Verspoor (2010) and Bulté & Housen (2020). Calculating moving averages is a form of detrending to smooth the data from the noise. Without detrending, there is the risk of overestimating local variability concerning the developmental slope (Spoelman & Verspoor, 2010). In the visual inspections (data plots), both detrended (moving averages) and weekly averages were presented to also see the peaks and valleys of development.

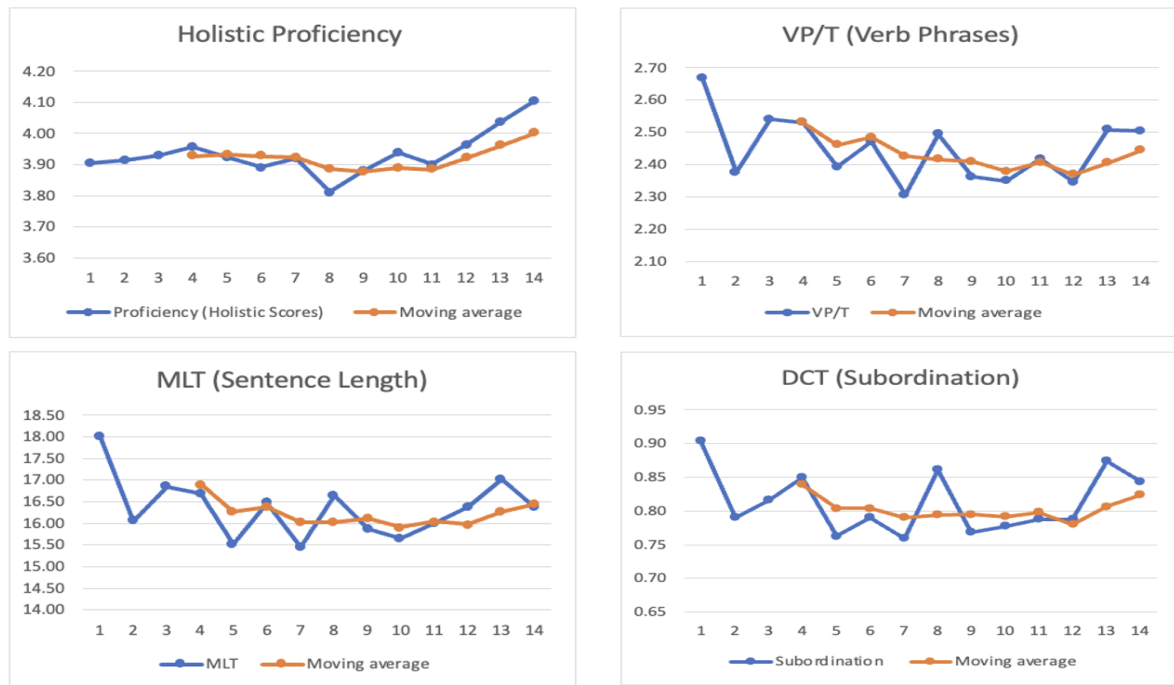
Results

Developmental Trajectories: Holistic Proficiency and Syntactic Complexity

To examine syntactic and lexical complexity over time, weekly data were visually plotted and smoothed with a 4-week moving window to reduce noise. Figure 1 shows the holistic and syntactic development of the group. As shown, the development of holistic proficiency has an increasing trend, meaning that the sojourners had a higher proficiency rate at the end of their sojourn. Conversely, the syntactic complexity indices - MLT (sentence level), DCT (subordination level) and VP/T (verb phrase level) all show a declining trend, meaning that the participants tended to write shorter sentences with less subordination and fewer verb phrases. These plots also confirm high amounts of variation throughout a semester abroad.

Figure 1

Holistic and Syntactic Development during SA

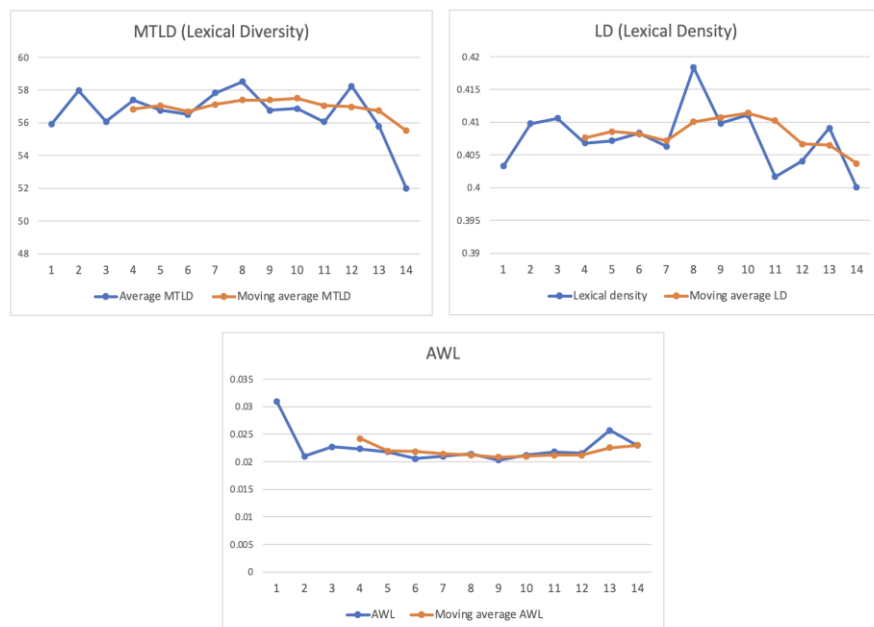


Developmental Trajectories: Lexical Complexity Indices

As seen in Figure 2, the participants used less diverse words from a higher frequency word band and fewer words from the academic word list in their diary entries. The lexical density also decreased. The plots for lexical development demonstrate large amounts of variation for all measures of lexical complexity except for AWL. In order to understand the course of development, charts comparing indices (representing a level - e.g., sentence length or lexical diversity) for each dimension (syntactic or lexical complexity) were cast using the moving averages for each index.

Figure 2

Syntactic and Lexical Complexity Indices



The next step was to visually inspect the developmental trajectories of syntactic and lexical indices through the moving averages. As a measure of syntactic complexity, the development of sentence length (MLT) was first compared to lexical complexity indices (see Figure 3). The charts indicated a competitive trend for MLT and lexical diversity (MTLD) and density (LD) as the syntactic measure tends to go down, but MTLT and LD increased. This trend is different for MLT and lexical sophistication (AWL) as both indices tend to go down over the course of a semester abroad. A follow up inspection of other syntactic indices representing subordination (DC/T) (see Figure 4) and verb phrases per T-Unit (VP/T) (see Figure 5), indicate the trajectories were almost the same when compared to lexical indices, with competition between all syntactic and lexical indices, except for lexical sophistication (AWL).

Figure 3

Development Of Moving Averages for Sentence Length (MLT) and Lexical Complexity Indices

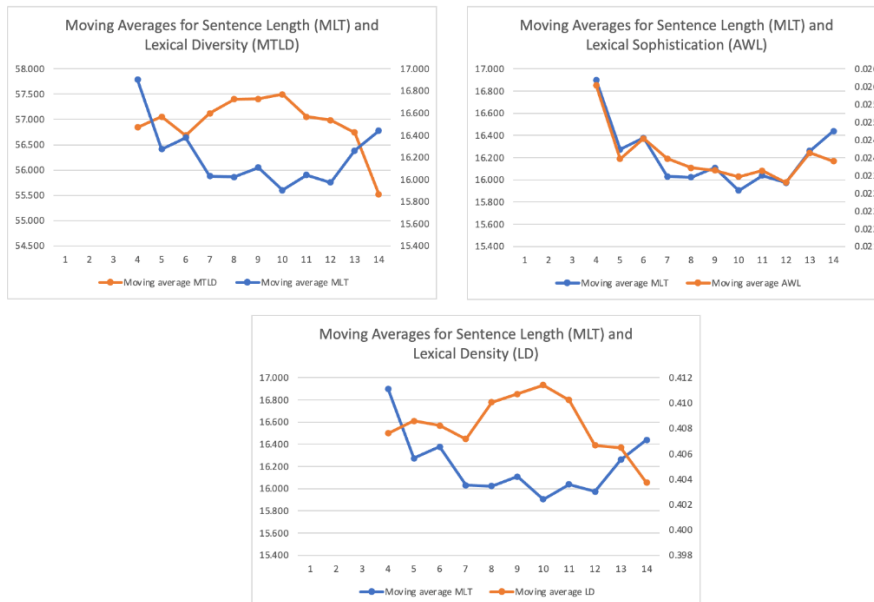


Figure 4

Development of Moving Averages for Subordination (DC/T) and Lexical Complexity Indices

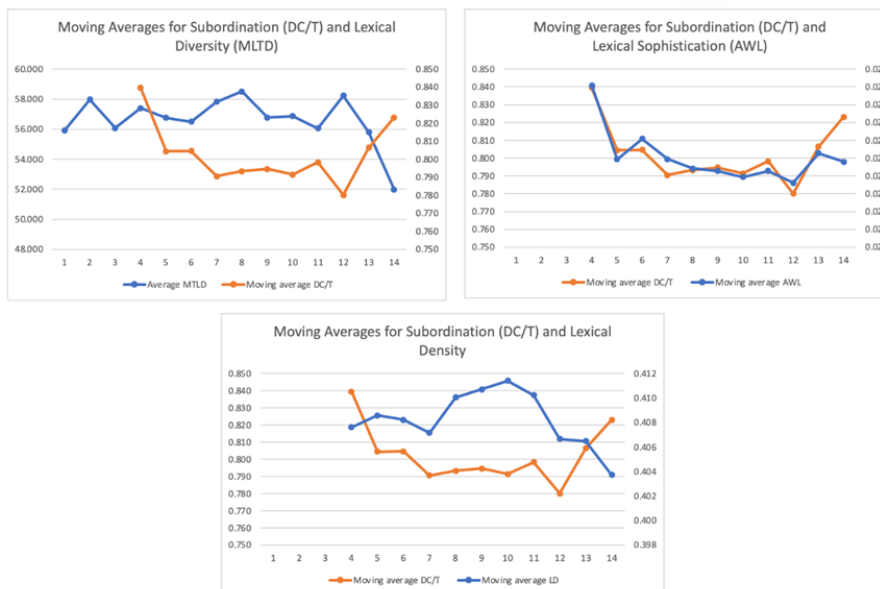
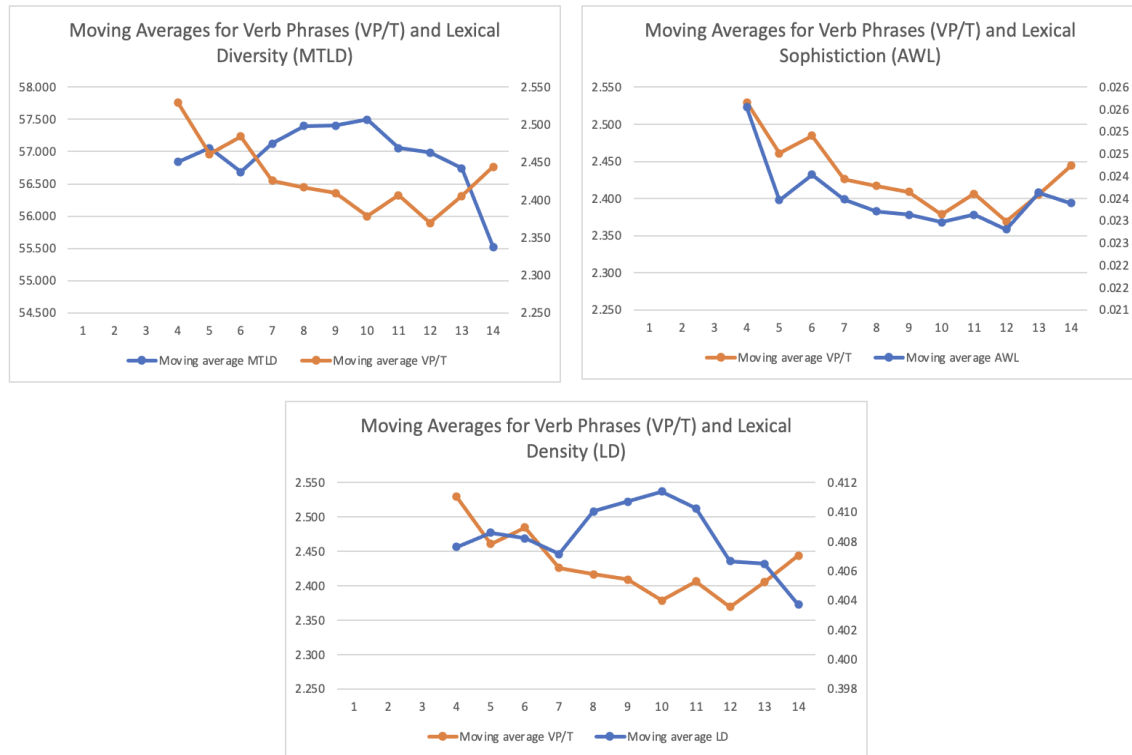


Figure 5

Development of Moving Averages for Verb Phrases (VP/T) and Lexical Complexity Indices



The Relationship between Holistic Proficiency and Complexity

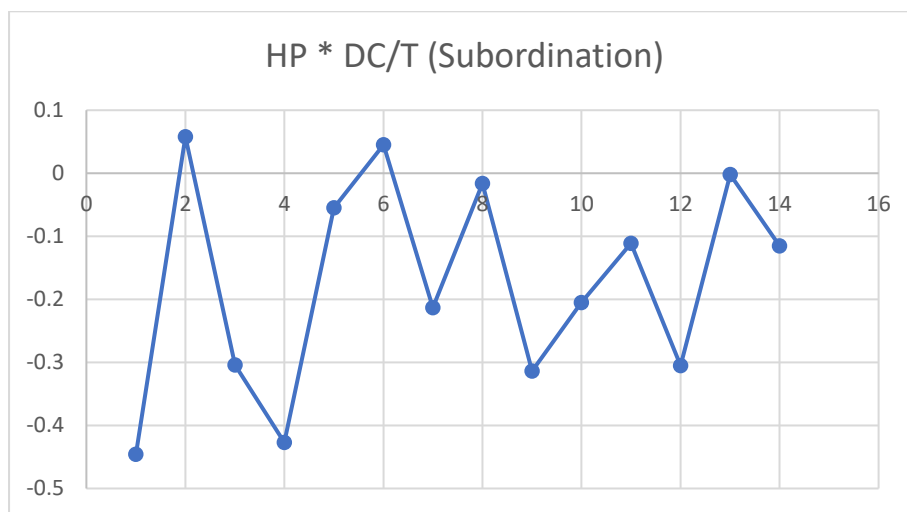
Following the inspection of development, the nature of relationships among the performance measures were inspected. Correlations and LCCAs were run to track how holistic proficiency related to syntactic and lexical complexity over time without further detrending.

Holistic Proficiency and Syntactic Complexity

Regarding holistic proficiency (HP), only the amount of subordination (DC/T) as a measure of syntactic complexity was found to have significant negative correlations with HP over 14 weeks. Figure 6 illustrates the pattern of changing relationships between these two measures.

Figure 6

Correlations between Holistic Proficiency (HP) and Subordination (DC/T)



At the beginning of the sojourn, there was a significant negative correlation ($r = -.446$, $p = .02$) between HP and DC/T, indicating a competitive relationship between the two. This trajectory did not change much throughout the 14-week period abroad. During weeks 2, 6, 8, and 13 the correlation coefficient was close to zero, indicating no relationship. During weeks 1 ($r = -.446$, $p = .02$), 3, 4 ($r = -.427$, $p = .03$), 6, 7, 9, 10, 11, and 12, the trend towards a negative relationship remained constant; however, the correlation coefficient varied, showing the changing magnitude of negative association. The nature of relationship between HP and DC/T thus remains to be a competitive one. The higher the degree of subordination, the lower the holistic scores. No significant relationships were found between HP and other syntactic measures.

Concerning the nature of the relationship between HP and lexical complexity, HP was found to have a significant correlation with lexical diversity (MTLD) and sophistication (AWL), but not with lexical density (LD). There was a significant positive relationship between MTLD and HP (see Figure 7), at the very beginning of the semester abroad ($r =$

.380, $p = .05$), while the rest of the semester these indices had mostly positive but nonsignificant relationships. As for lexical sophistication, there was a negative correlation between HP and AWL (see Figure 8), towards the very end of sojourn ($r = -.446, p = .03$), but overall, the course of relationship tended to support each other, like lexical diversity.

Holistic Proficiency and Lexical Complexity

Figure 7

Correlations between Holistic Proficiency (HP) and Lexical Variation (MTLD)

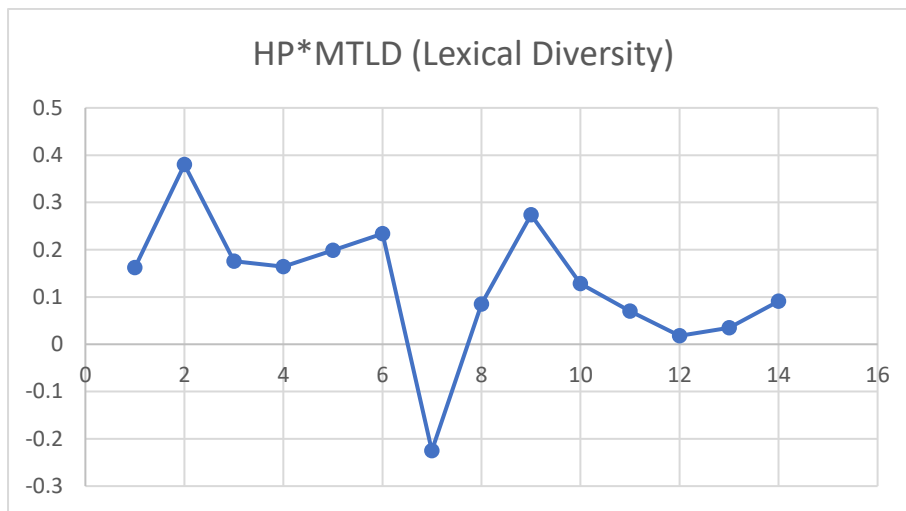
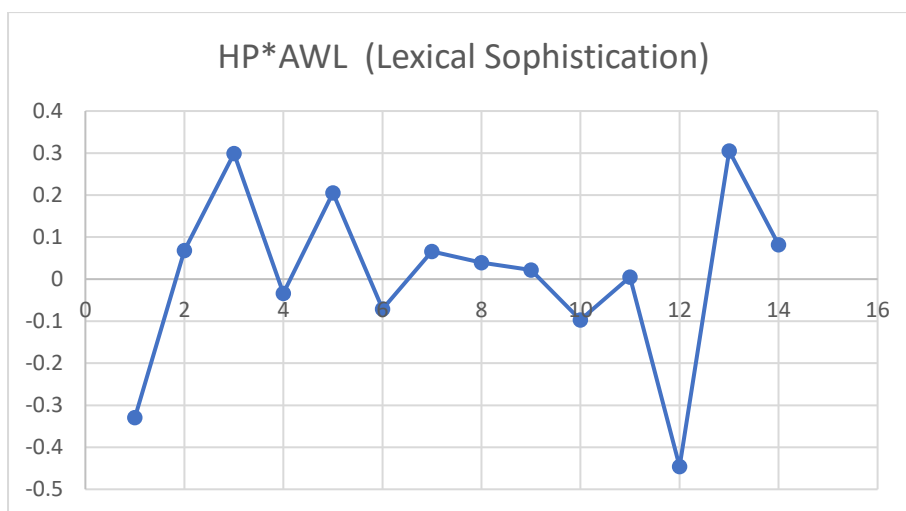


Figure 8

Correlations between Holistic Proficiency (HP) and Lexical Sophistication (AWL)



Syntactic and Lexical Complexity Indices

The inspection between syntactic and lexical complexity also revealed a dynamic picture.

MLT was found to have a positive correlation with AWL (see Figure 9), reaching significance at week 9 ($r = .571, p = .002$), following an overall supportive trend. Figure 10 illustrates the evolution of the relationship between MLT and AWL. The longer the sentences, the more words from the academic word list (AWL) were used indicating lexical sophistication. No other significant relationships were found between AWL and other syntactic indices.

Considering lexical density, DC/T (see Figure 7) was found to have significant negative correlations with LD, reaching significance at weeks 5 ($r = -.435, p = .03$) and 7 ($r = -.455, p = .02$). Although the strength of this inverse relationship changes during the sojourn, the negative trend remains constant, meaning the more the number of dependent clauses is, the lower the lexical diversity in the weekly diary entries. We can conclude that the nature of relationship between these two constructs shows a competitive one. The nature of relationship between verb phrases (VP/T) and lexical density (see Figure 11) showed significant negative correlations at weeks 6 ($r = -.436, p = .03$), 7 ($r = -.431, p = .03$), and 13 ($r = -.598, p = .003$).

Figure 9

Correlations between Sentence Length (MLT) and Lexical Sophistication (AWL)

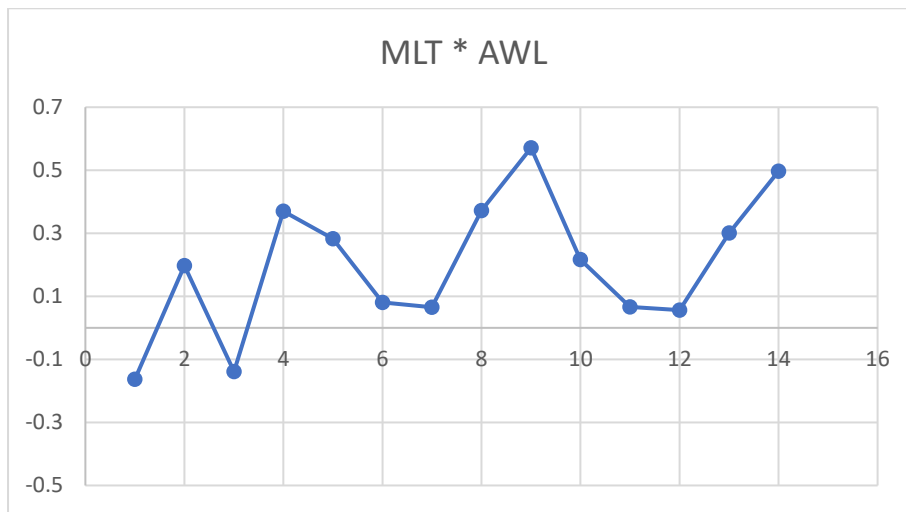


Figure 10

Correlations between Subordination and Lexical Density (LD)

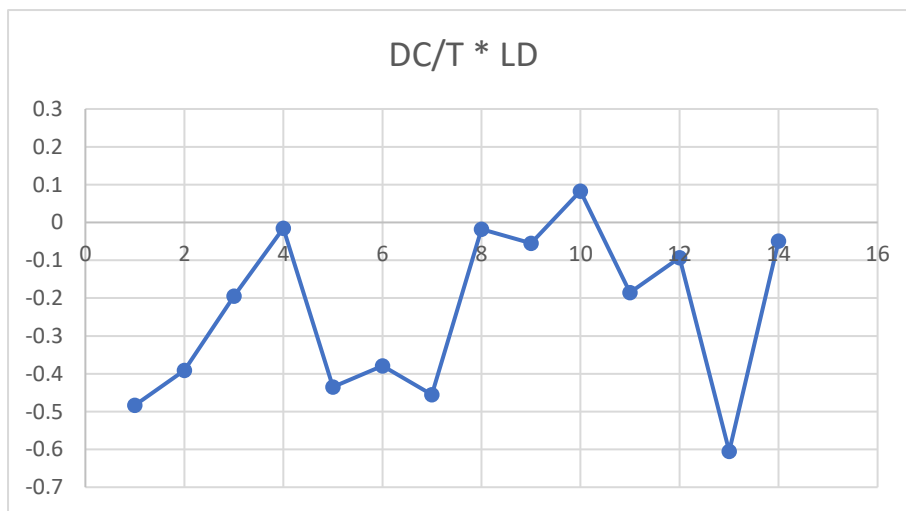
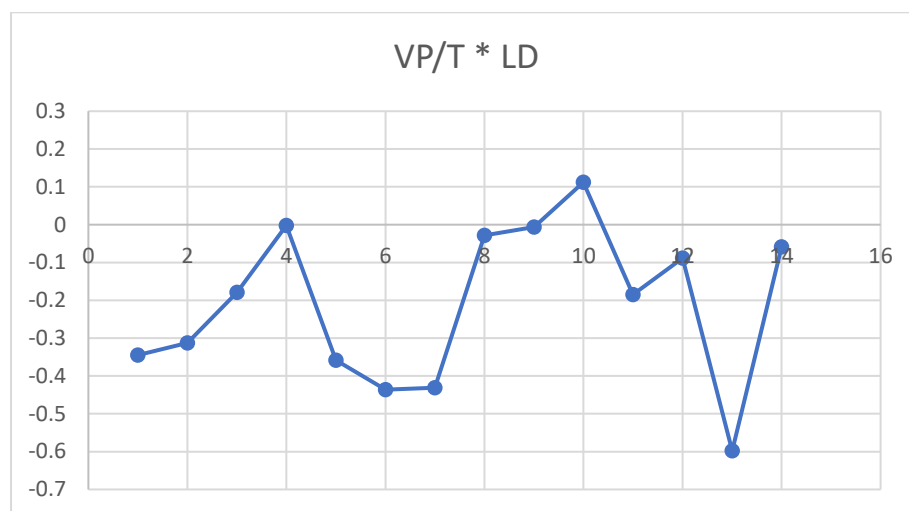


Figure 11

Correlations between Verb Phrases (VP/T) and Lexical Density (LD)



Change of Relationship over Time

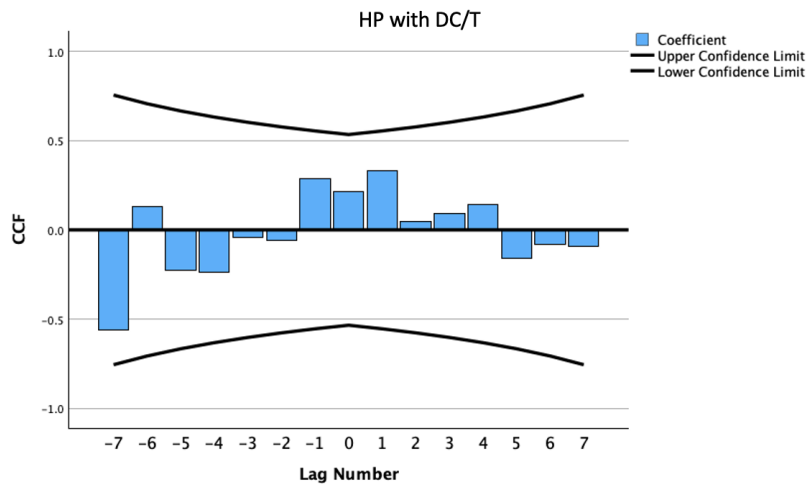
These visual representations of developmental patterns and significant inter-relationships were checked through a series of lagged cross correlational analysis (LCCA). In the LCCA, the weekly aggregated scores over 14 weeks were used following a 7-lag solution.

The first LCCA was run between holistic proficiency (HP) and subordination (DC/T). As seen in Figure 12, the correlogram between HP and DC/T indicates a dynamic pattern throughout the sojourn period. There is a significant inverse relationship between the amount of subordination at week 1 and holistic scores at week 7 (lag -7) ($r = -0.563$, $p = .003$). This means that the higher the holistic scores are at the beginning of sojourn, the lower the amount of subordination will be around mid-sojourn. In an alternative scenario, the lower the holistic score is at the beginning of sojourn, the higher the amount of subordination will be observed during week 7 or so. We might interpret that higher holistic scores would predict fewer examples of dependent clauses in written production as learners continue their sojourn. This negative correlational pattern alters into a positive one towards mid-sojourn (between week 6-11) (lag 1-4), when both measures are positively correlated, yet the relationship is not

significant ($r = .330, p = .099$ for lag 1). Towards the end of sojourn, the inverse relationship emerges again.

Figure 12

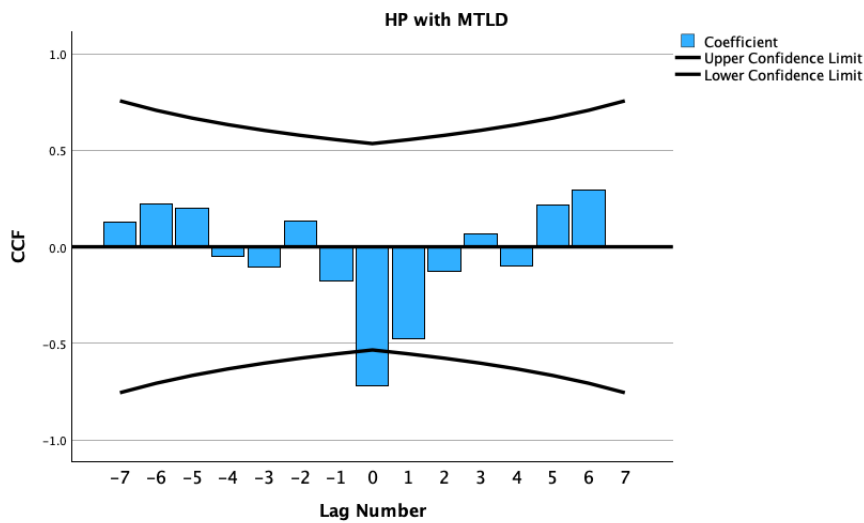
Cross Correlations between Holistic Proficiency and Subordination (DC/T)



Next, a series of LCCAs was computed to assess development of the relationship between HP and lexical complexity. The correlogram shows a positive pattern for HP and lexical diversity (MTLD) in early sojourn towards week 4, but the result is not significant. This pattern evolves into a negative one around week 6 (lag -1), reaching significance at week 7 ($r = -.722, p = .01$). Towards the end of sojourn, the pattern tends to be positive, but not significant. Early high HP scores might predict lower MTLD scores in mid-sojourn. The major trend could be that the higher HP towards mid-sojourn, the lower the MTLD at early sojourn (see Figure 13).

Figure 13

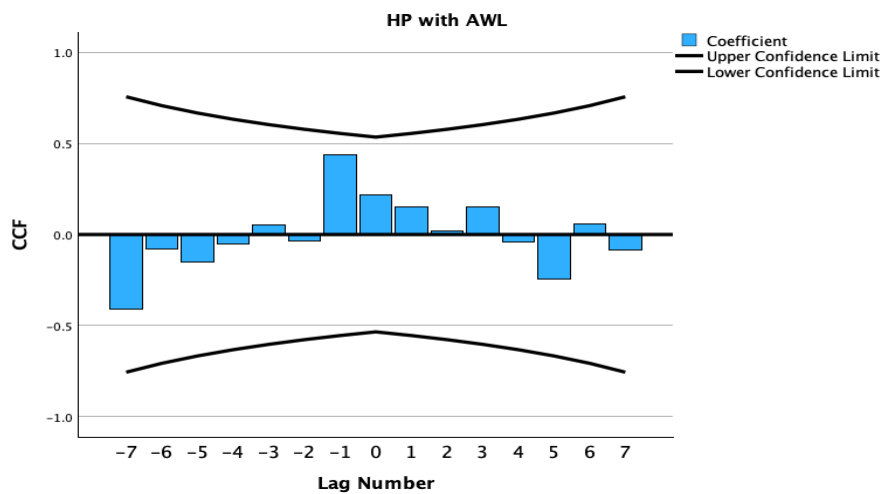
Cross Correlations between Holistic Proficiency and Lexical Variation (MTLD)



The LCCA between HP and AWL (see Figure 14) shows a surprisingly different pattern. Up until week 7, there is an inverse relationship between early lexical sophistication and mid HP scores. The higher the HP scores, the lower the AWL probably due to the competition between the two ($r = -.409, p = .05$ for lag 0). After mid-sojourn (week 6 and onwards), both HP and AWL mostly present a positive relational pattern and develop simultaneously within a supportive relationship ($r = .436, p = .03$ at lag -1). In other words, lexical sophistication only influences sojourners' holistic performance scores starting from approximately week 6.

Figure 14

Cross Correlations between Holistic Proficiency and Lexical Sophistication

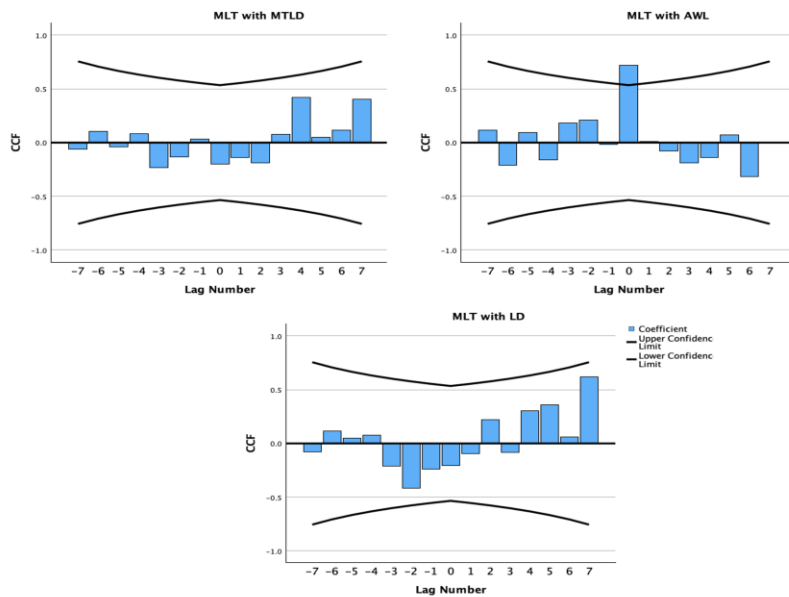


The final series of LCCA was run to investigate the relational trajectory between syntactic and lexical complexity indices. The relational trajectory between MLT and lexical complexity indices mostly showed a negative relationship. Sentence length and lexical diversity (MTLD) were found to have a significant positive relation only at around week 11 (lag 4) ($r = .420, p = .03$), showing a general trend for a dynamic relationship, but not reaching significance. The association between MLT and lexical sophistication was similar, with a significant positive correlation at week 7 (lag 0) only ($r = .721, p = .01$), keeping the dynamic pattern. These trends suggest that long sentences at early sojourn indicate lower scores for lexical diversity and sophistication until later. But around mid-sojourn, there are moments of support for sentence length and MTLD and AWL, which again set onto a competitive relationship only for lexical sophistication (AWL). Considering lexical density, the course of relationship between MLT and LD is similar to that for lexical diversity. At first, sentence length competes with lexical density (lag -2 $r = -.416, p = .03$) until around week 9 (lag 2). From week 9 onwards, longer sentences at mid-sojourn mostly indicate

higher lexical density towards the end, reaching significance at week 14 (lag 7) ($r = .618, p = .01$) (see Figure 15).

Figure 15

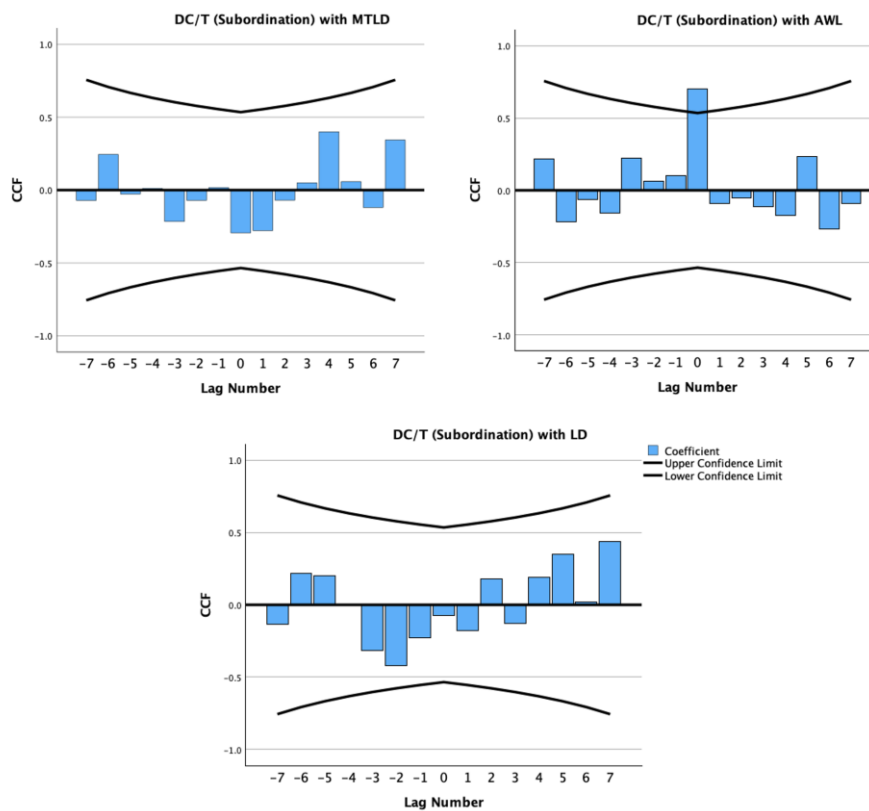
Cross Correlations between Sentence Length (MLT) and Lexical Complexity Indices



Moving to the relationship between subordination (DC/T) and lexical complexity, there are differentiated time-related patterns for each subdimension with an overall competitive relationship (see Figure 16). Lexical diversity and DC/T had a negative relationship until week 11 (lag 4), reaching significance for a positive correlation ($r = .400, p = .03$), indicating that producing subordinate clauses at early-sojourn indicates a higher AWL score towards the end of sojourn. This pattern is similar for lexical density insofar as the competitive relationship in mid-sojourn evolves into a positive one after week 9, reaching significance at week 14 ($r = .440, p = .03$). Higher amounts of subordination lead to increased lexical density later in the sojourn.

Figure 16

Cross Correlations between Subordination (DC/T) and Lexical Complexity Indices



The trajectory between verb phrases and lexical complexity is similar to lexical diversity and density, showing a competitive relationship through negative correlation until after mid-sojourn. This evolves positively after week 9 for both dimensions, reaching significance at week 11 (lag 4) for VP/T and MTLD ($r = .473, p = .02$) and at week 14 (lag 7) ($r = .448, p = .02$). In other words, more instances of subordination led to more diverse vocabulary later in the sojourn when subordination supported lexical diversity. As for VP/T and LD, the trajectory turns supportive at week 12 ($r = .426, p = .03$) and week 14 ($r = .487, p = .02$) despite the negative onset at week 5, indicating an inverse relationship ($r = -.451, p = .02$). Thus, more verb phrases at early sojourn predicted lower lexical density. However, this trajectory changed, indicating that more verb phrases lead to greater lexical density (see Figure 17).

Figure 17

Cross Correlations between Verb Phrases (VP/T) and Lexical Complexity Indices

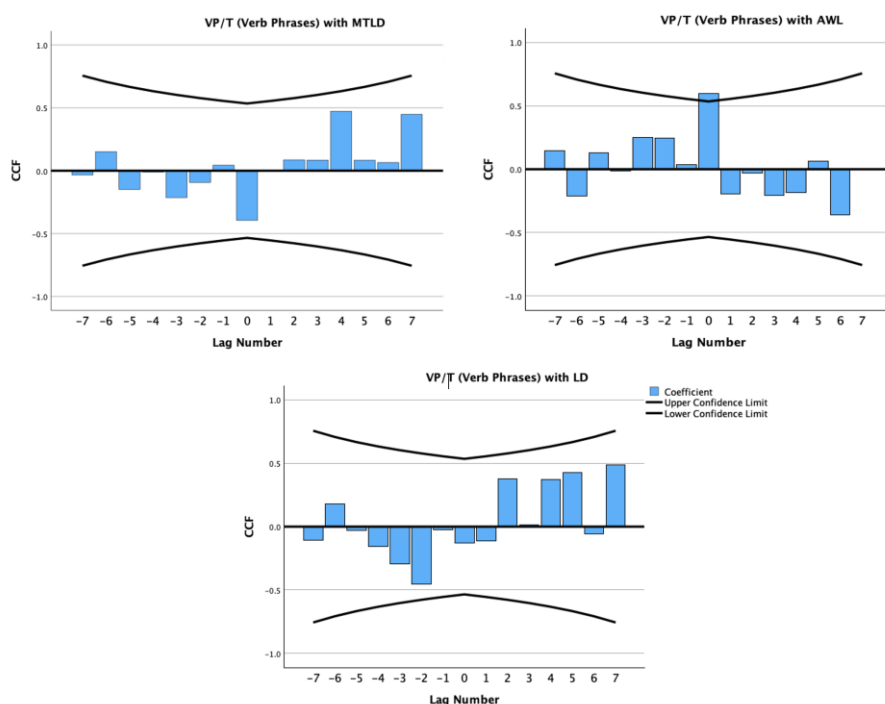


Table 3 provides a summary of all results.

Table 3

Summary of Results

Type of relationship	Visual inspection of moving averages	Correlations
Supportive relationship	Between sentence length (MLT), subordination (DC/T), and verb phrases (VP/T) with lexical sophistication (AWL)	HP and MTL (+) MLT and AWL (+)
Competitive relationship	Between sentence length (MLT), subordination (DC/T), and verb phrases (VP/T) and lexical diversity (MTLD) density (LD)	HP and subordination (-) DCT and LD (-) VPT and LD (-)

<p>Precursor Relationship</p>	<ul style="list-style-type: none"> • Early sojourn lower HP predicts mid-sojourn higher MLT. • Early sojourn lower HP predicts mid-sojourn higher DC/T. • Early sojourn lower HP predicts mid-sojourn higher VP/T. • Early-sojourn higher HP predicts mid-sojourn higher AWL. • Early sojourn higher HP predicts higher LD mid-sojourn. • Early sojourn higher MLT, DC/T and VP/T – predict mid-sojourn higher AWL. 	<p>Lower levels of proficiency tend to focus more on syntactic complexity and this is reflected in their MLT, DC/T, and VP/T scores.</p>
-------------------------------	---	--

Discussion

This study aims to investigate the written developmental trajectories of a group of L2 learners of English over the course of a semester abroad considering both syntactic and lexical complexity, along with holistic proficiency, in a multilayered design. Additionally, it aims to explicate the dynamic relationship among the written performance constructs investigated.

The visual inspection of the time series data indicated that holistic proficiency (HP) increased over the course of sojourn, while all syntactic and lexical indices followed a declining trend. As discussed in a previous paper from this longitudinal project, we know that only the increase for HP and the decrease for VP/T and MTLTD were significant (Köylü et al., 2023). Overall, our participants might have tended to focus more on other aspects of their written production such as formulaicity (Köylü et al., 2024) and verb argument constructions

(Köylü, forthcoming), while they tended to write shorter but more elaborated sentences including fewer verb phrases with words indicating lower levels of lexical diversity.

Concerning the dynamic relationship among these constructs, the results of the correlations per week indicated a few striking patterns for HP and subordination (DC/T). These indices were in a competitive relationship from early on, showing significance at week 1. Higher amounts of subordination in participants' writing received lower holistic proficiency scores from human raters. This might be relevant to the selection of subordinating conjunctions used, such as *because* or *although*, which are typical of academic writing, but when repetitive and lengthy, these might be linked to lower competence levels in learner languages. Use of subordination, or clausal complexity, might be a good indicator of a developing interlanguage at earlier proficiencies, but in more advanced levels, raters might expect shorter sentences with more embedding (e.g., participle reduction) and elaboration, in other words, phrasal complexity (e.g., noun phrase complexity) (Lambert & Nakamura, 2018). Turning to the LCCAs, lower HP scores at early sojourn predicted higher scores for syntactic complexity indices in mid-sojourn. In other words, sojourners with lower levels of initial proficiency might tend to focus more on syntactic properties at least during the first half of their stay abroad.

The dynamic relationship between HP and DC/T might also indicate that raters, when assessing written performance holistically, do not solely consider syntactic properties. The nature of the significant inverse relationship between early sojourn HP and mid-sojourn syntactic complexity might reflect raters' assessment that lengthier sentences with more subordination and verb phrases are not a characteristic of higher or more advanced competence in the L2. This finding is in line with Biber et al. (2011, 2021) as shorter and more elaborated sentences are reported to be a characteristic of advanced academic writing in English L2.

The next significant relationship was found between holistic proficiency (HP) and lexical variation (MTLD). There was a dynamic pattern for these constructs throughout the sojourn with some supportive moments reaching significance at week 2. Yet, towards mid-sojourn, the nature of the relationship was inverse. The higher the HP scores, the lower the MTLD ($r = -.409, p = .05$ for lag 0) probably due to the competition between the two. Sojourners might focus more on one aspect of their production, resulting in no or minimal changes in the other. Conversely, higher HP scores at early sojourn predicted lower levels of lexical diversity and density in mid-sojourn, along with lower scores for all syntactic complexity indices. The correlational pattern for lexical sophistication (AWL) indicated variation and dynamism but mostly a positive relationship with syntactic properties despite a significant inverse correlation at Week 12. Overall, lexical diversity could be a positive indicator of better holistic scores predominantly during early weeks as human raters tend to pay particular attention to participants' lexical properties (Crossley et al., 2011). Also, sojourners' focus could be different during mid-sojourn. Alternatively, raters' criteria could be more linked to the syntactic aspects of L2 writing rather than the lexical aspects until later weeks, which could be reflected in the fact that between early to mid-sojourn LD is negatively correlated to mid to late HP.

The inspection of the relationship between syntactic and lexical indices indicated a positive correlation between sentence length (MLT) and lexical sophistication (AWL). The lengthier the sentences the more words from the academic word list were used. This is an expected result as length, likewise subordination and coordination, would allow for more space to use words from AWL. Considering subordination (DC/T) and verb phrases (VP/T), both indices were found to have a negative relationship with lexical density. This is also natural as when a sentence gets longer, the sentence can include words repeated several times or more content words (e.g., pronouns). These patterns between syntactic and lexical

complexity make sense when length as a syntactic property is taken into consideration. This is also a developmental feature as learners with lower proficiencies tend to write longer stretches with more subordination or verb phrases, which leads to repetitive lexical choices or uses of more function words (Wolfer et al., 2015).

The results of the lagged cross correlations (LCCAs) confirmed the nature of these dynamic relational patterns. Learners might have focused on different aspects of their written production at different stages, but never on both syntactic and lexical complexity simultaneously. Another justification could be raters' assessment criteria. Some indices, such as AWL, might fail to be a major criterion for the raters to build their assessment on. Given the nature of diary writing, even if these advanced learners tended to use an essay-like style, AWL might fail to properly capture lexical sophistication. But HP is more than the sum of its parts, in other words, the dimensions of syntactic or lexical complexity. The human raters take a holistic view and may evaluate the product with reference to other concepts such as formulaicity (Köylü et al., 2024) and this study might have failed to capture every aspect of advanced written L2 production. Thus, it is imperative to incorporate the notion of human judgment through holistic assessment to compensate for the aspects that individual indices miss. The LCCAs also indicated that certain scores could predict performances at different times over the sojourn experience. The predictive power of LCCAs indicated a precursor relationship between these constructs. Writing more elaborate sentences with fewer verb phrases might lead to more lexically dense sentences, potentially paving the way for more target-like use of formulaicity over a semester abroad (Köylü et al., 2024). These results align with those in the literature as different dimensions of L2 performance, even those of the same component, might follow a unique developmental trajectory (Verspoor et al., 2021). Moreover, they corroborate the results of the seminal study by Verspoor et al. (2008) given that the competitive relationship between syntactic and lexical complexity is now confirmed

for advanced level sojourners. However, the direction of the precursor relationship seems to be different than those in Caspi (2010) and Penris & Verspoor (2017) as the current participants tended to focus more on lexical aspects before changing focus to syntactic properties. This might be caused by the nature of the study abroad context, where the participants did not receive any language instruction but were exposed to lots of oral input in meaningful conversations with expert or native speakers of English. Overall, advanced L2 English learners in a study abroad program focus on different aspects of their performances at different stages of the sojourn. Nevertheless, more care is needed when selecting the indices with which to measure change longitudinally (Köylü et al., 2023; 2024; Köylü, forthcoming).

Conclusion

Overall, this paper is the first study investigating developmental trajectories of advanced learners' writing in a semester-long sojourn. The major take away is that the participants focused on different aspects of their written L2 performance at different stages of their sojourn, but not on both syntactic and lexical complexity simultaneously. Furthermore, human judgment provides a more complete picture as raters focus on several aspects of L2 performance at once, which the complexity approach using specific individual indices might fail to capture. Thus, future research should investigate what aspects are critical for human judgment when L2 performance is holistically assessed. By the same token, it is recommended to include an index for formulaicity (e.g., formulaic sequences, multiword constructions) to the analysis especially when exploring the study abroad context with potentially much exposure to oral input but limited or no opportunities for formal language instruction.

References

- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35. <https://doi.org/10.5054/tq.2011.244483>
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2021). *The Register-Functional Approach to Grammatical Complexity*. Routledge <https://doi.org/10.4324/9781003087991>
- Bulté, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics*, 28(1), 147-164. <https://doi.org/10.1111/ijal.12196>
- Bulté, B. & Housen, A. (2020). A DUB-inspired case study of multidimensional L2 complexity development: Competing or connecting growers? In W. Lowie, M. Michel, A. Rousse-Malpat, M. Keijzer, and R. Steinkrauss, *Usage-Based Dynamics in Second Language Development* (pp. 50-87). Bristol: Multilingual Matters. <https://doi.org/10.2307/jj.22730678.8>
- Caspi, T. (2010). A dynamic perspective on second language development (Unpublished doctoral dissertation). Retrieved from: <http://irs.ub.rug.nl/ppn/329338412>.
- Köylü, Z. (Forthcoming). Verb argument constructions in advanced L2 English: The case of sojourners. In K. Madlener Charpentier, Mirjam Weder, and Marjolijn Verspoor (Eds.), *Usage-based perspectives on language and language acquisition*. Routledge.
- Köylü, Z., Eryılmaz, N., & Pérez-Vidal, C. (2023). A dynamic usage-based analysis of L2 proficiency: Syntactic and lexical complexity development of sojourners. *Journal of Second Language Writing*. <https://doi.org/10.1016/j.jslw.2023.101002>
- Köylü, Z., Eryılmaz, N., Pérez-Vidal, C., Verspoor, M., & Gustafsson, H. (2024). Longitudinal development of holistic formulaicity, formulaic sequences, and lexical

- complexity in sojourner diaries: A dynamic usage-based perspective. *Language Learning*, 74(S2), 116-147. <https://doi.org/10.1111/lang.12680>
- Lambert, C., & Nakamura, S. (2019). Proficiency-related variation in syntactic complexity: A study of English L1 and L2 oral descriptive discourse. *International Journal of Applied Linguistics*, 29(2), 248-264. <https://doi.org/10.1111/ijal.12224>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15, 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum Associates.
- 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. <https://doi.org/10.1093/applin/amp044>
- Penris W., Verspoor M. (2017). Academic writing development: A complex, dynamic process, in S. E. Pfenninger and J. Navracsics (Eds): *Future Research Directions for Applied Linguistics* (pp. 215-242). Multilingual Matters. <https://doi.org/10.21832/9781783097135-012>
- Pérez-Vidal, C. (2014). Language acquisition in study abroad and formal instruction contexts. John Benjamins.
- Ortega, L. (2012) Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann and B. Szmrecsanyi (Eds), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact* (pp. 127–155). de Gruyter. <https://doi.org/10.1515/9783110229226.127>

- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31, 532–553. <https://doi.org/10.1093/applin/amq001>
- Verspoor, M., Lowie, W., Chan, H., & Vahtrick, L. (2017). Linguistic complexity in second language development: Variability and variation at advanced stages. *Recherches En Didactique Des Langues Et Des Cultures*, 14(1), 1/27. <https://doi.org/10.4000/rdlc.1450>
- Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231. <https://doi.org/10.1111/j.1540-4781.2008.00715.x>
- Verspoor, M., Lowie, W., & Wieling, M. (2021). L2 Developmental Measures from a Dynamic Perspective. In B. Le Bruyn & M. Paquot (Eds.), *Learner Corpus Research Meets Second Language Acquisition* (pp 172-190). Cambridge University Press. <https://doi.org/10.1017/9781108674577.009>
- Verspoor, M. H., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239-263. <https://doi.org/10.1016/j.jslw.2012.03.007>
- Wolfer, S., Hansen-Morath, S., & Konieczny, L. (2015). Are Shorter Sentences Always Simpler? Discourse Level Processing Consequences of Reformulating Jurisdictional Texts . In K. Maksymski, S. Hansen-Schirra, & S. Gutermuth (Eds.), *Translation and Comprehensibility*. (pp. 263–287). Frank & Timme.