Second Language Writing Development

From A Complex Dynamic Systems Theory Perspective:

A Multiple Case-Study of Hungarian Learners of English

Attila Miklós Wind

This thesis is submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy in Applied Linguistics

August 2018

Department of Linguistics and English Language

Ph.D. Viva Voce

Date:	25/10/2018
Location:	Department of Linguistics and English Language, County South Lancaster University Lancaster United Kingdom LA1 4YL
Supervisor:	Dr Luke Harding (Lancaster University)
Internal examiner:	Dr Judit Kormos (Lancaster University)
External examiner:	Dr Marjolijn Verspoor (University of Groningen)

Abstract

Second language (L2) writing (especially English as a foreign language) is generally considered as one of the most important skills that language learners need to acquire during their language education (Lee, 2016). However, L2 writing might be one of the most difficult skills for English as a foreign language (EFL) learners to develop throughout their language learning process (Barkaoui, 2007; Frydrychova Klimova, 2014). Consequently, writing has often been found to be one of the weakest skills of EFL learners around the world ("IELTS Test taker performance 2016", 2016), including among Hungarian language learners (Csapó & Nikolov, 2001, 2009; Mihaljević Djigunović, Nikolov, & Ottó, 2008). In Hungary, as in other contexts, although writing might be one of the most difficult skills to acquire, language learners have been found to spend an insufficient amount of time learning to write in the foreign language classroom (Árva, 2007; Cook, 2005, 2008; Nikolov, 2002). In addition, research has shown that Hungarian university students make limited use of self-regulatory strategies during writing, despite the important role they play in second language writing development (Kormos, 2012; Lam, 2015; Nitta & Baba, 2015, 2018). Research on L2 writing development in the Hungarian context is therefore important to establish where particular challenges may exist, and how these might be addressed in language education programs.

L2 writing development has been mainly investigated by studies adopting a two-wave longitudinal research design (Barkaoui, 2016; Bulté & Housen, 2014; Knoch, Rouhshad, & Storch, 2014; Knoch, Rouhshad, Oon, & Storch, 2015; Mazgutova & Kormos, 2015; Storch, 2009; Storch & Tapper, 2009). In a two-wave longitudinal research design, changes in complexity and accuracy indices are measured at two points in time from a large population. Although two-wave longitudinal studies are useful in pointing to tendencies and making generalisations in L2 writing development, most recent research suggests that L2 writing

development is idiosyncratic, that is, no two learners exhibit similar developmental patterns (Chan, 2015; Rosmawati, 2016). The 2000s saw the emergence of an influential approach to second language acquisition, the Complex Dynamic Systems Theory (CDST), which acknowledges that there is no average language learner (de Bot, Lowie, & Verspoor, 2007). According to the CDST, language is seen “as a dynamic, complex and nonlinear process” (Larsen-Freeman, 1997, p. 142). CDST studies usually adopt a multi-wave research design to gain insight into the dynamic and complex nature of language development (Caspi, 2010; Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor, de Bot, & Lowie, 2004; Verspoor, Lowie, & van Dijk, 2008; Verspoor, Lowie, Chan, & Vahtrick, 2017; Verspoor, Lowie, & Wieling, 2018).

This study investigates the second language writing development of four Hungarian EFL learners over a nine-month period by adopting a multi-wave mixed-methods research design. The participants were enrolled in an EAP programme offered by a university in Budapest, Hungary. Two argumentative essays were composed each month by the four participants in their naturalistic setting. Furthermore, one argumentative essay was written by the four learners on a monthly basis under a controlled setting. The final mini corpus consisted of 92 argumentative essays collected from the four participants over the nine-month period. The four participants were also interviewed on their self-regulatory processes over the nine-month investigation after each controlled written sample was collected.

This study found that lexical and syntactic complexity indices developed nonlinearly in the four participants’ written data over the nine-month investigation, substantiating previous studies on L2 writing development (Caspi, 2010; Larsen-Freeman, 2006; Rosmawati, 2016; Spoealman & Verspoor, 2010; Verspoor et al., 2004, 2008, 2017, 2018). However, the directions of the trends were different for all lexical and syntactic complexity and accuracy indices in the four EFL learners’ written data. Lexical and syntactic complexity and accuracy indices showed a great deal of variability in the four learners' written data over the nine-month period. The amount of variability constantly changed in the complexity and accuracy indices over time. In addition, the degree of variability was also different in the four participants' written data which supports previous studies on L2 writing development (Caspi, 2010; Larsen-Freeman, 2006; van Dijk, Verspoor & Lowie, 2011). This study found that there was a general improvement in one of the participants' lexicons and in another participant's accuracy over time as demonstrated by the statistically significant developmental peaks (van Geert & van Dijk, 2002). The interactions between lexical and syntactic complexity and accuracy were dynamic over the nine-month investigation. The polarity of the interactions between lexical and syntactic complexity and accuracy changed from negative to positive and vice versa over time. Moreover, the magnitude of the interactions oscillated over time, ranging from weak to strong associations. The polarity and the magnitude of the interactions within lexical and syntactic complexity and between lexical and syntactic complexity and accuracy were different in the four participants' written data which supports previous studies on L2 writing development (Spoelman & Verspoor, 2010; Verspoor & van Dijk, 2011). The self-regulatory processes developed nonlinearly and at different rates in the four participants' learning journey. The four participants' focus shifted from self-observation to self-evaluation processes at different points in time which substantiates earlier studies on self-regulation (Nitta & Baba, 2015; Sasaki, Mizumoto & Murakami, 2018). In addition, only two learners' focus shifted from self-evaluation to goal-setting over the nine-month investigation.

The findings of this study bear important theoretical, methodological, and pedagogical significance and implications. First, the findings of this study support the Complex Dynamic Systems Theory perspective on second language development. Second, this study demonstrated how quantitative and qualitative CDST methods can be combined. Third, this study presented pedagogical implications for teaching and language learning. By demonstrating that language

development is dynamic and idiosyncratic and language development is the outcome of the self-organisation of the language systems, this study questions the validity and the effectiveness of an item-processing approach.

This study offers a new approach to the study of second language writing development and advances our understanding of the complex and dynamic nature of second language development itself.

Acknowledgements

First and foremost, I would like to thank my PhD supervisor, Dr. Luke Harding. His comments on my work have been insightful and his support throughout the research process invaluable. I would also like to thank Jane Sunderland who accepted me as a PhD student and who was my supervisor for the first six months of my PhD study.

I also want to acknowledge Dr. Paul van Geert from the University of Groningen, for his comments and feedback on setting up the mathematical models of this study. His comments and feedback improved the quality of my research.

I am also grateful to two special friends whom I met at Lancaster University for their constant encouragement (Jean Jimenez and Joan Oakley). I am especially grateful to Joan Oakley for acting as a coder of this study.

This research would not have been possible without the help of my four research participants who provided the written and the interview data. Furthermore, I would like to thank the teacher of the EAP course for facilitating the data collection and the coding procedure.

Last but not least, I am deeply thankful to my family for their continuous support. I am taking a chronological approach to mentioning them. First of all, I would like to thank my mother for her continuous support and motivation and my father who unfortunately could not see the completion of this study. I am deeply grateful to my wife (Viki) for her infinite patience and encouragement. Finally, I would like to thank my son (Alex) for being one of the biggest motivational factors.

Declaration

I hereby declare that this thesis has been written by myself and the work is entirely on my own. The research has not been previously submitted for a degree in this or any other form.

14 August 2018

Attila Miklós Wind

Dedication

To my father – to whom I promised to dedicate this thesis before he left this world.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AG	Advanced Guiraud
ANPL	Average noun phrase length
AveWL	Average word length
AWL	Academic Word List
C/T	Clause per T-unit
CAF	Complexity, accuracy, and fluency
CC	Coordinating conjunction
ConC	Conditional clause
CCR	Coordinate clause ratio
CDST	Complex Dynamic Systems Theory
CEFR	Common European Framework of Reference
CI	Coordination Index
CLU	Correct lexical use
CN	Complex nominal
CN/C	Complex nominal per clause
CN/T	Complex nominal per T-unit
CP/C	Coordinate phrase per clause
CP/T	Coordinate phrase per T-unit
C/S	Clause per sentence
CTTR	Corrected type token ratio
CWO	Correct word order
CWP	Controlled writing prompt
DC/C	Dependent clause per clause

DC/T	Dependent clause per T-unit
DRPP	Incidence of prepositional phrase
DST	Dynamic Systems Theory
E/W	Errors per words
EAP	English for Academic Purposes
EFC	Error-free clause
EFC/C	Error-free clause per clause
EFL	English as a Foreign Language
EFT	Error-free T-unit
EFT/T	Error-free T-unit per T-unit
FVR	Finite verb ratio
HMM	Hidden-Markov Model
ID	Individual difference
IELTS	International English Language Testing System
InfC	Infinitive clause
L1	First language
L2	Second language
L3	Third language
LSA	Latent semantic analysis
LW	Learning to write
MLC	Mean length of clause
MLCU	Mean length of c-unit
MLNP	Mean length of noun phrase
MLS	Mean length of sentence
MLTU	Mean length of T-unit

MLU	Mean length of utterance
MSTTR	Mean segmental type token ratio
MTLD	Measure of textual lexical diversity
NPD	Noun phrase density
RelC	Relative clause
SC	Subordinating conjunctions
SFL	Systemic functional linguistics
SL	Sentence length
SLA	Second language acquisition
STRUT	Syntax similarity
SYNLE	Mean number of words before noun phrase
SYNNP	Mean number of modifiers per noun phrase
T/S	T-unit per sentence
TELC	The European Language Certificates
TOEFL	Test of English as a Foreign Language
TTR	Type token ratio
VP/T	Verb phrase per T-unit
W/T	Words per T-unit
WL	Writing to learn
WLC	Writing to learn content
WLL	Writing to learn language
WP	Writing prompt
WRDFRQc	Word frequency

"I think the next century will be the century of complexity".

(Hawking, 2000)

Chapter 1. Introduction

Chapter Overview

Chapter 1 presents a broad introduction to this study. There are five main sections: (1) second language writing, (2) second language writing in the Hungarian language classroom, (3) second language writing research, (4) the aim of the study and (5) the overall structure of the thesis.

1.1 L2 Writing

Second language (L2) writing (especially English as a foreign language) is usually seen as one of the most important skills that language learners need to acquire during their language education (Lee, 2016). A high degree of writing proficiency can be the key to passing a foreign language exam, receiving a scholarship to study abroad, or being accepted for a master's degree programme for undergraduate university students (Jordan, 2002). In addition, a good command of L2 writing can be crucial for successful employment in our globalised world (Parks, 2016). Conversely, poor writing skills may be detrimental in an application process for a Master's degree programme or may impede employment prospects (Parks, 2016) and harm business (Long, 1998).

Although L2 writing is important, it might be one of the most difficult skills that English as a foreign language (EFL) learners need to acquire during their language development (Barkaoui, 2007; Frydrychova Klimova, 2014). According to the International English Language Testing System (IELTS) Test taker performance 2016, of the four skills tested, the lowest mean band scores were recorded for the writing tasks (both academic and general). These results were consistent for males and females across the sample of participants from forty countries ("IELTS Test taker performance 2016", 2016). Writing is usually seen as a psychological problem-solving activity akin to solving mathematical problems or interpreting challenging texts since EFL writers need to employ a number of different writing strategies such as planning and editing while composing a text (Cumming, 2016). L2 writing is a complex process that involves “the full range of psychological, cultural, linguistic, political, and educational variables in which humans engage” (Cumming, 2010; p. 19). Furthermore, writing is a time-consuming activity that needs determination and concentration (Kormos, 2012). Therefore, the EFL learners’ motivational profile and self-regulation might play an important role in L2 writing development.

1.2 L2 Writing in the Hungarian Language Classroom

Writing in English might also be a highly complex task for Hungarian language learners. Csapó and Nikolov (2001) found that writing was the weakest skill among 12-year-old, 14-year-old, and 16-year-old Hungarian EFL students. The representative sample of students (almost 28,598 students) were asked to compose an informal letter which was evaluated according to a marking scheme. Furthermore, Mihaljević Djigunović, Nikolov and Ottó (2008) found that writing was the weakest skill among Hungarian 8th graders’ performances in a comparative study of Croatian and Hungarian EFL students. Csapó and Nikolov (2009) concluded that “L2 literacy skills are cognitively more demanding than aural” among Hungarian EFL learners. According to the IELTS Test taker performance 2016, test takers from Hungary received their lowest mean band score in writing (academic: listening: 7.42, reading: 7.19, writing: 6.31, and speaking: 6.97, overall: 7.03; general: listening: 6.59, reading: 6.27, writing: 6.00, and speaking: 6.60, overall: 6.43) (“IELTS Test taker performance”, 2016). There could be four main reasons why Hungarian EFL learners struggle with writing:

1. the lack of exposure to comparable and parallel texts,
2. the lack of explicit instruction on academic writing in L1,

3. (2) the insufficient amount of time spent learning to write,
4. (3) the inefficient use of self-regulatory processes.

1.2.1 The Lack of Exposure to Comparable and Parallel Texts

The use of first language (L1) in the second language classroom might be associated with second language writing development. One form of L1 use in the second language classroom is the composition of comparable and parallel texts. However, there has been no agreement on why and how the use of L1 might assist L2 development. The field of second language acquisition (SLA) has been divided into two opposing camps. The proponents of the monolingual teaching approach oppose the use of L1, while the supporters of the bilingual teaching approach promote the use of L1 in the second language classroom (Hall & Cook, 2012). According to the monolingual teaching approach, students learn an L2 in the same way as they acquired their L1 since there are natural and universal orders for SLA (Krashen, 1981). In addition, interference from L1 might be counter-productive (Cook, 2001) and the use of L2 should be maximised in classrooms since students have little exposure to it after class (Tang, 2002). Cook (2009) noted that the monolingual teaching approach has been dominant since the rise of the Communicative Language Teaching (CLT) approach. In contrast, the proponents of the bilingual teaching approach do not exclude the use of L1 in the second language classroom. For instance, Stern (1992) and Macaro (2006) claimed that inter-lingual and intra-lingual skills and practices might complement one another if they are combined adequately.

The incorporation of parallel and comparable corpora (including translation) in the second language classroom is one way how the students can benefit from the L1 use during second language learning. Comparable corpora are a collection of independently produced authentic texts in more than one language. These texts are comparable to each other in terms of subject matter, text type, style, communicative function, and register. On the other hand, parallel corpora is comprised of original source texts and their translations. Li (2018) pointed out that both comparable and parallel texts can be important sources in language teaching and translation and contrastive studies because they promote students' language-specific, typological, and cultural knowledge and explain inter-lingual and intercultural differences between source and target texts and between native and non-native texts at both macro and micro levels.

1.2.2 The Lack of Explicit Instruction on Academic Writing in L1

The second reason why Hungarian learners struggle with L2 writing might be that there is little explicit instruction on academic writing besides the dissertation writing seminars. Furthermore, these seminars also tend to focus more generally on academic and research skills and not on writing. According to the regulation of the Ministry of Education in Hungary (15/2006, IV. 3.), both undergraduate and postgraduate university students are required to write a dissertation in order to obtain their university degrees. Consequently, both undergraduate and postgraduate students are obliged to participate in dissertation writing courses (szakdolgozat szeminárium in Hungarian) provided by their university. Although there have not been any studies on the effectiveness of dissertation courses at Hungarian universities, it can be speculated that Hungarian university students struggle with academic writing. First, there has been a decrease in the quality of the dissertations written in Hungarian universities (Molnár, 2007). Second, there has been an increase in the number university students who order their dissertations from illegal dissertation writers (Kovacsik, 2016). However, academic writing needs to be learned and developed through explicit instruction for all novice writers regardless of their native or non-native speaker status (Zhao, 2017).

1.2.3 The Insufficient Amount of Time Spent Learning to Write

Although EFL writing is seen as an important element of success at university and work, language learners still spend only a limited amount of time learning to write in the foreign language classroom during their elementary, secondary and university years (Cook, 2005). Nikolov (2002) found that 10^{th}-, 11^{th}-, and 12^{th}-grade EFL students spent the least amount of time writing compared to other skills such as listening, reading, speaking, grammar, vocabulary in 118 observed classes in secondary schools in Hungary. Furthermore, EFL students spent the most time copying text, gap filling, translating text from English into Hungarian. Nikolov (2002) concluded that the EFL students completed predominantly language-focused writing tasks in her study which reflect the grammar-translation and audio-lingual tradition in Hungarian language education practices. Árva (2007) also found that Hungarian EFL teachers considered speaking skills as the priority and thus provided little time to the development of L2 writing skills. Furthermore, Árva (2007) found that Hungarian EFL teachers were not properly trained how to teach writing. Cook (2008) also pointed out that most teaching methods in the twentieth century did not see speech and writing as being equally important. In addition, most course books devote only a few pages to writing tasks. For example, only a tenth of the tasks are related to writing activities in the Oxford's *New English File Intermediate* (Oxenden, Latham-Koenig & Storton, 2006), a frequently used textbook in Hungarian state and public schools. Csizér and Tankó (2017) also noted that Hungarian EFL learners do not complete a sufficient number of writing tasks during their secondary school education, although they are required to compose two letter subgenres (letter or e-mail) as part of their final school leaving examination.

There is no obligatory language education at universities in Hungary. However, according to the Act CCIV of 2011 on Higher Education, undergraduate students have to pass a language exam in order to obtain their bachelor's degree. In addition, there are certain university programmes, such as international economy and business, where students have to pass language exams both in their first and second foreign languages. For those students who wish to participate in study abroad programmes such as the Erasmus, both undergraduates and postgraduates have to take either the International English Language Testing System (IELTS) or the Test of English as a Foreign Language (TOEFL) language exams, as stipulated by the accepting institution. Some universities in Budapest provide both English for Academic Purposes (EAP) and language exam preparatory courses for universities students.

1.2.4 The Inefficient Use of Self-Regulatory Processes

EFL writers must coordinate cognitive, metacognitive, and linguistic processes when composing extended texts. In other words, writers must self-regulate at several levels (Boscolo & Hidi, 2007). Therefore, self-regulation plays an important role in second language writing development (Kormos, 2012; Nitta & Baba, 2015, 2018). However, there is growing evidence that EFL learners do not use self-regulatory strategies efficiently (Kormos & Csizér, 2014; Csizér & Tankó, 2017; Göy, 2017; Seker, 2016). For example, Kormos and Csizér (2014) investigated the relationship between the motivation, learner autonomy and self-regulation strategies of 638 Hungarian EFL secondary school learners and found that there was a mismatch between a higher level of motivation to learn English and lower levels of autonomy and self-regulatory strategy use. Kormos and Csizér (2014) also found that learning goals in connection with the international status of English, instrumental orientation, and positive self-related beliefs are prerequisites for the efficient use of self-regulatory processes. Furthermore, Csizér and Tankó (2017) found that 222 first-year English majors from a Hungarian university were motivated to increase their abilities in professional writing, although only a third of them possessed the ability and the willingness to control their writing processes.

1.3 L2 Writing Research

L2 writing development has been investigated from a number of different theoretical perspectives such as sociocultural theories of language learning (Wigglesworth & Storch, 2012), theories of multicompetence in language learning studies (Kobayashi & Rinnert, 2012), goal theories in education and psychology (Cumming, 2006; 2012), genre theories in L2 writing research (Tardy, 2012) and systemic functional linguistics (Byrnes, 2012). Nevertheless, Cumming (2010) pointed out that "no single theory might ever explain such complex phenomena as second language writing" (p. 19). In addition, Manchón (2012) argued that "the development of L2 writing capacities is intrinsically a multifaceted phenomenon that is mediated by a wide range of varied personal and situational variables" (p. 5). Cumming (2016) concluded that "the challenge for theories of L2 writing, however, is to account for the multi-faceted complexity of L2 writing comprehensively" (p. 79).

Research on L2 writing development has a relatively long history in the field of second language acquisition (SLA) dating back to the 1930s (Anderson, 1937; Frogner, 1933; LaBrant, 1933). Most previous studies focused on the linguistic aspect of development by adopting two-wave longitudinal research designs, also known as panel studies. In a two-wave longitudinal research design, changes in complexity and accuracy indices are measured at two points in time (Barkaoui, 2016; Bulté & Housen, 2014; Knoch et al., 2014, 2015; Mazgutova & Kormos, 2015; Storch, 2009; Storch & Tapper, 2009). However, a two-wave research design cannot explore the nonlinear and complex nature of L2 writing development.

The 2000s saw the emergence of a highly influential perspective on second language development, the Dynamic Systems Theory (DST) or Complex Dynamic Systems Theory (CDST)[1], in the field of SLA (de Bot, Lowie, & Verspoor, 2005a). According to the CDST, language is seen "as a dynamic, complex and nonlinear process" (Larsen-Freeman, 1997, p.

[1] de Bot (2017) recommended the label Complex Dynamic Systems Theory (CDST) to refer to both Complexity Theory and Dynamic Systems Theory.

142). By adopting a multi-wave research design, the CDST approach provided new insights into the complex nature of L2 writing development. Within the CDST perspective, changes in complexity and accuracy indices are measured multiple times in order to capture the dynamic, complex nonlinear process of L2 writing development over time (Caspi, 2010; Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor et al., 2004, 2008). Furthermore, Manchón (2016) argued that CDST is “a suitable theoretical framework to inform future research on diverse facets of development of L2 writers and their texts” (p. 9-10).

Although the body of research on L2 writing development from a CDST perspective is continually growing, there is an urgent need to validate and develop new research methodologies and techniques and to conduct research in as many different contexts as possible. Second language writing development from a CDST perspective has mainly been investigated in the EFL context and the populations included mainly Dutch (Ma, 2012; Penris & Verspoor, 2017; Verspoor et al., 2004; 2008), and Asian learners of English (Caspi 2010; Hou, 2017; Larsen-Freeman, 2006; Rosmawati, 2016). However, there is a dearth of research on the L2 writing development of EFL writers with a Finno-Ugric language background such as Hungarian.

In addition, future studies from a CDST perspective need to expand the focus of research to other closely related fields such as self-regulation and motivation (Rosmawati, 2016). Although a few studies investigated the development of linguistic complexity and motivational variables simultaneously (e.g. Kormos & Dörnyei, 2004; Mazgutova, 2015), these studies adopted a two-wave research designs failing to gain insight into the complexity of L2 writing development.

1.4 The Aim of the Study

In this section I indicate the potential significance of the study. This study provides theoretical and methodological contributions to the research on L2 writing development and has important pedagogical implications.

The first contribution of my study is theoretical. Although there is a growing body of research on L2 writing development from a CDST approach (Caspi, 2010; Penris & Verspoor, 2017; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2004; 2008), there is still a dearth of studies in this vein. In addition, most previous studies focused only on the linguistic aspect of L2 writing development, thus not exploiting the holistic and overarching nature of the CDST. Therefore, this present study expanded the scope of the existing body of research on L2 writing development by including the writers' motivational and contextual factors.

The second contribution of my study is methodological. Although previous studies indicated that a mixed-methods approach would more fruitfully capture the complex nature of L2 writing development, they employed mainly quantitative methods of data collection (Caspi, 2010; Rosmawati, 2016; Verspoor et al., 2004; 2008). However, Nitta and Baba's (2015) study is an exception since they applied a mixed-method approach to the evolution of self-regulation and L2 writing. However, Nitta and Baba (2015) focused only on the fluency aspect of linguistic development. The present study adopted a mixed-methods research design to investigate four Hungarian EFL writers' linguistic development and the evolution of self-regulation.

Finally, my study made a pedagogical contribution by drawing attention to the idiosyncratic nature of development. Although linguistic development in L2 writing is considered to be a general phenomenon (see e.g. Biber, Gray, & Poonpon, 2011; Halliday & Matthiessen, 1999), this study emphasises that no two learners go through similar developmental paths.

1.5 The Overall Structure of the Thesis

In Chapter 1, I introduced the background to my study. First, I described the context of my study and then I provided a broad view of research on L2 writing development. I pointed out that longitudinal studies are still rare, although they can be very revealing of developmental patterns. I then provided an overview of my study, indicating the significance of the research.

In Chapter 2, I present the theoretical background to the study. I explain the Complex Dynamic Systems Theory, the theoretical framework of this study, and justify its relevance to my research.

In Chapter 3, I focus on the current issues on second language writing research, focusing on lexical and syntactic complexity and accuracy. I provide an overview of research on L2 writing development. The chapter finishes with a review of the literature on L2 writing motivation, focusing on self-regulation.

In Chapter 4, I describe the methodology of this study. I begin with the conceptual foundation of the study by presenting the main characteristics of CDST research methodology. The central body of the chapter includes the description of the pilot study, the data collection, the data coding, and the data exploration of this study.

In Chapter 5 to 8, I present the results of this study, case by case. All four chapters are arranged in analogous format. The results are presented in the order of the research questions they answer and each chapter is divided into six main sections: (1) the participant and the context, (2) the developmental profile, (3) the degree of variability, (4) interactions, (5) modelling and (6) motivation. The first section briefly describes the participants. The second and the third sections present the developmental trends and the degree of variability of complexity and accuracy respectively. The fourth section presents the interactions between complexity and accuracy, while in the fifth section the interactions are simulated by modelling. The last section presents the evolution of the participants' self-regulatory processes.

In Chapter 10, I discuss the results of the case studies collectively. I point out the major findings of this study and the implications they might have for advancing our understanding of L2 writing development and L2 writing motivation.

In Chapter 11, I present the theoretical, methodological, pedagogical significance of this study. I provide implications for future research and present the limitations of this study. Chapter 11 finishes with the conclusions of this study.

Chapter 2. Complex Dynamic Systems Theory

Chapter Overview

Chapter 2 presents the theoretical framework of this study: the Complex Dynamic Systems Theory. There are three main sections: (1) theories of second language acquisition, (2) Complex Dynamic Systems Theory and (3) features of dynamic systems.

2.1 Theories of L2 Acquisition

Various theories and hypotheses have been proposed concerning how humans acquire a second language. Modern SLA research started with two seminal publications: (1) Corder's (1967) The Significance of Learners' Errors, and (2) Selinker's (1972) Interlanguage. The 1970s saw the emergence of error analysis, studies on transitional stages of second language ability, and on natural order of morpheme acquisition (Brown, 1973). The 1980s were dominated by Krashen's (1977) Input hypothesis, White's descriptions of learner competence, and Pienemann's speech processing models. In the 1990s several new hypotheses were suggested such as Schmidt's (1990) Noticing hypothesis, Swain's (1985) Output hypothesis, and Long's (1996) Interaction hypothesis. Nevertheless, two major approaches were predominant in the field of SLA: (1) the application of linguistic theory and (2) the application of certain psychological approaches. The linguistic theoretical approach focuses on the adequate description and explanation of interlanguage. Proponents of the linguistic theoretical approach focused on the nature of the learner's internal mental representation and what restricted it. The main principle of the linguistic theoretical approach is that language is special and uniquely human, and language is encapsulated in its own module in the mind or brain. Humans are equipped from birth with a set of language-specific constraints called Universal Grammar (Chomsky, 1986). Consequently, language acquisition is a special type of experience for humans that involves the interactions of Universal Grammar with data form the outside world. In contrast, the advocates of the application of certain psychological approaches avoided any linguistic description of an interlanguage. Instead, proponents of the psychological approaches viewed language “as just another instance of human behaviour” (VanPatten & Benati, 2010, p. 5). Language was not seen as special anymore but rather “it was an artefact of learning, a latent structure that emerged based on data the learner encountered in the environment” (VanPatten & Benati, 2010, p. 5). In the 1990s, several other approaches emerged such as Processability Theory and Input processing, which were compatible with either linguistic theory or cognitive theory. Furthermore, the 1990s saw the emergence of Sociocultural Theory that rejected both linguistic theory and cognitive theory due to their mind/brain orientation. Instead, Sociocultural Theory situated the learner as an active agent in learning within special social contexts.

In the 2000s SLA research was split into two main camps of linguistic theory and psychological approaches. Van Patten and Benati (2010) “do not envision this changing in the near future” (p. 5). Nevertheless, the early 2010s saw the emergence of usage-based or emergentist theories. According to usage-based theories, language is viewed as a system that emerges through perception and language use. In contrast, Universal Grammar views language as a genetically pre-determined and inborn system. According to Chomsky’s (1980) argument, referred to as ‘the poverty of stimulus’, children are not exposed to rich enough data within their linguistic environment to acquire every characteristic of language. Lowie (2017) pointed out that the poverty of stimulus demonstrates that languages cannot be learnt from input and interaction. In contrast, usage-based or Emergentist approaches have shown that language can be acquired from input and interaction (Tomasello, 2003). According to N. C. Ellis (1998), language emerges “from interactions at all levels from brain to society” and “simple mechanisms suffice to drive the emergence of complex language representations” (p. 631). A novel contribution to the usage-based and Emergentist framework is the Complex Dynamic Systems Theory.

2.2 Complex Dynamic Systems Theory

Complex Dynamic Systems Theory (CDST) originated as an area of mathematics. It deals with the long-term behaviour of dynamical systems and studies the nature of the equations of motion of systems that are usually mechanical or physical in nature such as planetary orbits. Although Complex Dynamic Systems Theory was initially applied to describe the behaviour of the complex dynamical systems in applied mathematics, it has applications to a wide variety of branches of natural science such as biology, chemistry, physics, and astronomy. Complex Dynamic Systems Theory has also been applied in the field of social sciences such as cognitive development (Thelen & Ulrich, 1991; Smith & Thelen, 1993) and applied linguistics (Larsen-Freeman & Cameron, 2008). For example, the CDST was applied to study infants' development of perception and action (Thelen & Smith, 1994), treadmill stepping (Thelen & Ulrich, 1991), and reaching (Thelen, Corbetta, & Spencer, 1996).

The possible application of the CDST to study phenomena in second language acquisition (SLA) was first recommended by Diane Larsen-Freeman in her seminal paper 'Chaos/Complexity Science and Second Language Acquisition' in 1997. In this oft-cited article, Larsen-Freeman (1997) characterised the systems studied by CDST as "dynamic, complex, nonlinear, chaotic, unpredictable, sensitive to initial conditions, open, self-organizing, feedback sensitive, and adaptive" (p. 142). According to de Bot and Larsen-Freeman (2011), the basic features of the dynamic systems include: (1) sensitive dependence on initial conditions, (2) complete interconnectedness, (3) nonlinearity in development, (4) change through self-organisation and interaction with the environment, (5) dependence on internal and external resources, (6) constant change, with occasional chaotic variation, in which the systems only momentarily settle into attractor states, (7) iterative process, and (8) emergent properties. The first empirical application of the CDST to the study of language development was carried out by Verspoor et al. (2004), published in their influential article entitled 'Dynamic Systems Theory and Variation: A Case Study in L2 Writing'.

Complex Dynamic Systems Theory is closely aligned with a number of theories and approaches such as cognitive linguistics (Dirven & Verspoor, 2004; Langacker, 2008), emergentist and connectionist approaches (e.g. Elman, 1995), Grammaticalization theory (e.g. Bybee, 2008) and Activation theory (Rumelhart & McClelland, 1987). The meaningful interrelationships within the language system in general are one of the primary foci of cognitive linguistics (Verspoor & Behrens, 2011). Cognitive linguistics recognizes the three major propositions of CDST: (1) the language system is complex and dynamic, (2) the language system emerges through social interaction, and (3) the language system is in constant change (Verspoor & Behrens, 2011). CDST is also closely aligned with emergentist and connectionist approaches (e.g. Elman, 1995). The main focus of both Emergentism and Connectionism is on how it is possible that complexities can emerge in language through simple iterations. Complex Dynamic Systems Theory is also closely related to Grammaticalization theory (e.g. Bybee, 2008) which focuses on the detailed processes that might occur in language change. The idea that language learning is an iterative process is also congruent with Activation theory (Rumelhart & McClelland, 1987). Furthermore, MacWhinney's (2008) unified model focuses on the factors (such as the frequency of occurrence) that might play a role in language acquisition and use. All of these usage-based approaches either implicitly or explicitly recognize that patterns might emerge at the general system level (a language in general), and at a more specific level (individual level) (Verspoor & Behrens, 2011).

2.3 Features of Dynamic Systems

In the following sub-sections, the main features of CDST are presented: nonlinearity, variability, self-organisation, and complete interconnectedness. Furthermore, additional characteristics are also described such as iterative process, co-adaptation, attractor states, and sensitive dependence on initial conditions. The descriptions are also complemented with two relevant CDST abstractions: phase shift, and perturbation.

2.3.1 Nonlinearity

Dynamic systems are nonlinear "in which the effect is disproportionate to the cause" (Larsen-Freeman, 1997, p. 143). In contrast, in a linear system a cause of a specific strength leads to an effect of identical strength. Sometimes nonlinear systems might also show linearity, but at other times they might react in a way that is disproportionate to the cause. Vocabulary development is a clear example of nonlinearity in language development. For example, if L2 learners invest twice as much time in learning new vocabulary, they may not learn twice as many new words (Caspi, 2010; Caspi & Lowie, 2013). Nonlinearity is in close relation with the complete interconnectedness of the system (de Bot & Larsen-Freeman, 2011). Since there might be many interacting elements of the system, it is impossible to predict how the system will change. For example, vocabulary acquisition might be affected by many different factors such as the language learner's motivation, the time invested, and working memory.

Nonlinear patterns of development have been demonstrated by research on second language writing (Caspi, 2010; Larsen-Freeman, 2006; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2008, 2017) and second language vocabulary (Zheng, 2016). For example, Verspoor et al. (2008) found that the average word length (AveWL) index showed a nonlinear pattern of development in an advanced EFL learner's written data over three years. If the AveWL index had been measured at the beginning and at the end of the data collection, a linear increase could have been detected. However, the multi-wave research design demonstrated that the average word length index first declined, reaching the zenith at around the first third of the data collection. However, from the middle of the data collection the average word length index started to increase. The nonlinear pattern of the average word length index would not have been detected by a classic two-wave research design.

2.3.2 Variability

Variability in development has been at the heart of research on second language development (van Dijk, Verspoor, & Lowie, 2011). Variability exists between L2 learners (inter-individual variability). In other words, L2 learners differ from each other in many respects. For example, some L2 learners are more motivated to learn languages than others. Some L2 learners may acquire new words much faster than other L2 learners and so on. Furthermore, variability exists within L2 learners (intra-individual variability). In other words, L2 learners' language can also be different from one point in time to another. For example, sometimes L2 learners are capable of composing a text without any errors, but at other times they tend to make more errors in writing. Van Dijk et al. (2011) pointed out that the differences between and within L2 learners have often been considered "as unwanted by-products of the real data" in traditional two-wave longitudinal or panel studies and it was "implicitly assumed that the 'underlying' development must be a rather smooth, linear process and that the description of such processes must be able to generalize across learners" (p. 55).

Both inter-individual and intra-individual variability occur constantly in any complex dynamic system or any sub-system (Verspoor et al., 2008). The degree of variability might change depending on the stability of the system. Both free and systematic variability will be relatively high during self-organisation and low during system stability. Thelen and Smith (1994) pointed out that "variability is a metric of stability and a harbinger of change" (p. 342).

Verspoor et al. (2008) argued that the degree and the patterns of variability might provide useful information about the developmental process. Thelen and Smith (1994) also claimed that both intra- and inter-individual variability are essential characteristics of complex dynamic systems. Consequently, variability should be analysed rather than treated as measurement error (van Geert & van Dijk, 2002).

Variability in language development has been demonstrated by numerous studies (Caspi, 2010; Larsen-Freeman, 2006; Rosmawati, 2016; Verspoor et al., 2008). For example, Larsen-Freeman (2006) examined the oral and written production of five Chinese leaners of English over six months. Although the group averages of measures of complexity, accuracy and fluency (CAF) displayed an increasing trend, the individual growth trajectories showed ebbs and flows over time. The CAF measures showed both inter-individual and intra-individual variability in the data of the five Chinese learners.

2.3.3 Phase Shift

Intra-individual variability – both systematic and free – constantly occurs in complex dynamic systems. However, the degree of variability may change depending on the stability of the system (Verspoor et al., 2008). Unstable periods can be a sign of change in a complex system. Conversely, stable periods may suggest that the system might have settled into an attractor state (for a discussion see section 2.3.9). Dramatic and sudden changes in complex systems are called phase shifts or bifurcations (Larsen-Freeman & Cameron, 2008). Nonlinearity, self-organisation and emergence might be the source of a phase shift (Larsen-Freeman & Cameron, 2008). In language development, phase shifts or phase transitions refer to significant changes in the development of a system. Van Dijk et al. (2011) pointed out that variability reflects actual data that should be scrutinised to detect how systems change from one phase to the next.

Phase transitions or significant developmental peaks were detected in a number of studies on second language development from a CDST perspective (Piniel & Csizér, 2015; Spoelman & Verspoor, 2010; van Dijk et al., 2011; Zheng, 2016). Verspoor et al. (2018) also found near statistically significant developmental peaks in the mean length of T-unit (MLTU) and Guiraud's indices in the written data of two learners in a study including 22 Dutch learners of English, aged 12-13.

2.3.4 Self-Organisation

Self-organisation, sometimes referred to as spontaneous order, is a process in which some type of overall order emerges from interactions among elements of a disordered system. Self-organisation and emergence are different ways of describing the source of phase shifts in the behaviour of complex dynamic systems (Larsen-Freeman & Cameron, 2008). A dynamic system is believed to self-organise into a new pattern of behaviour after a phase shift. The phase shift is self-organised without any external organising force. Self-organisation might occur due to the adaptation of a system in response to changes and it might result in new phenomena on a different scale or level, called emergence (Larsen-Freeman & Cameron, 2008). Van Geert (2003) claimed that "it is highly likely that language acquisition, similar to many other processes that involve growth and development, is such a self-organizational process" (p. 659).

De Bot and Larsen-Freeman (2011) exemplified self-organisation in language learning by the acquisition of synonyms. When a language learner knows only one word for consuming beverages, 'drink', all type of drinking will be referred to using this word. When the language learner acquires a synonym for drinking such as sipping, the lexical system will self-organise itself by differentiating between different types of drinking.

2.3.5 Complete Interconnectedness

All elements are connected to all other elements in a dynamic system. In addition, the elements or processes of a complex dynamic system are themselves complex dynamic sub-systems. For example, the linguistic system is a complex dynamic system containing many different complex dynamic sub-systems such as the lexical system, syntactic system and phonological system. Since complex dynamic systems are completely interconnected, a change in one system might have an effect on all the other sub-systems. However, different connections have different strengths in a dynamic system (de Bot & Larsen-Freeman, 2011). Some connections might be loose, while other connections might be much stronger. In dynamic parlance, some variables might support or compete with each other, and one might be conditional upon another. For example, at beginner stages of both first and second language development it might be presumed that the interaction between the lexical and the syntactic systems is conditional. A certain number of words must be learnt before the language learner can start developing his or her syntactic system.

The complete interconnectedness of complex dynamic systems has been extensively explored by previous studies on language development. A moving window of correlations is a frequently applied technique to visualise the interactions between the complex dynamic systems (see 4.8.1 for a detailed discussion). For example, Verspoor et al. (2008) found a non-significant overall correlation between sentence length (SL) and type-token ratio (TTR) over time in the written data of their advanced learner. However, a moving window of correlations revealed dramatic oscillations between the two variables suggesting that there was a complex dynamic interaction between TTR and SL. The interaction between SL and TTR changed from moderately positive values to strong negative values, and then back to moderately positive values over time. The shift between moderately positive and strong negative values indicates that SL and TTR share a competitive relationship with each other. Verpsoor et al. (2008)

concluded that when academic writing proficiency increases, there might be a trade-off between more varied word use and longer sentences at different stages in the developmental process.

2.3.6 Iterative Process

Growth is considered to be an iterative process from a Complex Dynamic Systems Theory perspective. The current level of development is crucially dependent on the previous level of development (Van Geert, 1994). De Bot and Larsen-Freeman (2011) illustrated the iterative process with dance. Dance as a system is not just the repetition of steps, but rather the complex dynamic interaction between the two dancers, their steps, and the environment. The dance is the emergence of the repetition of the steps in a constantly changing environment. The rhythm and the melody of the music, the intentions of the dancers are in constant change from step to step. The dance metaphor is a clear example of the iterative process.

An iterative process can be best demonstrated by dynamic, non-linear and stochastic models of development and change. A model is considered dynamic when it incorporates the iterative process. The best-known example of a dynamic model is the logistic map generated by the logistic equation. Although the logistic equation, discovered by Verhulst in 1938, was originally used to describe population development, in 1976 Robert May discovered that the logistic equation can reproduce the chaotic behaviour of dynamic systems. The logistic equation is given in Equation 2.1:

Equation 2.1 Logistic equation

$$\frac{\Delta P}{\Delta t} = rP(1 - \frac{P}{K})$$

In Equation 2.1, P is the population size, while t denotes the time. Δ denotes the change or difference, while r is the growth rate. The carrying capacity, “the final attainable growth level based on the resources available" (van Geert, 1995, p. 316), is represented by K. The left term of Equation 2.1 (rP) expresses the early growth. However, the early growth is curbed when the second term (1-P/K) becomes larger than the previous term. In other words, the population size increases towards its carrying capacity (K). The resulting damping effect reduces r until the population stops increasing. The logistic equation generates an S-shape curve which is a prototype for numerous developmental and learning phenomena.

Language development has been infrequently modelled by using the logistic equation due to its complicated applications. Van Geert (1991) simulated the nonparallel development of lexicon and syntax in spoken early L1 data with the precursor model. Caspi (2010) successfully managed to support the hypothesis that complex interactions between vocabulary knowledge levels give rise to the receptive-productive gap by the precursor model. Four knowledge levels (recognition, recall, controlled production, and free production) of English as a second language (ESL) vocabulary were traced over a 36-week period and detailed analyses revealed that knowledge levels interact in a complex way, simultaneously competing for learner resources and conditionally supporting each other's growth. Caspi (2010) also applied the precursor model to simulate the L2 writing data of the same advanced learners used in the study on vocabulary development. The model specified lexical complexity as a precursor to lexical accuracy, which is in turn a precursor to syntactic complexity, itself a precursor to syntactic accuracy.

Previous studies related inverse variability patterns across lexicon and syntax with dynamic precursor interactions (Verspoor et al., 2008). Caspi (2010) claimed that the optimized model values confirmed the hypotheses concerning the interactions between the levels in the precursor hierarchy for the Portuguese speaker in a multiple case study of vocabulary development with learners of different first languages. However, the hypotheses were only partially confirmed for the Mandarin and Indonesian speaker and the values of the relational control parameters were mostly incongruent with the predictions for the Vietnamese speaker. Caspi (2010) explained the incongruity between the model and the predictions as follows: “This may be due to the distinction of this participant, who was generally more proficient, and slightly older than the others” (p. 165). Furthermore, Caspi (2010) claimed that the basic precursor model structure is upheld, although the exact threshold of the order parameters was not addressed in her study. It can be presumed that the precursor model works well for data of beginner learners, where there could be a hierarchy in the development. However, at advanced stages a coupled logistic equation might work better.

2.3.7 Chaotic Variation

The path of a dynamic system might be difficult to predict due to the interaction of variables over time. Language learners might experience phases of fluency and disfluency for no apparent or discernible reason. “Chaos refers simply to the period of complete randomness that complex nonlinear systems enter into irregularly and unpredictably” (Larsen-Freeman, 1997, p.143). Van Geert (1994) attributed chaotic oscillations to “deep intrinsic property of the processes themselves” and not to the consequence of external, independent factors (p. 84).

2.3.8 Co-Adaptation

Co-adaptation or co-evolution refers to the interaction of two or more complex dynamic systems (Larsen-Freeman & Cameron, 2008). In other words, co-adaptation describes “a kind of mutual causality, in which change in one system leads to change in another system connected to it, and this mutual influencing continues over time” (Larsen-Freeman & Cameron, 2008, p. 233). Larsen-Freeman and Cameron (2008) mentioned numerous examples of co-adaptation in applied linguistics. For instance, infants and care-givers co-adapt their language and communication, speakers accommodate to one another in accent and vocabulary in conversation, and teacher and students might co-adapt in classroom behaviour and discourse (Larsen-Freeman & Cameron, 2008). Co-adaptation can also occur within the complex dynamic system and not only between two closely related systems (Oyama, Griffith, & Gray, 2001). For instance, cancer cells might mutate or multiply in response to perturbations triggered by surgery or taking anticancer drugs (see section 2.3.10 for a discussion).

Co-adaptation might occur between and within the linguistic and the motivational systems. For example, Nitta and Baba (2015) investigated how the ideal L2 self develops and evolves through repeated engagements in language learning tasks over an academic year. Self-regulation was adopted to represent the third component (L2 learning experience) in the L2 Motivational Self System. The analysis of the changes in the participants' self-regulatory processes and L2 writing revealed that the ideal L2 self evolves co-adaptively with the development of micro-level functioning of self-regulation. Furthermore, it was found that the self-evaluation process played a more important role than the self-observation processes in facilitating the development of second language writing. Nitta and Baba (2015) argued that co-adaptation or co-evolution might occur between the ideal L2 self and self-regulation processes in a sense that the "ideal L2 self functions as an evaluative standard for self-regulation, while self-regulatory processes influence the formation and revision of the ideal L2 self" (p. 372).

2.3.9 Attractor States

Complex dynamic systems are in constant change. However, they or their sub-systems sometimes settle into attractor states. According to Larsen-Freeman and Cameron (2008), "an attractor is a region of a system's state space into which the system tends to move" (p. 50). In second language development, fossilisation is considered to be an attractor state (de Bot and Larsen-Freeman, 2011). Fossilisation describes stagnation in second language development. For instance, some L2 learners manage to acquire one element of the language system but fail to develop it further. In some cases, a large amount of energy is required to move the system out of an attractor state.

Motivation has been shown to fluctuate over time (MacIntyre & Serroul, 2015). Nonetheless, motivation also tends to self-organise and settle into relatively stable attractor states like the ideal L2 self (Hiver, 2015). Dörnyei (2009c) has proposed that the conglomerates of specific learner characteristics might form attractor states such as extroversion and introversion (Matthews & Zeinder, 2004), and the ideal L2 self (Dörnyei, 2009c) on the longer timescales, and interest (Ainley, 2006; Dörnyei, 2009c; Hidi, Renninger & Krapp, 2004) on the shorter timescales.

Waninge (2015) focused on identifying the main attractor states that make up the learning experiences in a language classroom. The participants, enrolled in a language development module, were second year students of English at a British university. The interviews were kept short (from seven to nine minutes) intentionally because it was expected that a dominant attractor state would surface in the first couple of minutes of an interview. From the content analysis of a sufficient sample of focused qualitative interviews, four distinct categories emerged: interest, boredom, neutral attention and anxiety. Waninge (2015) claimed that interest, the most frequently mentioned state, is one of the most important attractor states in the L2 learning experience since it is closely connected to "learners' active engagement and future motivational intention in the classroom" (p. 201). Boredom, a passive and negative affective dimension, was also frequently mentioned by the participants. Boredom has been found to be negatively connected to academic achievement (Baker, D'Mello, Mercedes, Rodrigo, & Graesser, 2010; Dewaele, 2011; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). Neutral attention was also reported by the participants, although less frequently than interest and boredom. Anxiety was the fourth salient attractor state mentioned by the participants. Anxiety was found to be more detrimental for learning than boredom (Waninge, 2015). It was also found that anxiety appeared to wax and wane during lessons. Although some participants reported switching from one state to another within a single lesson, there was consistent data that pointed to relatively stable states.

2.3.10 Perturbations

Events that disrupt the stability of a complex dynamic system are called perturbations. Perturbations to the systems might lead to phase shifts (Larsen-Freeman & Cameron, 2008). Larsen-Freeman and Cameron (2008) pointed out that even the smallest perturbations might cause the complex dynamic system to move from one state, particularly from a chaotic or strange attractor state to another. Hiver (2015) indicated that feedback and perturbations are the most common types of inputs. Iterations of positive feedback might push the system from one attractor state to another. Hiver (2015) exemplified perturbations with the sudden news of an unexpected language exam, the subsequent poor test results, and the teacher's use of an extrinsic reward or prize.

Henry (2015) examined processes of change in six participants' motivational systems. The participants came from a class of 22 first-year students studying French at the A2.2 Common European Framework of Reference (CEFR) level. A series of semi-structured interviews was carried out five times with three participants and six times with the other three participants over an eight-month period. The interviews were analysed by using the abstractions of Complex Dynamic Systems Theory and possible selves theories as a compass. In the interviews, indications of changes in attractor states, perturbations, phase shifts, co-adaptation between the L3 and L2 motivational systems and features of emergence were sought. It was found that the motivational system displayed considerable fluctuations and variation over time and the momentary shifts in students' motivation could be caused by a wide range of different factors. These factors included perceptions concerning the enjoyment, meaningfulness, relevance of a particular activity, the effect of perturbations, such as receiving test results, and the continuous interaction with other complex dynamic systems. Henry (2015) found that perturbations triggered a process in which two identity-transforming phase shifts occurred in one student's L3 motivational system.

2.3.11 Sensitive Dependence on Initial Conditions

There is sensitive dependence on initial conditions in dynamic systems. This phenomenon has become well-known as the "butterfly effect", which refers to the famous example of the meteorologist Edward Lorenz, who found that minimal differences in initial conditions of systems might lead to huge effects later on. The same phenomenon can be observed in language development. For example, minimal differences between language learners might result in different learning outcomes even if the learning experience is similar. Verspoor (2015) defined initial conditions as "the conditions sub-systems are in when the researcher starts measuring" (p. 45). However, Larsen-Freeman (2015) also indicated that sensitive dependence on initial conditions does not necessarily mean that the system is only sensitive at the onset of the development because even minor influences can lead the system in a different direction.

Chan (2015) and Chan, Verspoor and Vahtrick (2015) found that even identical twins showed inverse trends of syntactic development over eight months. A great deal of variability was found within each learner, although many different variables were controlled in the study. When the twin learners were compared on measures of syntactic complexity over time employing the Hidden-Markov Model (HMM), there were some resemblances but also clear dissimilarities between the two learners. Both learners' data displayed more syntactic complexity in speaking early on, and one of the participants' syntactic complexity in speaking remained stable over time, while the other participant displayed a shift to more complexity in writing than speaking in the second and third stages as indicated by the Hidden-Markov Model. Chan et al. (2015) attributed the dissimilar development of written syntactic complexity between the twins to the competition between lexical and syntactic processing. Chan et al. (2015) concluded that "even identical twins with similar personalities and interests who are exposed to similar input within the same environment may demonstrate different developmental paths" (p. 318).

Henry (2015) observed students' social interactions by tracing where they sat during the lesson. Among the six participant there was one student (Tim) whose seating patterns were more fluid. Tim felt that the student who was sitting next to him influenced the state of his motivation. Henry (2015) claimed that Tim's choice functioned as an initial condition that would have a determining influence on the trajectory of the system over the following period.

2.4 Summary

In Chapter 2, I focused on the theoretical framework of this study: the Complex Dynamic Systems Theory. First, I provided a brief overview of previous influential theories of second language acquisition. Second, I presented the Complex Dynamic Systems Theory and then I listed its key features. In Chapter 3, I will present the literature review on second language writing development and self-regulation.

Chapter 3. L2 Writing Development and Motivation

Chapter Overview

Chapter 3 reviews the literature on L2 writing development and L2 writing motivation. There are seven main sections: (1) second language writing development, (2) linguistic measures in second language writing development, (3) linguistic development in two-wave longitudinal studies, (4) linguistic development in multi-wave longitudinal studies, (5) interactions between complexity and accuracy, (6) second language writing motivation, and (7) co-adaptation of second language writing and motivation.

3.1 L2 Writing Development

L2 writing is both a cognitive process and a situated activity (Polio & Friedman, 2017). The cognitive process refers to the writer's "set of internalized skills and knowledge to produce a text" (Polio & Friedman, 2017, p. 1), while the situated activity denotes the specific context in which writing takes place. Consequently, L2 writing research includes a large array of topics while employing a great variety of research traditions. According to Cumming (2016), "L2 writing is inherently multi-faceted, involving multiple issues and orientations that may not even be commensurable with each other" (p. 65). Consequently, the types of L2 writing research are varied and broad, covering many different areas of second language acquisition and applied linguistics (Polio & Friedman, 2017).

Manchón (2011) made a bipartite pedagogical distinction between the types of L2 writing research: (1) learning-to-write (LW), and (2) writing-to-learn (WL). Furthermore, she distinguished between two types of L2 writing research within the writing-to-learn orientation: (2a) writing-to-learn-content (WLC), and (2b) writing-to-learn-language (WLL). The learning-to-write (LW) orientation refers to "the manner in which second and foreign (L2) users learn to express themselves in writing", while the writing-to-learn (WL) orientation denotes "the way in which the engagement with L2 writing tasks and activities can contribute to development in areas other than writing itself", "be it content knowledge" (WLC), or "language knowledge and skills" (WLL) (p. 3).

According to Cumming (2001), the learning-to-write orientation subsumes three dimensions: (1) textual features, (2) composing processes, and (3) sociocultural context. Earlier L2 writing research on the textual features dimension investigated the development of lexis, syntax and morphology at the micro level, and cohesive devices and text structure at the macro level. L2 writing research on the composing processes dimension investigated how writers search for words and syntax and pay attention to ideas and language concurrently at the micro level, and how L2 writers plan and revise at the macro level. Past studies on the contextual factors investigated the L2 writers' individual development and self-image or identity at the micro level, while at the same time investigated the L2 writers' participation in a discourse community and social change at the macro level. Cumming (2001) pointed out that previous research has focused on the three dimensions separately, although they are "integrally interrelated" (p. 1). Figure 3.1 illustrates the bipartite orientations of L2 writing research, highlighted in blue, identified by Manchón (2011) and the three dimensions of the writing-to-learn orientation, highlighted in red, identified by Cumming (2001).

Writing development is most frequently equated with linguistic development. However, L2 writing development has been investigated from multiple perspectives (Manchón, 2012), such as from a goal theory perspective (Cumming, 2012), a rhetorical genre theory perspective (Tardy, 2012), a genre-based systemic functional linguistic approach (Byrnes, 2012) and a socio-cultural approach (Wigglesworth & Storch, 2012). Wolfe-Quintero, Inagaki, and Kim (1998) defined language development as "characteristics of a learner's output that reveal some pointed or stage along a developmental continuum" (p. 2).

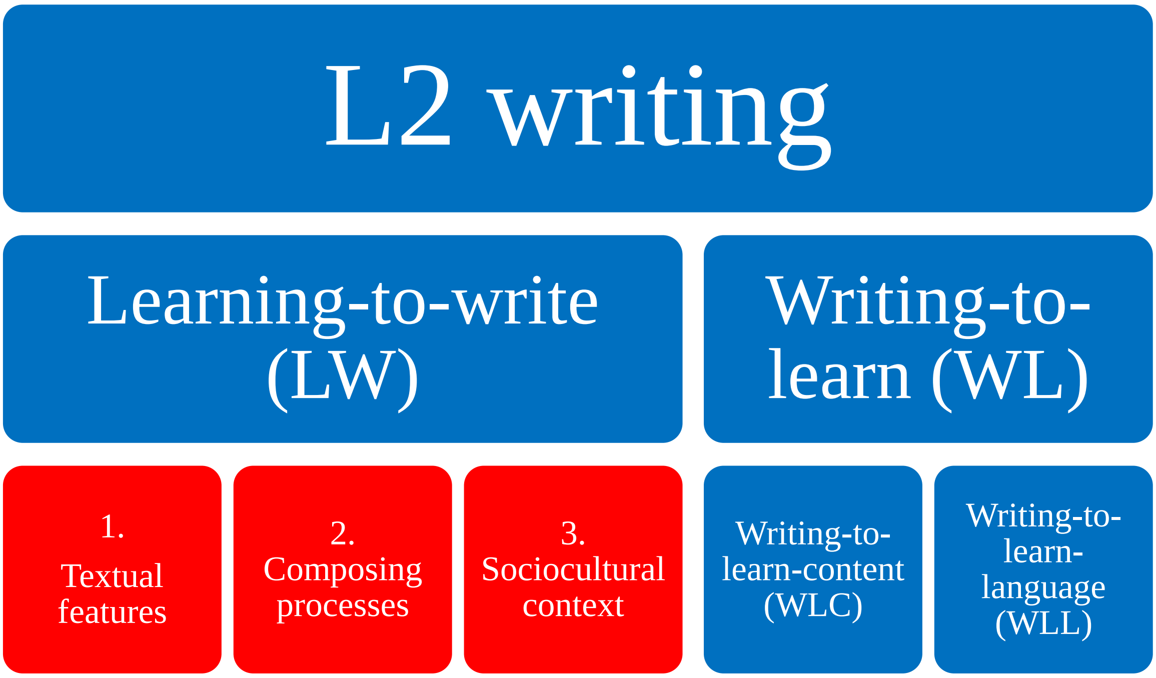


Figure 3.1 The orientations and dimensions of L2 writing research based on Manchón (2011) and Cumming (2001).

Their definition made a clear distinction between proficiency and development. Interestingly, Manchón (2012) refrained from providing a definition of development in her book "L2 Writing development: Multiple Perspectives". A more recent definition of language development was proposed by Polio (2017), who defined writing development as "change over time in any of the following areas related to written text production: language (e.g., complexity, accuracy, fluency, cohesion, mechanics); knowledge of different genres; text production processes; metacognitive knowledge and strategy use; and writing goals and motivation" (p. 261). It is also important to note that development does not necessarily mean improvement in quality or movement toward a target norm. Nonetheless, development most frequently refers to progress toward some target that might vary in relation to context and the purpose of writing (Polio, 2017).

3.2 Linguistic Measures in L2 Writing Development

It is now widely accepted that the constructs of L2 performance and L2 proficiency can be sufficiently captured by the notions of complexity, accuracy and fluency (CAF) (R. Ellis, 2003, 2008; R. Ellis & Barkhuizen, 2005; Skehan, 1998). The CAF has been used for many different purposes such as performance descriptors for the oral and written assessment of language learners, for indicators of students' proficiency underlying their performance, and for the investigation of second language development (Housen & Kuiken, 2009).

L2 learners' texts can be analysed from different approaches. Previous studies focused on the following categories: (1) lexical complexity (e.g. Gebril & Plakans, 2013; Kormos, 2011; Li, 2000; Verspoor, Schmid, & Xu, 2012), (2) syntactic complexity (e.g. Lu, 2011; Vyatkina, 2012), (3) accuracy (e.g. R. Ellis &Yuan, 2004); (4) fluency (e.g. Storch, 2005), (5) formulaic sequences (e.g. Durrant & Schmitt, 2009; Li & Schmitt, 2009), (6) cohesion (e.g. Crossley, Weston, Sullivan, & McNamara, 2011), (7) paraphrasing and text copying (e.g. Petrić, 2012), and (8) revision process (e.g. Lavolette, Polio, & Kahng, 2015; Suzuki, 2012).

3.2.1 Complexity

In the SLA research literature, complexity refers to two different types of complexities: (1) L2 complexity and (2) task complexity. L2 or linguistic complexity is generally measured by tracing changes in the constructs of lexicon and syntax. However, there are different dimensions of L2 complexity (see Bulté & Housen, 2012 for a taxonomy). Complexity is the most ambiguous and complex construct of the CAF triad (Housen & Kuiken, 2009) and it is generally defined as "[t]he extent to which the language produced in performing a task is elaborate and varied" (R. Ellis, 2003, p. 340). According to Bulté and Housen (2012), system complexity refers to "the degree of elaboration, the size, the breadth, the width, or richness of the learners' L2 system" (p. 25). In line with Bulté and Housen (2012, 2014), Verspoor et al.

(2017) defined linguistic complexity as a "quantitative property of language units" that is "the greater the number of components a construction has and the more levels of embedding it contains, the more complex it is" (p. 1).

Nonetheless, Mazgutova and Kormos (2015) took a novel approach to analyse complexity since they argued that "complexity of learners' output should be considered with reference to the mode, genre, and communicative demands of the particular task to be performed" (p. 2). Mazgutova and Kormos's (2015) approach challenged Bulté and Housen's (2012, 2014) definitions of absolute and relative complexities. The former refers to "*objective* inherent properties of linguistic units and/or systems thereof", while the latter denotes the "cost and difficulty of processing or learning" (Bulté & Housen, 2014, p. 43, italics in original).

3.2.2 Lexical Complexity

The range, variety, or diversity of words found in L2 learners' language use is considered to reflect the complexity of their vocabulary knowledge and the level of their language proficiency (Jarvis, 2013a). There is a consensus that lexical complexity is a multifaceted construct (Bulté & Housen, 2012; Jarvis, 2013b; Schmitt, 2010). Nevertheless, there has been no agreement in the categorisation, definition, and operationalisation of the construct of lexical complexity. Lexical complexity generally includes lexical diversity, lexical density and lexical sophistication such as in Bulté and Housen's (2012) framework. Conversely, according to Jarvis's (2013b) model, lexical diversity is used as an umbrella term to include six different properties: variability, volume, evenness, dispersion, rarity and disparity. Furthermore, Kim, Crossley, and Kyle (2018) claimed that lexical sophistication is also a multidimensional phenomenon.

Lexical complexity in L2 learners' written production is generally measured by the constructs of lexical diversity, lexical sophistication, and lexical density. The definition and the operationalisation of lexical sophistication and lexical density are unambiguous. The former refers to the language learner's command of less-frequent words, while the latter denotes the proportion of content words in a text. However, the definition and operationalisation of lexical diversity are more problematic. Lexical diversity is generally defined as the variety of words or relative absence of lexical repetition found in a language learner's written or oral product (Carroll, 1938). Nevertheless, there has been a long controversy and confusion in connection with the measurement of the construct of lexical diversity due to the absence of a clear definition in the literature (Jarvis, 2013b). The term lexical diversity has been used interchangeably with the terms lexical variation, lexical variety, lexical variability, and lexical flexibility (see for example Engber, 1995; Johnson, 1944; Read, 2000). Furthermore, lexical richness, originally defined as "wealth of words at [the author's] command" (Yule, 1944, p. 83), has also been used as a synonym of lexical diversity referring to the variety of words found in a language sample (e.g. Daller, Van Hout, & Treffers-Daller, 2003). In addition, lexical richness has been used as a superordinate term to cover all lexical constructs including lexical diversity and lexical sophistication (Engber, 1995; Read, 2000). However, Jarvis (2013a) argued that lexical diversity is not a subset of lexical richness.

A wide range of lexical diversity and sophistication measures have been proposed in the second language literature. Lexical diversity is generally measured by the number of word types, type-token ratio (TTR) (Johnson, 1939, 1944; Templin, 1957), mean segmental type-token ratio (MSTTR), Guiraud index (Guiraud, 1954), word types per words, and D, also known as VOCD, (Malvern & Richards, 1997; Malvern, Richards, Chipere, & Durán, 2004), and measure of textual lexical diversity (MTLD) (McCarthy, 2005). Lexical density is usually measured by the lexical word per function words or the lexical words per total words indices. Lexical sophistication is generally gauged by the less frequent words per total words index. Numerous developmental studies have investigated the extent to which these measures can be reliable indicators of L2 proficiency (Bulté & Housen, 2014; Wolfe-Quintero et al., 1998). Previous research has shown that language learners with higher language proficiency produce texts with greater lexical diversity (e.g. Engber, 1995; Grant & Ginther, 2000; Jarvis, 2002; Reppen, 1994). The vast majority of lexical diversity indices involve statistical relationships between types and tokens. Jarvis (2013b) pointed out that the validation of most lexical diversity indices has been based on two criteria: (1) how well the index can overcome sample-size effects (text length), and (2) how well the index can predict vocabulary knowledge. However, most lexical diversity indices have not been validated in accordance with how well they actually measure the construct of lexical diversity (Jarvis, 2013b).

Lexical diversity has long been considered as the inverse of lexical repetition. However, Jarvis (2013b) argued that lexical diversity is a perception-based phenomenon such as lexical redundancy, whereas lexical repetition is a purely statistical phenomenon such as lexical variability. For example, lexical indices that are calculated only from type and token frequencies (e.g. type-token ratio, Guiraud's index) reduce the text to only two categories: (1) types, and (2) tokens or (1) "first occurrences" and (2) "repetitions" (Jarvis, 2013b, p.18). "Tokens are the number of running words in a text, while types are the number of different words" (Schmitt, 2010, p.188). For instance, Schmitt's 18-word long definition is comprised of eighteen tokens. However, three words ('are', 'the', and 'of') are repeated twice in his citation which reduces the number of types to 14. An inequality between the number of types and tokens is the result of the repetition. The less repetition there is in a specific text, the more lexically diverse the text is. However, certain words are repeated due to grammatical or pragmatic constraints. For example, in Schmitt's 18-word long definition (see above) the word 'are' is repeated due to grammatical constraints. However, certain words are repeated redundantly in a specific text (Bazzanella, 2011). For example, if we paraphrase Schmitt's definition as - tokens are the number of running words in a text, while types are the number of different words in a text - the last three words (in a text) are redundantly repeated. The paraphrase exemplifies that repetition is an objective phenomenon, while redundancy is fundamentally subjective. Jarvis (2013b) concluded that the type-token ratio-based indices do not measure lexical diversity, but lexical variability.

Jarvis (2013b) argued that “lexical diversity exhibits a great deal of multidimensional complexity” (p. 37). Consequently, lexical diversity cannot be captured sufficiently by measures that reduce texts to first occurrences and repetitions such as the type-token ratio or the measure of textual lexical diversity (MTLD). Jarvis’s (2013b) model of lexical diversity, grounded in human perception, includes six properties associated with the human perception of diversity. These properties include (1) variability, (2) volume, (3) evenness, (4) rarity, (5) dispersion, and (6) disparity. Variability can be best captured by a type-token ratio-based index such as the measure of textual lexical diversity index which does not vary as a function of volume (McCarthy, 2005; McCarthy & Jarvis, 2010). However, the most common measure of lexical variability is the type-token ratio (Johnson, 1939, 1944) which received much criticism due to its sensitivity to text length. Volume can be calculated by adding the number of words (tokens) in a text, while evenness can be measured by the standard deviation of the number of tokens per type in a text. Rarity can be captured by assessing the overall frequency of the words in a text in relation to how frequently those words can be found in the language. Mazgutova and Kormos (2015) recommended the log frequency of content words index to measure rarity. Dispersion can be measured by the mean distance between tokens and types, while disparity can be measured by the mean number of words per sense or the latent semantic analysis (LSA) index. However, Jarvis (2013b) suggested that volume (i.e. the total number of words produced in the text), evenness (i.e. “how evenly word tokens are distributed across types”) (p. 23), and dispersion (i.e. “the mean distance between different tokens of the same type”) (p. 30) suffer from multicollinearity.

Lambert and Kormos (2014) proposed some directions for identifying more developmentally based measures of L2 task performance and they claimed that "both cross-sectional and longitudinal research on L2 development across proficiency levels, discourse genres, production modes, and target languages is important in accurately conceptualizing and measuring the effects of task performance on L2 development" (p. 6). For example, to measure lexical complexity in academic writing Mazgutova and Kormos (2015) claimed that Academic Word List (AWL) was a reliable indicator to estimate the percentage of academic words in written texts. Furthermore, a multidimensional measurement of L2 lexical complexity also involves the addition of a general L2 lexical complexity index. Verspoor et al. (2017) found that average word length (AveWL) is the best measure of general lexical complexity. Figure 3.2 illustrates a multidimensional measurement of lexical complexity based on Jarvis (2013b), Verspoor et al. (2017), and Mazgutova and Kormos (2015).

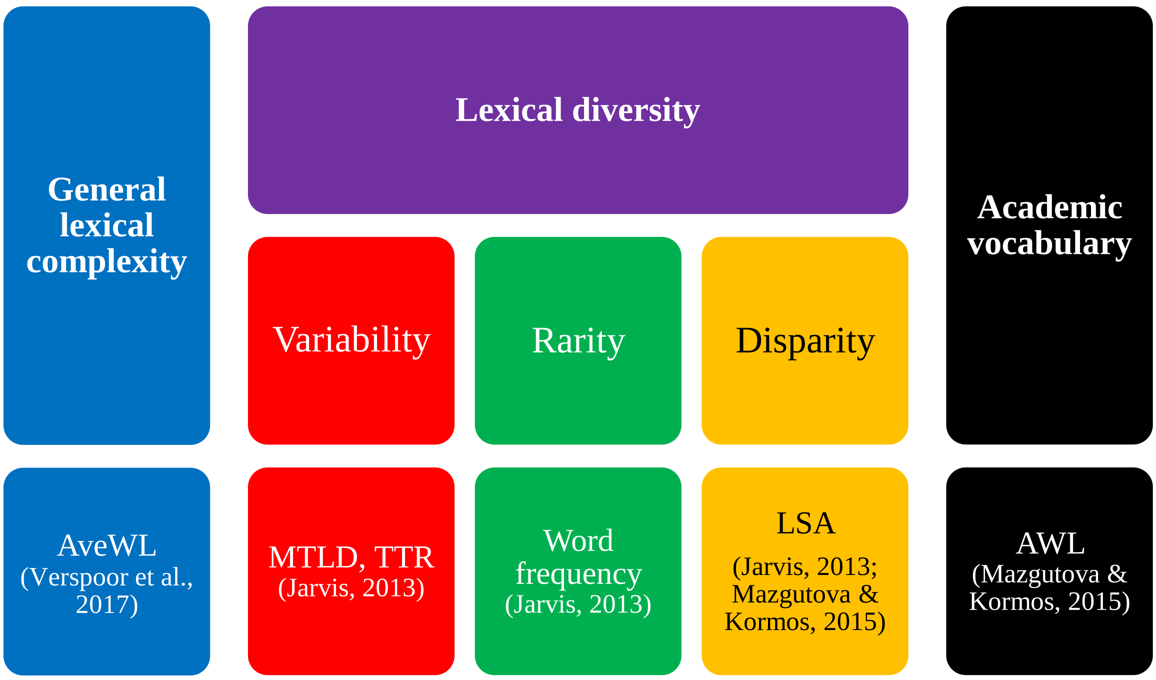


Figure 3.2 A multidimensional measurement of L2 lexical complexity as suggested by Jarvis (2013b), Verspoor et al. (2017), and Mazgutova and Kormos (2015)

3.2.3 Syntactic Complexity

Syntactic complexity has been conceptualised as a multifaceted or multidimensional construct similar to lexical complexity (Bulté & Housen, 2014; Lu, 2011, 2017; Norris & Ortega, 2009). Syntactic complexity development is most frequently traced by the quantification of specific grammatical structures such as the number of dependent clauses in a given text. The quantification of certain syntactic structures is usually expressed in ratios (for example dividing the dependent clauses by the total number of clauses in a text). If a specific measure increases between two points in time, its development or growth is detected. Numerous metrics have been employed to trace syntactic complexity development to date and they are well-documented by several synthesis analyses (Wolfe-Quintero et al., 1998; Ortega, 2003).

Length-based indices, or general complexity indices, are usually calculated by dividing words, morphemes or characters by a specific production unit such as the T-unit which is the “shortest grammatically allowable sentences into which the theme could be segmented” (Hunt 1965, p.12). Length-based measures are extensively used in child language acquisition (Brown, 1973) and in L1 writing (Hunt, 1965; Loban, 1976; Scott, 1988). Examples of length-based measures are the mean length of utterance (MLU) (Brown, 1973), the mean length of T-unit (MLT) (Hunt, 1965), the mean length of c-unit (MLCU) (Loban, 1976) and the mean length of clause (MLC) (Scott, 1988). Norris and Ortega (2009) suggested employing any length-based indices to measure global syntactic complexity with a multiple-clausal unit of production (for example the T-unit) in the denominator. Lu (2017) recommended the number of clauses per sentence (C/S) index to measure overall sentence complexity, the mean length of clause (MLC), the mean length of sentence (MLS), or the mean length of T-unit (MLT) indices to measure clause length. Verspoor et al. (2017) recommended finite verb ratio (FVR) as the best general syntactic complexity index.

Subordination is usually calculated by counting all clauses and dividing them by a specific production unit. Subordination measures usually include the mean number of clauses per T-unit (Elder & Iwashita, 2005), the mean number of clauses per c-unit (Skehan & Foster, 2005), the mean number of clauses per analysis of speech unit (Michel, Kuiken, & Vedder, 2007), or the mean number of dependent or subordinate clauses per total clauses. Norris and Ortega (2009) recommended using any index to gauge subordination with clause in the numerator. Lu (2017) suggested using the number of clauses per T-unit (C/T), the complex T-unit ratio (CT/T), the number of dependent clauses per clause (DC/C), or the number of dependent clauses per T-unit (DC/T) to gauge the amount of subordination.

Coordination measures, such as the coordination index (CI) (Bardovi-Harlig, 1992), were found more reliable at beginner levels of L2 competence. Lu (2017) suggested using the number of T-units per sentence (T/S) to measure the amount of clausal coordination, while the number of coordinate phrases per clause (CP/C), the number of coordinate phrases per T-unit (CP/T) to gauge phrasal coordination.

Phrasal complexity has been a rather neglected domain of syntactic complexity. Norris and Ortega (2009) proposed measuring phrasal complexity or “subclausal complexity” (p. 561) by the mean length of clause (MLC). However, Lu (2017) recommended the number of complex nominals per clause (CN/C), the number of complex nominals per T-unit (CN/T), or the number of verb phrases per T-unit (VP/T) to measure the degree of phrasal sophistication. Mazgutova and Kormos (2015) found that the mean number of modifiers per noun phrase (SYNNP), and the mean number of complex nominals (CN) were reliable indicators of development of phrasal complexity in academic writing.

Genre-specific syntactic measures are infrequently employed in the field of L2 writing development. However, Mazgutova and Kormos (2015) found that the frequency of conditional and relative clauses increased, while the frequency of infinitive clauses and prepositional phrases decreased in the academic argumentative texts of intermediate EFL learners over a one-month intensive English for Academic Purposes (EAP) programme at a British university.

Research on syntactic complexity development has been influenced by four major claims over the past ten years:

1. the development of syntactic complexity is claimed to be measured multi-dimensionally (Norris & Ortega, 2009),
2. academic writing is characterised by phrasal complexity instead of clausal elaboration (Biber et al., 2011),
3. the clause is a more informative unit of analysis than the T-unit (Lu, 2011),
4. complexity, accuracy, and fluency are claimed to be measured by developmentally oriented indices (Lambert & Kormos, 2014).

The first major claim is that syntactic complexity needs to be measured multidimensionally (Norris & Ortega, 2009). In other words, in order to capture the multifaceted nature of syntactic complexity, multiple measures are necessary. However, the selection of the indices to measure syntactic complexity should be warranted both theoretically and empirically. Norris and Ortega (2009) pointed out that the use of more than one index from the same type of measures can be redundant since they tap into the same dimension of syntactic complexity. Conversely, some indices are distinct, hence they should be used together with other types of measures. Lu (2011) also recommended the selection of multiple measures that do not correlate highly with each other. He found that complex nominals per clause (CN/C), mean length of clause (MLC), mean length of sentence (MLS), and coordinate phrase per clause (CP/C) significantly discriminated between developmental levels. Therefore, he recommended these indices as best options for developmental indices. Biber et al. (2011) also claimed that syntactic complexity is not a unified construct. Therefore, it is not reasonable to presume that any single measure will sufficiently represent this construct. Apart from empirical findings, there are also theoretical justifications for a multidimensional measurement of syntactic complexity. According to the Systemic Functional Linguistics (SFL) approach (Halliday & Matthiessen, 1999), syntactic development proceeds from coordination (parataxis), via subordination (hypotaxis) to nominalisation. Consequently, L2 writers tend to rely on coordination at beginner levels (Bardovi-Harlig, 1992), on subordination at intermediate levels, and on nominalisation at advanced levels.

The second major claim is that academic writing is characterised by phrasal complexity (Biber et al., 2011). Biber (2006) found that L1 English academic registers are characterised by higher levels of nominalisation. Biber et al. (2011) in their corpus-based analysis also found that subordination is more common in conversation, while complex noun phrase constituents and complex phrases are more prevalent in academic writing. Staples, Egbert, Biber and Gray (2016) also found that as academic level and phrasal complexity increase, clausal complexity features, particularly finite dependent clauses, decrease in student writing. Furthermore, Kyle and Crossley (2018) found in a written proficiency corpus made up of 480 argumentative essays that fine-grained indices of phrasal complexity such as the number of dependents per prepositional object were stronger predictors of writing quality than either traditional syntactic complexity indices such as the MLC or fine-grained clausal complexity indices such as the number of objects per clause.

The third major claim is that the clause is a more informative unit of analysis than the T-unit. Lu (2011) pointed out that there are several pieces of evidence in support of the clause as a possibly more informative unit of analysis than the T-unit. For instance, Gaies (1980) and Bardovi-Harlig (1992) challenged the validity and the usefulness of T-unit analysis. Biber et al. (2011) also added that the T-unit can generally capture the extent to which a writer uses dependent clauses. Consequently, the T-unit should not be used to study academic writing since it fails to capture non-clausal features embedded in noun phrases.

The fourth major claim is that L2 performance in task-based research need to be measured by developmentally oriented indices (Lambert & Kormos, 2014; Mazgutova & Kormos, 2015). Mazgutova and Kormos (2015) claimed that complexity in written and spoken performance need to be operationalized in harmony with the linguistic features of the given genre or task-type. Previous studies found that different linguistic features characterise speech and writing (Biber et al., 2011; Halliday & Martin, 1993, 1996). For example, phrasal embedding, complex nominalisation and the use of abstract and compound nouns occur more frequently in academic writing than in speech (Fang, Schleppegrell, & Cox, 2006; Norris & Ortega, 2009). Furthermore, Mazgutova and Kormos (2015) pointed out that the complexity demands of writing and speech differ across genres. For example, expository texts are characterised by a higher number of relative clauses and passive constructions and more complex noun phrases than narrative texts (Nippold, 2004; Berman & Nir-Sagiv, 2007). Way, Joiner, and Seaman (2000) found in French beginner learners' written data that syntactic complexity was the highest for the descriptive task, while the lowest was for the expository task. Beers and Nagy (2009) found that words per clause correlated positively with quality for expository essays, whereas clauses per T-unit correlated positively with quality of narratives in the written data produced by middle school writers. Lu (2011) found that argumentative essays demonstrated higher syntactic complexity than narrative essays. Figure 3.3 illustrates a multidimensional, clause-based, developmentally oriented measurement of L2 syntactic complexity as recommended by Norris and Ortega (2009), Biber et al., (2011), Lu (2011), and Mazgutova and Kormos (2015).

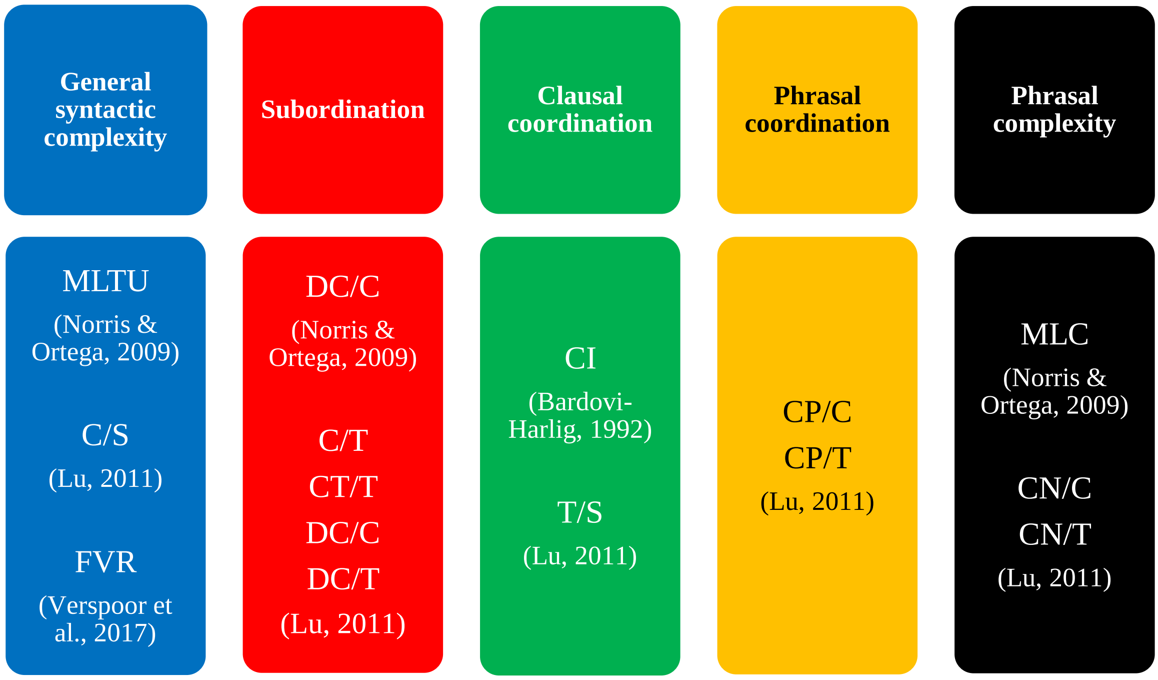


Figure 3.3 Multidimensional L2 syntactic complexity measurement as recommended by Norris and Ortega (2009), Lu (2011), and Verspoor et al. (2017)

3.2.4 Accuracy

Accuracy generally refers to the absence of error in a text. There is a general agreement on the definition of accuracy in the literature. For example, Foster and Skehan (1996) defined accuracy as "freedom from error" (p. 304), while Wolfe-Quintero et al. (1998) defined accuracy as "the ability to be free from errors while using language" (p. 33). However, researchers do not agree on what constitutes an error. For example, Lennon (1991) defined error as "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterpart" (p. 182). Therefore, coding reliability may be especially problematic to achieve.

Polio and Shea (2014) identified three different types of accuracy measures: (1) holistic scores or scales, (2) error-free units, and (3) error counts. Holistic scales are used to assess linguistic or grammatical accuracy as one element among others in a composition rating scale. The different scales consist of descriptors in connection with vocabulary, spelling, punctuation, syntax, morphology, idiom use, paragraph indentation, and word form. Some studies using holistic scales attempt to quantify the number of errors, while other studies try to characterize the quality of language. Polio (1997) pointed out that holistic scales can go beyond the quantification of the number of errors since the severity of the errors can also be considered.

Polio (1997) claimed that error-free T-unit (EFTs) and error-free clause (EFCs) measures are more objective than holistic ratings. In addition, error-free units are more purely a measure of accuracy since they are separate from complexity. In other words, an essay might contain many error-free T-units but can contain very simple sentences. Nevertheless, error-free unit measures (EFTs and EFCs) ignore the severity of the error and the number of errors within one T-unit. In addition, using a measure such as EFT or EFC requires the precise definition of the production unit (for example the T-unit or the clause). The definition of the T-unit is clear cut (an independent clause and its dependent clauses) (Hunt 1965), while the definition of a clause is quite problematic as was previously mentioned. Furthermore, Polio (1997) pointed out that the definition of what counts as error-free is also problematic. She concluded that both EFTs and EFCs are suitable for the quantification of errors, but they are not appropriate for the identification of the quality of errors.

Wigglesworth and Foster (2008) proposed the weighted clause ratio (WCR) to measure linguistic accuracy. The WCR was designed to indicate the linguistic accuracy of written language since it was based on the clause rather than the T-unit. The basis of the WCR is that there is a general agreement that inaccuracies in language units might influence comprehensibility to varying degrees (Kuiken & Vedder, 2007, 2008; Palotti, 2009). The WCR is calculated by dividing the texts into clauses and then assigning a weight to each clause based on the ease of retrieving meaning. The WCR index is calculated by aggregating the values of each weighted clause and dividing it by the total number of clauses (see Foster and Wigglesworth (2016) for a detailed discussion). Wigglesworth and Foster (2008) claimed that the WCR accounts for both the severity and frequency of errors. However, Polio and Shea

(2014) found that the weighted T-unit measure was not more reliable than the error-free unit measures in an investigation into measures of linguistic accuracy in second language writing research. Evans, Hartshorn, Cox, and Martin de Jel (2014) also examined the validity of the WCR by comparing the WCR to the EFT and the EFC indices. Evans et al. (2014) concluded that the WCR might function better in the evaluation of language produced by lower-proficiency learners. In addition, the authors found that the EFC index distinguished learners' varying accuracy levels more reliably than the WCR index.

The third group of linguistic accuracy measures are error counts without classification (Carlisle, 1989; Zhang, 1987). This type of measure obtains the quantity of errors more effectively than a measure such as the EFT because the latter does not differentiate between one and more than one errors per T-unit. Polio and Friedman (2017) emphasised that counting the number of errors can be challenging since sentences can be revised in many different ways that might result in different error counts. Errors are usually tagged according to a manual. One of the most extensively used guidelines is the Louvain Error Tagging Manual (Dagneux, Dennes, & Granger, 1998). The hierarchical Louvain error tagging system includes seven main error domains (see Figure 3.4). Each domain is further broken down into subcategories. The Louvain system distinguishes between 54 error types.

Polio and Shea (2014) also pointed out that neither of the four different types of accuracy measures (see above) is more valid than the other because all of these measures gauge the same construct. Furthermore, the accuracy measures correlate highly with each other (Polio & Friedman, 2017). Polio and Shea (2014) also claimed that EFT/T and EFC/C are "relatively practical, but information is lost because the number and type of errors is not identified" (p. 22). Therefore, researchers should employ more than one measure in their studies. However, Polio and Shea (2014) doubted whether a multidimensional measurement, as proposed by Norris and Ortega (2009) for the measurement of syntactic complexity, can be "extended to accuracy because the various accuracy measures are more redundant and are all related to the number of errors in some way" (p.22). Figure 3.4 illustrates a proposed measurement of L2 accuracy, based on Polio and Shea (2014).

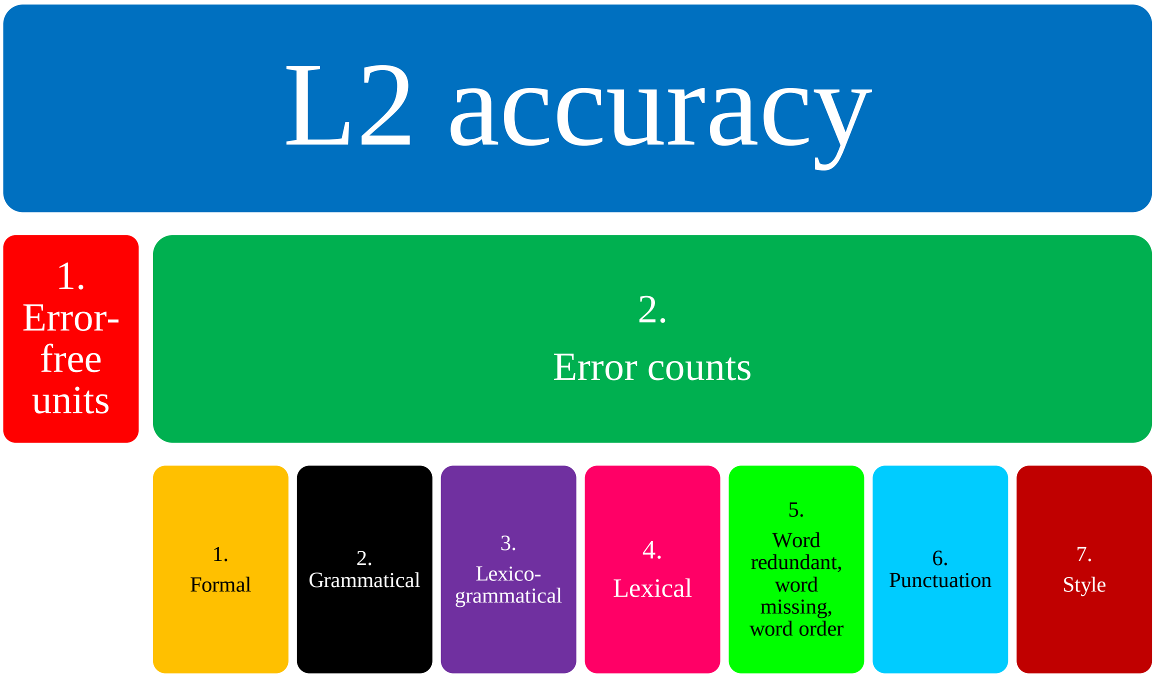


Figure 3.4 A proposed measurement of L2 accuracy based on Polio and Shea (2014)

3.3 Linguistic Development in Two-Wave Longitudinal Studies

Previous studies adopting a two-wave longitudinal research design of L2 writing development generally take samples at two points in time from the same population. The population is generally large which permits generalization about the observed phenomenon. If there is a statistical significance between the two numbers of the same index, development is confirmed. However, longitudinal studies adopting a two-wave research design can confirm only three possible outcomes: (1) an increase, (2) stagnation, or (3) a decrease in a specific linguistic index. Another characteristic of two-wave longitudinal studies is that they usually group participants by proficiency level (Mazgutova & Kormos, 2015; Verspoor, Schmid & Xu, 2012) or class level (Lu, 2011).

3.3.1 Lexical Development in Two-Wave Longitudinal Studies

Average word length (AveWL) is a very general measure of lexical complexity since both less frequent and more complex lexical items with different affixes tend to be longer than more frequent and morphologically simple lexical items (Verspoor & van Dijk, 2011). The AveWL index tends to increase as L2 writers become more proficient. Verspoor et al. (2012) found that the AveWL index increased slowly across five different stages of development (from A1.1 to B1.2 CEFR levels) in the written data of 437 Dutch learners of English. However, Knoch et al. (2015) found that the AveWL index decreased from Time 1 to Time 2 in the written data of 31 undergraduate students enrolled in a three-year degree study in an Australian university. Furthermore, Barkaoui (2016) found that the AveWL index did not change significantly across three test occasions in the written data of 78 IELTS test repeaters. The participants with diverse L1 background completed the IELTS Writing Task 2 three times. However, the length of time between the first and third test ranged between 14 and 219 days.

Lexical variability is generally measured by calculating a type-token ratio-based index such as the TTR or the MTLD index. Lexical variability tends to increase as L2 writers become more proficient. Mazgutova and Kormos (2015) found that the MTLD index increased in the academic writing of two groups from Time 1 to Time 2 over the course of a one-month long intensive English for Academic Purposes (EAP) programme at a British university. The participants were 25 postgraduate upper-intermediate and 14 undergraduate intermediate students of diverse L1 backgrounds (Chinese, Japanese and Thai). Likewise, Barkaoui (2016) found that the MTLD index increased across test occasions. However, Bulté and Housen (2014) found that the lexical diversity index D (Malvern et al., 2004) decreased slightly in 45 adult ESL learners' written data over four months. However, the decrease was not statistically significant.

Lexical rarity is usually measured by the calculation of a frequency-based index. Lexical rarity tends to decrease as L2 writers use less frequent lexical items. For example, Mazgutova and Kormos (2015) found that the log frequency of content words decreased in the writing of their two groups. Likewise, Bulté and Housen (2014) found that the advanced Guiraud (AG) index (Daller, van Hout, & Treffers-Daller, 2003) increased over four months. However, Barkaoui (2016) found that word frequency did not change significantly across test occasions.

Lexical disparity is generally measured by the calculation of the latent semantic analysis (LSA) index (Jarvis, 2013b). The latent semantic analysis index provides measures of semantic overlap between sentences or paragraphs. The latent semantic analysis index tends to increase as L2 writers are able to address a writing prompt more specifically. Mazgutova and Kormos (2015) found that the latent semantic analysis index increased from Time 1 to Time 2 in the group of intermediate students. However, the LSA index stagnated in the group of upper-intermediate students. Barkaoui (2016) found that three of the four latent semantic analysis indices (connectives density, argument overlap for adjacent sentences and mean LSA overlap for adjacent sentences) did not change significantly across test occasions. However, the latent semantic analysis overlap for adjacent paragraphs increased significantly. Crossley, Kyle and McNamara (2016) found that the latent semantic analysis index between all sentences and paragraphs index increased in the written data of 47 university level students enrolled in upper-level EAP courses at an American university over one semester. However, the strongest growth was found in the noun overlap between paragraphs index indicted by the effect size.

Academic Word List (AWL) index, computed by the Vocabprofiler BNC (Cobb, 1994; Heatley & Nation, 1994), is generally used to calculate the percentage of academic words in a specific text. The Academic Word List index tends to increase as L2 writers use more academic vocabulary. Storch and Tapper (2009) found that the Academic Word List index showed statistically significant improvements over 10 weeks in 69 postgraduates' written data.

Likewise, Mazgutova and Kormos (2015) found that the Academic Word List index increased significantly in the written data of both groups in their study. Likewise, Knoch et al. (2015) found that the Academic Word List index increased over three years. However, Storch (2009) found that the Academic Word List index decreased in the written data of 25 graduate students from a university in Australia over 12 weeks. The 25 participants (17 female and eight male) were from Asia (such as China, Indonesia, and Vietnam). Table 3.1 is a summary the results of previous two-wave longitudinal studies on the development of lexical complexity in L2 writing.

Table 3.1

The Development of Lexical Complexity in Two-Wave Longitudinal Studies

Sub-construct	Study	Level	Index	Change
General	Verspoor et al. (2012)	A.1.1 to B1.2	AveWL	increase
	Knoch et al. (2015)	advanced	AveWL	decrease
	Barkaoui (2016)	mixed	AveWL	stagnation
Variability	Bulté & Housen (2014)	advanced	D	decrease
	Mazgutova	intermediate	MTLD	increase*
	& Kormos (2015)	upper-int.	MTLD	increase
	Barkaoui (2016)	mixed	MTLD	increased
Rarity	Bulté & Housen (2014)	advanced	AG	decrease
	Mazgutova	intermediate	WRDFRQc	decrease*
	& Kormos (2015)	upper-int.	WRDFRQc	decrease
	Barkaoui (2016)	mixed	WRDFRQc	stagnation
Disparity	Mazgutova	intermediate	LSA	increase*
	& Kormos (2015)	upper-int.	LSA	stagnation
	Barkaoui (2016)	mixed	LSA	stagnation
	Barkaoui (2016)	mixed	LSA**	increase*
	Crossley et al. (2016)	advanced	LSA	increase
Academic	Storch (2009)	advanced	AWL	decrease
	Storch & Tapper (2009)	advanced	AWL	increase
	Mazgutova	intermediate	AWL	increase*
	& Kormos (2015)	upper-int.	AWL	increase*
	Knoch et al. (2015)	advanced	AWL	increase

Note. *statistically significant; **LSA overlap for adjacent paragraphs.

3.3.2 Syntactic Development in Two-Wave Longitudinal Studies

General syntactic complexity is normally measured by a length-based index (Norris & Ortega, 2009) or by the finite verb ratio (FVR) (Verspoor et al., 2017). Vyatkina (2012) found that the FVR index did not display a monotonic increase for the cross-sectional cohort data of beginning learners of German. However, the correlation analysis showed a moderate positive correlation of the word per finite verb measure with time. Verspoor et al. (2012) found that the words per T-unit (W/T) measure increased across five different stages of development.

Subordination is usually measured by dividing the number of dependent or subordinate clauses by a unit of analysis such as the total number of clauses or T-units. Storch (2009) found that the dependent clause per clause (DC/C) index increased in the written data of 25 graduate students over 12 weeks. Mazgutova and Kormos (2015) found that the dependent clauses per T-unit (DC/T) index decreased in their upper-intermediate group, while the DC/T index increased in their intermediate group. Knoch et al. (2015) found that the DC/C index decreased in the data of 31 undergraduate students over three years. However, the differences did not reach significance. Hou, Verspoor and Loerts (2016) found that four different subordination measures (T-unit complexity ratio, complex T-unit ratio, dependent clause ratio, and dependent clauses per T-unit) decreased significantly in the written data of Chinese learners of English over 18 months.

Clausal coordination is generally measured by dividing the T-units per sentences in a text. Vyatkina (2012) found that the coordinating conjunction (CC) index decreased in the written data of beginner learners of German. Bulté and Housen (2014) found that the coordinate clause ratio (coordinate clauses per sentence) (CCR) index increased significantly over four months. However, Hou et al. (2016) did not find significant changes in the writing of Chinese learners over 18 months. Phrasal coordination is usually calculated by dividing the number of coordinate phrases per a unit of analysis such as the total number of clauses or T-units. Hou et al. (2016) found that both the coordinate phrase per clause (CP/C) and the coordinate phrase per T-unit (CP/T) indices increased significantly in the data of 23 Chinese senior school students and in the data of 18 Chinese university students.

Phrasal complexity is generally measured by calculating the complex nominals per clause or T-unit. Mazgutova and Kormos (2015) found that the complex nominals (CN) and modifiers per noun phrase (SYNNP) indices increased in both groups from Time 1 to Time 2 over a one-month long EAP course. Likewise, Bulté and Housen (2014) found that the mean length of noun phrases (MLNP) index increased significantly over four months. Similarly, Hou et al. (2016) found that the complex nominal per clause (CN/C) and the complex nominal per T-unit (CN/T) increased significantly in the written data of senior school and university students. However, the verb phrases per T-unit (VP/T) and the CN/T indices decreased significantly in the third group which included 18 students with higher language proficiency. Crossley and McNamara (2014) found that the number of modifiers per noun phrase (SYNNP), syntactic similarity (STRUT), and the number of words before the main verb (SYNLE) indices demonstrated statistically significant differences in the written data of 57 university-aged L2 writers over one semester. Furthermore, Crossley and McNamara (2014) found that the incidence of prepositional phrases (DRPP) increased over time. Barkaoui (2016) found that the left embeddedness (SYNLE), syntax similarity (STRUT) and noun phrase density (NPD) indices did not change significantly across test occasions. However, the SYNLE index increased slightly over time. Verspoor et al. (2012) also found that the distribution of the different types of sentences changed across five proficiency levels and that beginner learners used mainly simple and compound sentences. However, the proportion of simple sentences decreased quickly across proficiency levels and the proportion of complex sentences started to increase between A.1.2 and A2 CEFR levels. According to Verspoor et al. (2012), there could be a trade-off relationship between simple and complex sentences between A.1.2 and A2 CEFR levels. The authors also found that the proportion of compound sentences remained stable across the five levels. See Table 3.2 for a summary.

Table 3.2

The Development of Syntactic Complexity in Two-Wave Longitudinal Studies

	Study	Level	Index	Change
General	Vyatkina (2012)	beginner	FVR	increase
	Verspoor et al. (2012)	A1.1 to B1.2	W/T	increase
Subordination	Storch (2009)	advanced	DC/C	increase
	Mazguotva	intermediate	DC/T	increase
	& Kormos (2015)	upper-int.	DC/T	decrease
	Knoch et al. (2015)	advanced	DC/C	decrease
	Hou et al. (2016)	mixed	DC/C	decrease*
			DC/T	decrease*
Clausal	Vyatkina (2012)	beginner	CC	decrease
coordination	Bulté & Housen (2014)	advanced	CCR	increase*
	Hou et al. (2016)	mixed	T/S	stagnation
Phrasal	Hou et al. (2016)	mixed	CP/C	increase*
coordination			CP/T	increase*
Phrasal	Bulté & Housen (2014)	advanced	MLNP	increase*
complexity	Crossley	advanced	SYNNP	increase*
	& McNamara (2014)		STRUT	decrease*
			SYNLE	increase*
			DRPP	increase
	Mazgutova	intermediate	CN	increase*
	& Kormos (2015)	intermediate	SYNNP	increase*
		intermediate	STRUT	increase*
		upper-int.	CN	increase
		upper-int.	SYNNP	increase
		upper-int.	STRUT	increase*
	Barkaoui (2016)	mixed	SYNLE	increase
		mixed	STRUT	stagnation
		mixed	NPD	stagnation

Note. *statistically significant.

3.3.3 The Development of Accuracy in Two-Wave Longitudinal Studies

Accuracy is generally measured by the calculation of an error-free unit index. Storch (2009) found that grammatical accuracy did not change significantly in the written data of 25 university students over 12 weeks. The ratio of errors per words (E/W) increased, while the ratio of error-free clause per clause (EFC/C) decreased over time. However, the error-free T-units stagnated over 12 weeks. Likewise, Storch and Tapper (2009) found that the ratio of error-free T-units per total T-units (EFT/T) and the error-free clauses per total clauses (EFC/C) indices did not change significantly over 10 weeks. However, the difference between time 1 and time 2 on the ratio of total number of errors per total number of words (E/W) index was statistically significant. Knoch et al. (2015) found that there were slight changes across time and the differences were not significant in both EFT/T and EFC/C indices. Verspoor et al. (2012) calculated the distribution of types of errors at five levels of proficiency and found that both spelling and lexical errors decreased significantly across levels. However, spelling and lexical errors remained relatively high compared with other types of errors. Table 3.3 is a summary the results of previous two-wave longitudinal studies on the development of accuracy in L2 writing.

Table 3.3

The Development of Accuracy in Two-Wave Longitudinal Studies

Study	Level	Index	Change
Storch (2009)	advanced	EFT/T	stagnation
	advanced	EFC/C	decrease
	advanced	E/W	increase*
Storch & Tapper (2009)	advanced	EFT/T	stagnation
	advanced	EFC/C	stagnation
	advanced	E/W	increase
Knoch et al. (2015)	advanced	EFC/C	stagnation
	advanced	EFT/T	stagnation

Note. *statistically significant.

3.4 Linguistic Development in Multi-Wave Longitudinal Studies

Studies adopting a multi-wave longitudinal research design generally take samples at more than two points in time. However, there is no agreement on the exact frequency of data collection waves. Previous multi-wave longitudinal studies differ in the length of the study and in the frequency of the data collection waves (Ortega & Iberri-Shea, 2005). In addition to confirming an increase, decrease or stagnation in the development of a linguistic index under observation, multi-wave longitudinal studies can also provide an insight into the dynamic interactions between two or more constructs. From a CDST perspective, a multi-wave research design is generally preferred for studies on language development.

3.4.1 Lexical Development in Multi-Wave Longitudinal Studies

Verspoor et al. (2004) were not able "to discern any clear levelling off of variation indicating some development" in the average word length (AveWL) index in the written data of two beginner Dutch learners of English over a six-week period. Verspoor et al. (2008) found an overall increase of the average word length index in the academic writing (18 samples) of an advanced Dutch university learner of English during her three years of study. Similarly, Ma (2012) found an overall increase of the AveWL index in the academic written data of two advanced Dutch learners of English over four years. In addition, Verspoor et al. (2017) found an overall increase of the AveWL index in the writing (22 texts) of three advanced Dutch learners of English over three years in a university setting. Likewise, Penris and Verspoor (2017) found an overall increase of the AveWL index in the written data (49 samples) of a Dutch learner of English over 13 years. However, the growth trajectory of the AveWL index displayed a great deal of variability over time in the abovementioned studies.

Verspoor et al. (2004) were not able to discern any patterns in the development of the type-token ratio (TTR) index in the written data of two Dutch beginner learners of English. Verspoor et al. (2008) found a general increase in the type-token ratio index in the written data of a Dutch learner over three years. Furthermore, Larsen-Freeman (2006) found that the type-token ratio index increased in the written data of five Chinese learners of English over a six-month period. However, the individual performances showed regression and progression over time. In addition, Rosmawati (2016) found that the measure of textual lexical diversity (MTLD) index increased in the academic written data of a Chinese learner of English over one academic year. However, the measure of textual lexical diversity (MTLD) index developed without discernible patterns in the written data of the Thai, Japanese, and Korean learners of English. Penris and Verspoor (2017) also found a general increase of the type-token ratio (TTR) index in the written data of a Dutch learner over 13 years. Similarly, Yang and Sun (2015) found that the Guiraud's index displayed a U-shaped curve in the group averages of five third-year undergraduate Chinese students' writing. The group averages indicated a decrease from Time 1 to Time 5. However, a great deal of variability was found in the growth trajectories of the five individual learners. Verspoor et al. (2018) found a general upward trend in the Guiraud index in the written data of 22 Dutch learners over 23 weeks. However, the variability analyses showed that peaks occurred at different points in time in the 12-year-old leaners' written data.

Penris and Verspoor (2017) found a general decrease in the word frequency index over 13 years. Zheng (2016) found that lexical sophistication, calculated by the Beyond-2000 scores (B2000), increased in the academic written data of 15 university-level Chinese students of English over 10 months. However, the individual data of three participants, selected from the 15 students based on the scores in lexical sophistication, diversity and density, showed a great deal of variability in the lexical sophistication index.

Verspoor et al. (2008) found an overall increase in the Academic Word List (AWL) index in the written data of a Dutch learner of English. Likewise, Penris and Verspoor (2017) found a general increase in the Academic Word List index in the written data of a Dutch learner of English over 13 years. However, in both studies the growth trajectories displayed a great deal of variability over time. Table 3.4 is a summary of the results of previous multi-wave longitudinal studies on the development of lexical complexity.

Table 3.4

The Development of Lexical Complexity in Multi-Wave Longitudinal Studies

Sub-construct	Study	Level	Index	Change
General	Verspoor et al. (2004)	beginner	AveWL	
	Verspoor et al. (2008)	advanced	AveWL	upward
	Ma (2012)	advanced	AveWL	upward
	Verspoor et al. (2017)	advanced	AveWL	upward
	Penris & Verspoor (2017)	advanced	AveWL	upward
Variability	Verspoor et al. (2004)	beginner	TTR	
	Verspoor et al. (2008)	advanced	TTR	upward
	Yang & Sun (2015)	intermediate	Guiraud	downward
	Rosmawati (2016)	advanced	MTLD	upward
		advanced	MTLD	
	Penris & Verspoor (2017)	advanced	TTR	upward
	Verspoor et al. (2018)		Guiraud	upward
Rarity	Zheng (2016)	advanced	B2000	upward
	Penris & Verspoor (2017)	advanced	WRDFRQc	downward
Academic	Verspoor et al. (2008)	advanced	AWL	upward
	Penris & Verspoor (2017)	advanced	AWL	upward

3.4.2 Syntactic Development in Multi-Wave Longitudinal Studies

Verspoor et al. (2008) found an overall increase in the FVR index in the written data of a Dutch learner of English over three years. Vyatkina (2012) also took a deeper look into the written texts of two learners from her cohort data and she found mixed results. The FVR index displayed an incremental increase in the longitudinal data of a female beginner learner of German, while the FVR index decreased in the data of the male learner. Verspoor et al. (2017) found a general improvement in the FVR index in the data of three Dutch learners of English. Likewise, Penris and Verspoor (2017) found a general increase of the FVR index in the data of an advanced learner of English. However, the growth trajectories of the FVR indices showed variability over time in the abovementioned studies.

Rosmawati (2016) found mixed results on the development of subordination in the written data of her four learners. She found an overall decrease of the subordination (DC/T) index in the written data of two learners of English (Thai and Chinese), and an overall increase of the DC/T index in the written data of the other two learners of English (Japanese and Korean).

The growth trajectory of the DC/T index showed six inverse U-shaped peaks in the Japanese learner's data. Rosmawati (2016) found a more complex pattern in the DC/T trajectory of the Korean learner. The DC/T trajectory was stable in the first couple of months, but then the DC/T index displayed an inverse U-shaped pattern. In the last couple of months, the DC/T trajectory flattened out.

Rosmawati (2016) found an overall stagnation in the clausal coordination (T/S) index in the academic written data of four participants over one academic year. However, the growth trajectories of the T/S index showed different degrees of variability for the four participants. Rosmawati (2016) found a general stagnation of the phrasal coordination (CP/C) index in the academic written data of the Thai, Japanese and Chinese learners of English, while the CP/C index showed an overall increase in the written data of the Korean student. Conversely, Chan et al. (2015) found an overall decrease of the CP/T index in the written data of two Taiwanese beginner learners of English over eight months. In both studies the phrasal coordination indices displayed different degrees of variability over time.

Rosmawati (2016) found a general increase in the complex nominalisation indices in the written data of a Korean learner of English over one academic year. Likewise, Penris and Verspoor (2017) found an overall increase of the average noun phrase length (ANPL) in the written data of a Dutch learner of English over 13 years. In both studies the trajectories showed different degrees of variability over time. Table 3.5 is a summary of the results of previous multi-wave longitudinal studies on the development of syntactic complexity.

3.4.3 The Development of Accuracy in Multi-Wave Longitudinal Studies

Larsen-Freeman (2006) found that the proportion of error-free T-units to T-units (EFT/T) index increased in the data of five Chinese learners of English over a six-month period. Caspi (2010) distinguished between lexical and syntactic accuracy and she found that both lexical accuracy, measured by the correct lexical use (CLU) index and syntactic accuracy, measured by the correct word order (CWO) index, increased in the written data of a Portuguese learner of English over 36 weeks. Rosmawati (2016) found an overall increase of the EFC/C index in the academic written data of four learners of English over one academic year. Table 3.6 is a summary of the results of previous multi-wave longitudinal studies on the development of accuracy.

Table 3.5

The Development of Syntactic Complexity in Multi-Wave Longitudinal Studies

Sub-construct	Study	Level	Index	Change
General	Verspoor et al. (2008)	advanced	FVR	upward
	Vyatkina (2012)	beginner	FVR	upward
	Vyatkina (2012)	beginner	FVR	downward
	Verspoor et al. (2017)	advanced	FVR	upward
	Penris & Verspoor (2017)	advanced	FVR	upward
Subordination	Rosmawati (2016)	advanced	DC/T	upward
		advanced	DC/T	downward
Clausal	Vyatkina (2012)	beginner	CC	downward
coordination	Rosmawati (2016)	advanced	T/S	stagnation
Phrasal	Chan et al. (2015)	beginner	CP/T	downward
coordination	Rosmawati (2016)	advanced	CP/C	stagnation
		advanced	CP/C	upward
Phrasal	Rosmawati (2016)	advanced	CN	upward
complexity	Penris & Verspoor (2017)	advanced	ANPL	upward

Table 3.6

The Development of Accuracy in Multi-Wave Longitudinal Studies

Study	Level	Index	Change
Larsen-Freeman (2006)	intermediate	EFT/T	upward
Rosmawati (2016)	advanced	EFC/C	upward
Caspi (2010)	advanced	CLU	upward
	advanced	CWO	upward

3.5 Interactions Between Complexity and Accuracy

There are two competing hypotheses on the interactions of the CAF constructs: (1) Limited Attention Capacity (LAC) hypothesis and the (2) Cognition Hypothesis (CH). According to the LAC hypothesis (Skehan, 1996; 1998), L2 learners may not be able to focus on all aspects of performance (CAF) as task complexity increases. Conversely, the CH (Robinson, 2001; 2011) suggests that a more complex task may result in greater performance in complexity, accuracy and fluency. In this study, the LAC and the CH are applied to investigate division of attention within tasks and not across task.

The results of the interactions between complexity and accuracy are mixed. Previous studies adopting a cross-sectional or two-wave research design shed light on the overall interactions between constructs. For example, Kormos (2011) drew on both the Limited attention capacity and the Cognition hypotheses in her study to examine the effects of task complexity on two groups of learners' written text. The L2 participants were secondary school Hungarian learners of English, while the L1 students were British undergraduates. Kormos (2011) manipulated the complexity of the writing task and also compared L1 and L2 writers. However, she did not find a significant difference in the two types of narratives and only slight differences between L1 and L2 students, with no interaction effects.

A CDST approach to the development of complexity and accuracy can shed light on two different types of interactions: (1) surface and (2) underlying interactions. The former refers to the overall correlation, while the latter refers to the temporal correlations between two constructs. For example, Verspoor et al. (2008) found competition between average sentence length (AveSL) and type-token ratio (TTR) indices in the written data of one Dutch advanced learner of English over three years. However, the moving window of correlation plot showed that the interactions between average sentence length and type-token ratio indices were dynamic. Spoelman and Verspoor (2010) also found that word complexity and sentence complexity, and word complexity and noun phrase complexity develop simultaneously (support), while noun phrase complexity and sentence complexity develop asynchronously (competition). Caspi (2010) found a weak competition between lexical complexity (word variation) and lexical accuracy (correct lexical use), a moderate competition between lexical accuracy and syntactic complexity (DC/C), and a support between syntactic complexity and syntactic accuracy (correct word order) in the written data of a Portuguese learner of English over 36 weeks. Vyatkina (2012) found a supportive interaction between lexical variety (CTTR) and sentence length (SL) in the cohort data of beginner learners of German. Conversely, Vyatkina (2012) found a competitive interaction between coordinating conjunctions (CC) and subordinating conjunctions (SC) in the cohort data. Verspoor et al. (2017) found competition between the average word length (AveWL) and the finite verb ratio (FVR) indices in one of the three participants' written data. However, the competition between the average word length and finite verb ratio indices changed to support in the second half of the investigation. Interestingly, Verspoor et al. (2017) found inverse patterns in the written data of the other two participants. See Table 3.7 for a summary.

Table 3.7

Interactions Between Complexity and Accuracy

Study	Construct	Construct	Interaction
Verspoor et al. (2008)	ASL	TTR	competition
Spoelman	word complexity	sentence complexity	support
& Verspoor (2010)	word complexity	noun phrase complexity	support
	noun phrase complexity	sentence complexity	competition
Caspi (2010)	word variation	correct lexical use	competition
	correct lexical use	DC/C	competition
	DC/C	correct word order	support
Vyatkina (2012)	CTTR	SL	support
	CC	SC	competition
Verspoor et al.	AveWL	FVR	c. → s.
(2017)	AveWL	FVR	s. → c.

Note. c. = competition; s. = support; → = direction of change.

3.6 L2 Writing Motivation

Second language writing development is not an isolated process. There are several cognitive, motivational, and social factors that might shape the process of L2 writing. Ortega (2012) claimed that "writing is a mysteriously complex, irremediably social, deeply visceral activity" (p. x).

Boscolo and Hidi (2007) distinguished between three areas of motivational constructs that play an important role in L2 writing: (1) motives (goal orientation, needs, values, interest), (2) writers' perceptions of their ability to write regarding the difficulty of the task and the resources of the context (self-efficacy, self-concept, and self-perceptions), and (3) productive strategies (self-regulation). Furthermore, Boscolo and Hidi (2007) pointed out that "these areas are rarely, if ever, separate from one another" (p. 2). Kormos (2012) claimed that both cognitive factors such as language aptitude, working memory and phonological short memory, and motivational factors, such as language learning goals, self-efficacy beliefs, and self-regulatory processes might explain variations in L2 writing processes.

3.6.1 Self-Regulation in L2 Writing

Self-regulation has been demonstrated to play an important role in second language writing development. Zimmerman and Kitsantas (2007) pointed out that second language writing development depends on the successful acquisition of lexicon and syntax and "high levels of self-regulation and self-motivation" (p. 51). Boscolo and Hidi (2007) pointed out that "a writer has to coordinate cognitive, metacognitive, and linguistic processes when producing extended texts" (p. 8). Kormos (2012) defined self-regulation as "a process in which people organize and manage their learning" (p. 395). Self-regulation involves the language learners' control over their thoughts (e.g. competency beliefs), emotions (e.g. anxiety experienced during the learning process), behaviours (e.g. how they deal with a specific learning task) and the learning environment (Pintrich & De Groot, 1990; Zimmerman, 1998). In other words, self-regulation determines whether language learners will engage in writing activities, what type of writing they will undertake, with what level of effort and attention they will approach the different phases of the writing process and how they exploit the learning potential of writing tasks.

Previous theories of writing have identified several self-regulatory processes such as self-efficacy, self-observation, causal attribution, self-evaluation, text analysis, and goal-setting. Zimmerman (2000) proposed a cyclical social cognitive theory consisting of three self-regulatory phases: (1) forethought, (2) performance control, and (3) self-reflection. Zimmermman's (2000) model of self-regulation includes sources of motivation and cognitive processes "in a cyclical feedback loop wherein writing outcomes are used to modify and guide subsequent efforts to write" (Zimmerman & Kitsantas, 2007, p. 54).

According to Zimmermman's (2000) model of self-regulation, the forethought phase involves (1) task analysis processes (goal setting and strategic planning), and (2) sources of self-motivation (self-efficacy, outcome expectancy, task interest and goal orientation). The performance phase includes (1) self-control methods (self-instruction, mental imagery, task strategies, attention focusing, time management, environmental structuring, self-consequences, help-seeking), and (2) self-observation (metacognitive monitoring and self-recording). The self-reflection phase involves (1) self-judgement (self-evaluation and causal attribution) and (2) self-reaction (self-satisfaction, and adaptive/defensive). Nitta and Baba (2015) claimed that goal setting, self-observation and self-evaluation are the "three key self-regulatory processes" (p. 376). Figure 3.5 illustrates the three key self-regulatory processes proposed by Nitta and Baba (2015) in Zimmermann's (2000) cyclical model of self-regulation.

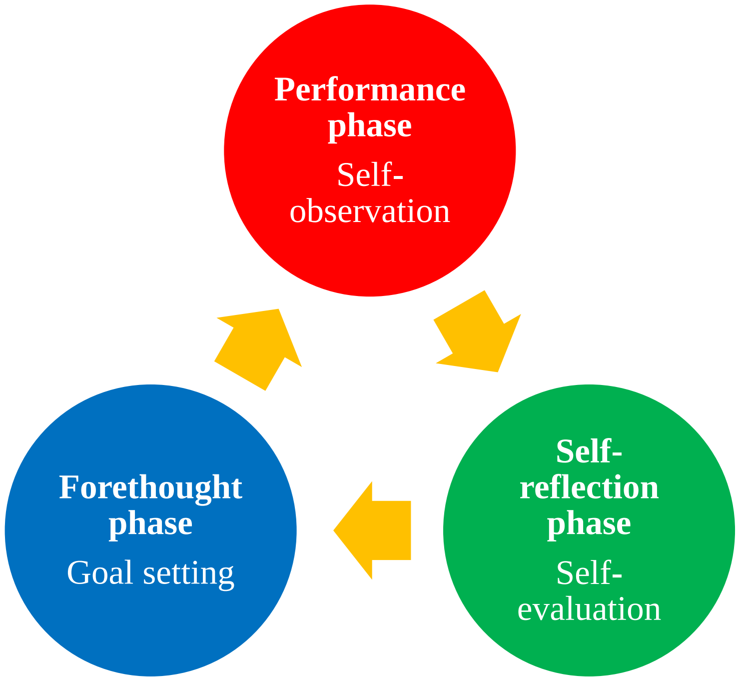


Figure 3.5 The three key self-regulatory processes proposed by Nitta and Baba (2015) based on Zimmerman's (2000) model of self-regulation

Zimmerman and Kitsantas (2007) pointed out that proficient writers "analyse their literary tasks and set specific writing goals" (p. 54) in the forethought phase. Self-observation refers to "metacognitive monitoring and physical record-keeping" (Zimmerman & Kitsantas, 2007, p. 57) of a given aspect of the writer's performance in the performance phase, whereas self-evaluation refers to "comparing self-monitored outcomes with a standard goal" (Zimmerman & Kitsantas, 2007, p. 58).

3.6.2 Motivational Factors in L2 Writing

In addition to self-regulation, interest, boredom, and anxiety also play an important role in second language writing development. Boscolo, Del Favero and Borghetto (2007) pointed out that the "request to write can elicit emotions such as anxiety, boredom, interest, etc." (p. 79). However, these emotions might be highly dependent on the writer's general attitude to writing. According to Boscolo et al. (2007) there are "different patterns of interactions between general motivation to write, text-based interest, and topic interest before and after writing" (p. 79).

It is reasonable to assume that writers more voluntarily engage in a writing activity if they find the writing prompt interesting. Boscolo and Hidi (2007) defined interest as "an individual's affective response to specific features of the environment" (p. 5). Research on interest in writing differentiates between individual and situational interest. The former is defined "as a stable aspect of the individual that can manifest across contexts", while the latter is "conceived as describing the interaction of the individual with an interest object in a particular context" (Bobbitt Nolen, 2007, p. 242). Previous studies on individual interest in writing found that topic attractiveness plays a role in writing quality (Hidi & McLaren, 1990, 1991; Benton, Corkill, Sharp, Downey, & Khramstova, 1995). However, these two studies viewed interest as a static motivational variable. In other words, students were either interested or uninterested in a specific writing topic. Previous studies on situational interest found that "interest would emerge in social activity viewed as meaningful by the students themselves" (Boscolo & Hidi, 2007, p.6).

Boredom is one of the most commonly experienced and potentially most destructive affective states in the classroom (Pekrun, Goetz, Hall, & Perry, 2014). Although there is a dearth of studies on the relationship between boredom and L2 writing development, it can be presumed that boring topics might influence the writing outcome negatively. Hidi and McLaren (1990, 1991) presumed that a source of situational interest (the interestingness of themes and topics), which were found to affect children's comprehension, might also affect children's production of expository texts. In their study, fourth- and fifth-graders and teachers rated the degree of interest of different topics and themes. The same students were required to compose a text on either a high-interest or low-interest topic. Hidi and McLaren (1990, 1991) did not find any positive effects of topic interest on the quality of written compositions. Furthermore, the authors found that students who composed a text on a low-interest topic produced longer texts than those students who composed a text on a high-interest topic.

3.6.3 A Macro Perspective on Self-Regulation

Previous studies on the role of self-regulation in L2 writing were mainly cross-sectional (Csizér & Tankó, 2017; Kormos & Csizér, 2014; Mehrabi, Kalantarian, & Boshrabadi, 2016; Nami, Enayati, & Ashouri, 2012; Samanian & Roohani, 2018; Teng & Zhang, 2016; Yang & Plakans, 2012; Zimmerman & Bandura, 1994). Consequently, self-regulation and other motivational factors were perceived as stable and monolithic. For example, Zimmerman and Bandura (1994) examined the predictiveness of the social cognitive theory of self-regulation for academic achievement. The authors found that different facets of perceived self-efficacy played an important role in writing course attainment. Self-efficacy raised the goals students set for themselves and the quality of writing with which they would be self-satisfied. Personal self-efficacy beliefs affected the level of writing attainments.

Similarly, Kormos and Csizér's (2014) study considered motivation, self-regulatory strategies and autonomous learning behaviour as stable constructs. The authors investigated the influence of motivation and self-regulation on the autonomous learning behaviour of 638 Hungarian EFL students. Kormos and Csizér (2014) found that strong instrumental goals, and international posture, together with positive future self-guides, are essential for the use of efficient self-regulatory strategies.

Likewise, Csizér and Tankó's (2017) study was a cross-sectional investigation of English majors' self-regulatory control strategy use and the relationship of control strategy use, motivational dispositions, and anxiety/self-efficacy beliefs. The authors collected data on standardized paper and pencil questionnaire from 222 first-year English majors at a large

Hungarian university. The authors found that only a third of the participants possessed the ability and willingness to control their writing processes.

3.6.4 A Micro-Perspective on Self-Regulation

Research on self-regulation in L2 writing was also influenced by the adoption of Complex Dynamic Systems Theory to L2 motivation research. Dörnyei (2009b, 2009c) pointed out that “most learner characteristics are complex, higher order mental attributes, resulting from the integrated operation of several subcomponents and subprocesses” (p.233). In other words, higher order individual difference (ID) variables (such as aptitude and motivation) might involve the cooperation of many different sub-components (such as cognitive, motivational or emotional factors). Dörnyei and Tseng’s (2009) study on motivational task processing well illustrates the componential mixture of higher order ID variables. Dörnyei and Tseng’s (2009) construct suggests that L2 learners are engaged in a cyclical appraisal and response process, including their constant monitoring and evaluating how well they are doing in a specific task and then making possible changes if something is going amiss. Dörnyei (2009b) pointed out that the motivational task processing (see Figure 3.6) can be represented through a dynamic system that includes three interconnected mechanisms: (1) task execution, (2) appraisal, and (3) action control.

According to Dörnyei and Tseng (2009), task execution refers to the learners’ engagement in task-based learning in accordance with the action plan that has been provided either by the teacher or the task designer. Task appraisal refers to the learner’s constant processing of the multitude of stimuli arriving from the environment in connection with their progress towards achieving the action outcome, and their comparing the actual performance with the predicted or likely performance. Action control refers to self-regulatory mechanisms that are activated to enhance, scaffold, or protect learning specific action.

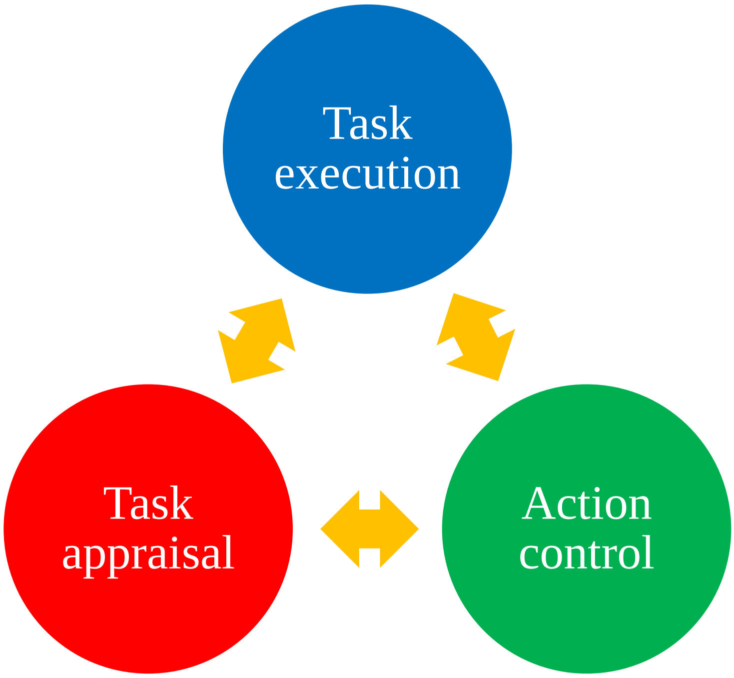


Figure 3.6 Motivational task processing based on Dörnyei & Tseng's (2009) study

Dörnyei and Tseng's (2009) construct is quite similar to Zimmerman's (2000) cyclical model of self-regulation. Furthermore, the constant functioning of self-regulatory processes is indispensable to maintain one's motivation in daily learning environments over longer periods of learning (Dörnyei & Ottó, 1998). Self-regulatory processes are believed to take place during a specific task. However, Nitta and Baba (2015) claimed that self-regulatory processes can also be "extended to longer periods of classroom learning" (p. 370).

Recent research on L2 motivation has also been fundamentally influenced by a powerful model called the L2 Motivational Self System (Dörnyei, 2005, 2009a). This model includes (1) the ideal L2 self, (2) the ought-to L2 self, and (3) L2 learning experience and provides a reinterpretation and expansion of the previous main separation between integrative and instrumental motivation (Gardner & Lambert, 1972). In Dörnyei's model integrativeness is incorporated in the construct of ideal L2 self, which is the L2 learner's ideal self-image expressing the wish to become a competent L2 user. The ought-to L2 self includes "attributes that one believes one ought to possess (i.e. various duties, obligations, or responsibilities) in order to avoid possible negative outcomes" (Dörnyei, 2005, p. 106). L2 learning experience contains “situation specific motives related to the immediate learning environment and experience” (Dörnyei, 2005, p. 106). Dörnyei’s L2 Motivational Self System model is based on the Self-discrepancy Theory (Higgins, 1987). According to the Self-discrepancy theory, motivation is the result of a person’s wish to reduce the discrepancy between the ideal self and the actual self. Motivation also originates from the intention to reduce the gap between the person’s actual self and the ought-to self, that is, the perception of what significant others would like her/him to become. From a macro-perspective the L2 Motivational Self System is seen as dynamic “cognition-emotion-motivation amalgam” (Dörnyei, 2009c, p. 251).

Although there is a strong suggestion that self-regulation and L2 writing motivation are not stable and monolithic constructs, only a few studies incorporated a time factor in their studies (Forbes, 2018; Han & Hiver, 2018; Nicolás-Conesa, Roca de Larios, & Coyle, 2014; Nitta & Baba, 2015, 2018; Sasaki et al., 2018; Serafini, 2017). For example, Nicolás-Conesa et al. (2014) explored EFL learners’ mental models and their effects on writing performance. The authors found that there are complex relationships between learners’ task representation, their formation of sub-goals, and engagement in writing tasks. Nicolás-Conesa et al.’s (2014) study adds empirical evidence for the relationship between a dynamic process of task conceptualisation and the maintenance of motivation and self-regulation to write. The results showed that EFL learners’ goals for writing from a motivational and self-regulatory perspective are influenced by the interplay of ID and contextual factors.

Nitta and Baba (2018) investigated the benefits of repeating a writing task over 30 weeks in an EFL classroom-based study from a CDST approach. Nitta and Baba (2018) investigated how 26 Japanese students’ use of self-regulatory processes related to changes in L2 writing. The participants of Nitta and Baba’s (2018) study were engaged in a 10-minute timed-writing task on a selected topic with immediate self-reflection on a weekly basis. Students’ essays were analysed using measures of lexical and syntactic complexity and fluency, while the students’ self-reflections written in L1 were analysed in terms of self-regulatory processes. The findings of two focal students showed that the first student demonstrated more elaborate engagement and employed self-regulatory cycles of goal-setting and self-evaluation, which improved his L2 writing over the 30 weeks. Conversely, the second student showed more limited engagement and employed less elaborate self-regulatory processes. Consequently, the second student reflected little change in his L2 writing. Nitta and Baba (2018) concluded that repeated encounters with tasks over longer periods of time might create a valuable pedagogic environment and within this context students' agentic attitudes towards the L2 writing task probably influence their learning significantly.

Han and Hiver (2018) examined processes of motivational change for 174 middle school language learners within classroom ecology of genre-based writing. The authors investigated changes in the students' writing-specific motivational profiles, supported by a time-series analysis of reflective journals and interviews with the participants. The findings showed that the students with respective profiles were able to develop a stronger capacity for writing self-regulation and sustain and consolidate their writing self-efficacy. Furthermore, Han and Hiver (2018) found that the final profile of several students was characterised by elevated levels of writing anxiety. Nonetheless, the elevated levels of anxiety was combined with moderate-to-strong levels of writing self-regulation and writing self-efficacy. Han and Hiver (2018) concluded that anxiety can co-exist in constructive configurations such as together with sufficient levels of self-regulation and self-efficacy that can offset this.

In another study, Sasaki et al. (2018) investigated how the development of three L2 writing strategies (global planning, local planning, and L1-to-L2 translation) used by 37 Japanese university students interacted with other cognitive and environmental factors over 3.5 years. The authors found that systematic patterns and individual uniqueness co-occurred in development. Study abroad experiences and other career-related environmental factors contributed to self-regulated learning over time.

3.6.5 A Micro-Perspective on L2 Writing Motivation

Research on motivational dynamics in L2 learning generally explains changes in the motivational system by the following three CDST features: (1) attractor states, (2) perturbations and (3) co-adaptation. For example, Waninge, de Bot and Dörnyei (2014) investigated the motivational dynamics of four language learners during their language lessons over two weeks. The authors investigated whether there is variability in students' in-class motivation and whether there is a detectable stable level or attractor states in students' in-class motivation. The results of their study showed that student motivation can be successfully explored using a Complex Dynamic Systems Theory framework. The authors also demonstrated how motivation changes over time on an individual level, while at the same time being characterised by predictable and stable phases. Furthermore, Waninge et al. (2014) showed that motivation is inseparable from the learner's individual learning context and motivation can be meaningfully studied at different interacting time scales. Waninge et al. (2014) concluded that classroom motivation is a mixture of dynamic stability, governed by attractor states, and individual variability, caused by a combination of multiple issues.

Waninge (2015) focused on identifying the main attractor states that make up the learning experiences in a language classroom. The 56 participants, enrolled in a language development module, were second year students of English at a British university. The interviews with the participants were kept short (from seven to 19 minutes) intentionally because it was expected that a dominant attractor state would surface in the first couple of minutes of an interview. From the content analysis of the focused qualitative interviews, four distinct categories emerged: interest, boredom, neutral attention, and anxiety. Waninge (2015) claimed that interest, the most frequently mentioned state, is one of the most important attractor states in the L2 learning experience since it is closely connected to “learners’ active engagement and future motivational intention in the classroom” (p. 201). Boredom, a passive and negative affective dimension, was also frequently mentioned by the participants. Boredom has been found to be negatively connected to academic achievement (Baker, D’Mello, Mercedes, Rodrigo, & Graesser, 2010; Dewaele, 2011; Pekrun et al., 2010). Neutral attention was also reported by the participants, although less frequently than interest and boredom. Anxiety was the fourth salient attractor state mentioned by the participants. Anxiety was found to be more detrimental for learning than boredom (Waninge, 2015). It was also found that anxiety appeared to wax and wane during lessons. Although some participants reported switching from one state to another within a single lesson, there were consistent data that pointed to relatively stable states.

In another study on motivational dynamics, Henry (2015) examined processes of change in six participants’ motivational systems. The participants came from a class of 22 first-year students studying French at the A2.2 CEFR level. A series of semi-structured interviews were carried out five times with three participants and six times with the other three participants over an eight-month period. The interviews were analysed using the abstractions of Complex Dynamic Systems (CDS) and possible selves theories as a compass. In the interviews, indications of changes in attractor states, perturbations, phase shifts, co-adaptation between the L3 (French) and L2 (English) motivational systems and features of emergence were sought. It was found that the motivational system displayed considerable fluctuations and variations over time and the momentary shifts in students’ motivation could be caused by a wide range of factors. These factors included perceptions concerning the enjoyment, meaningfulness, relevance of a particular activity, the effect of perturbations, such as receiving test results, and the continuous interaction with other complex dynamic systems. Henry (2015) found that perturbations triggered a process in which two identity-transforming phase shifts occurred in one student's L3 motivational system.

3.7 Co-Adaptation of Second Language Writing and Motivation

Co-adaptation refers to "change in a system that is motivated by change in another, connected, system" (Larsen-Freeman & Cameron, 2008, p. 66). A change in a learner's motivational system might trigger a change in his or her linguistic system. In dynamic parlance, a learner's motivational system may co-adapt with his or her linguistic system. Although the literature suggests that L2 writing development and self-regulation are closely connected, previous studies focused mainly on the two constructs separately.

Csizér, Kormos and Sarkadi (2010) proposed a dynamic model of language learning motivation, the Model of the Nested Systems in Motivation (MNSM). The authors conducted qualitative interviews with 15 Hungarian dyslexic students who studied foreign languages in diverse educational settings. Although Csizér et al. (2010) applied the MNSM to dyslexic learners, they claimed that their model "can also be applied to other learner populations because it includes internal and external factors that are relevant for other groups of learners" (p. 483).

Csizér et al. (2010) claimed that students' goals, attitudes, and motivated behaviour form three interrelated systems that are set in both the learner's milieu and the general social context for language learning (Figure 3.7). In addition, cognitive factors and self-perceptions are also added to the learner internal system. Language learning goals affect the attitudes students have towards language learning. The authors argued that goals, attitudes, and motivated behaviour are interrelated co-adaptive systems, in which change in one of these motivational systems might bring about change in the other closely related system. Among these three subsystems, language learning goals were found to be the most stable, while attitude and motivated behaviour tended to fluctuate during language learning. Dörnyei (2010) also claimed that cognition and motivation are dynamically interconnected. The language learner is situated in his or her close social environment (family and friends), which interacts with the language learner's attitudes and goals. Societal values and expectations concerning language learning are mediated by the student's milieu. Instructional setting also influences a learner's internal factors.

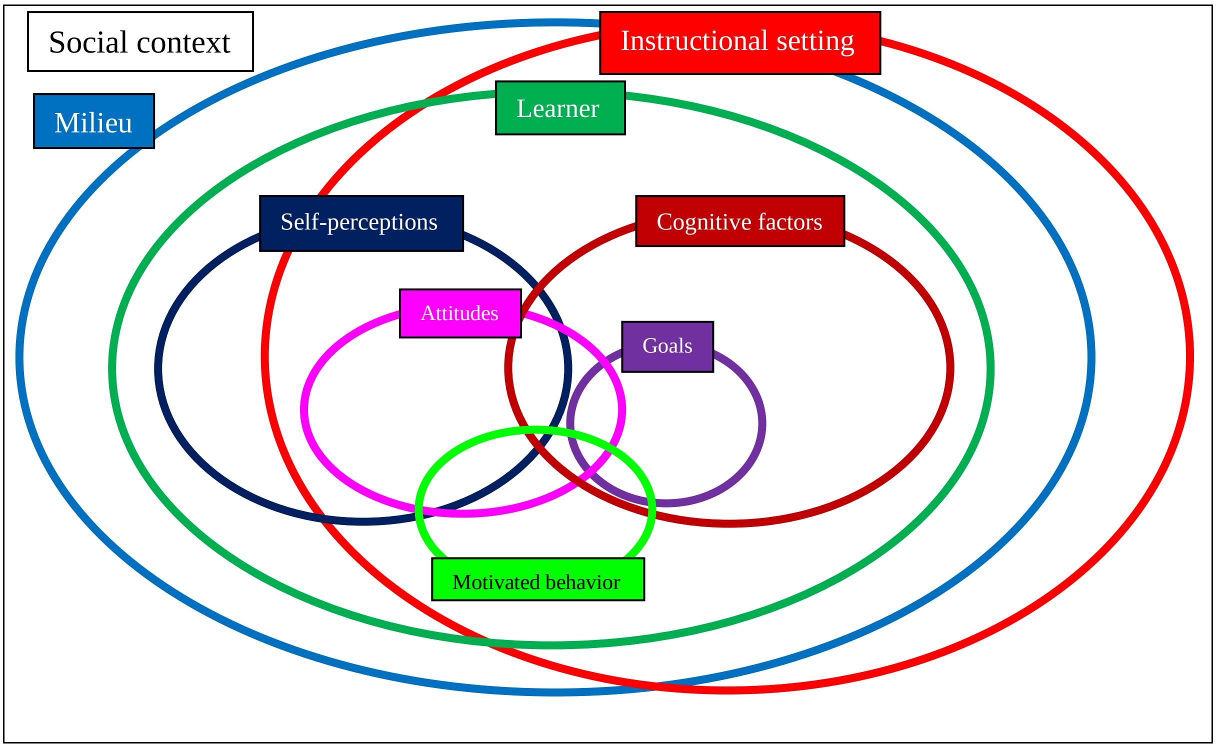


Figure 3.7 Model of the Nested Systems in Motivation (Csizér, Kormos & Sarkadi, 2010)

In this study, the MNSM was revised and extended to explain changes in the participants' linguistic and motivational systems. First, the systems of self-perceptions and cognitive factors were removed from the MNSM since the participants of this study were not dyslexic language learners. Furthermore, this study did not focus on the learners' self-perceptions (ideal L2 self) and cognitive factors. Instead, the system of self-regulatory processes was added to the MNSM. Self-regulation is generally considered as a motivational factor (Kormos, 2012) or a psychological factor (Han & Hiver, 2018). Previous studies on educational psychology found that motivation to attain a specific goal is presumed to trigger self-regulated learning behaviour (e.g. Heckhausen & Dweck, 1998; Lens & Vansteenkiste, 2008; Sansone & Smith, 2000; Wigfield, Hoa, & Klauda, 2008). Ryan and Deci (2000) also argued that identification with learning goals and integration of learning goals are prerequisites for self-regulated actions. According to the social cognitive theory of academic self-regulation, learners regulate motivational, affective, and cognitive aspects (Zimmerman, 1986, 1990a, 1990b). Ushioda (1996, 2003, 2006) also argued that students who take responsibility for their own learning are generally more intrinsically motivated and are capable of regulating their learning process more effectively. In addition, self-regulation changes over time as it interacts with cognitive, affective and social factors (Sasaki et al., 2018). Nitta and Baba (2015) also adopted self-regulation as a construct for characterising L2 learning experience in the L2 Motivational Self System and they found that the ideal L2 self evolves co-adaptively with the development of micro-level functioning of self-regulation. Consequently, since there is an overlap between motivation and self-regulation, the construct of self-regulatory processes is also placed in the motivational system in Figure 3.8.

Second, the MNSM was extended with the linguistic system since the focus of this study is linguistic development. Beckner, Blythe, Bybee, Christiansen, Croft, N. C. Ellis, Holland, Ke, Larsen-Freeman and Schoenemann (2009) claimed that language is a complex adaptive system. The linguistic system, comprised of the systems of complexity and accuracy, is placed in the language learner's internal system. Furthermore, the system of complexity includes two sub-systems: lexicon and syntax. De Bot and Larsen-Freeman (2011) claimed that language is as a dynamic system including highly interconnected subsystems such as the lexical system and the syntactical systems. In Figure 3.8, the lexical and the syntactical systems overlap one another since the lexical and the syntactical are intertwined rather than separate (Bybee, 1998; N. C. Ellis, 2008; Goldberg, 2006; Halliday, 1994; Langacker, 1987).

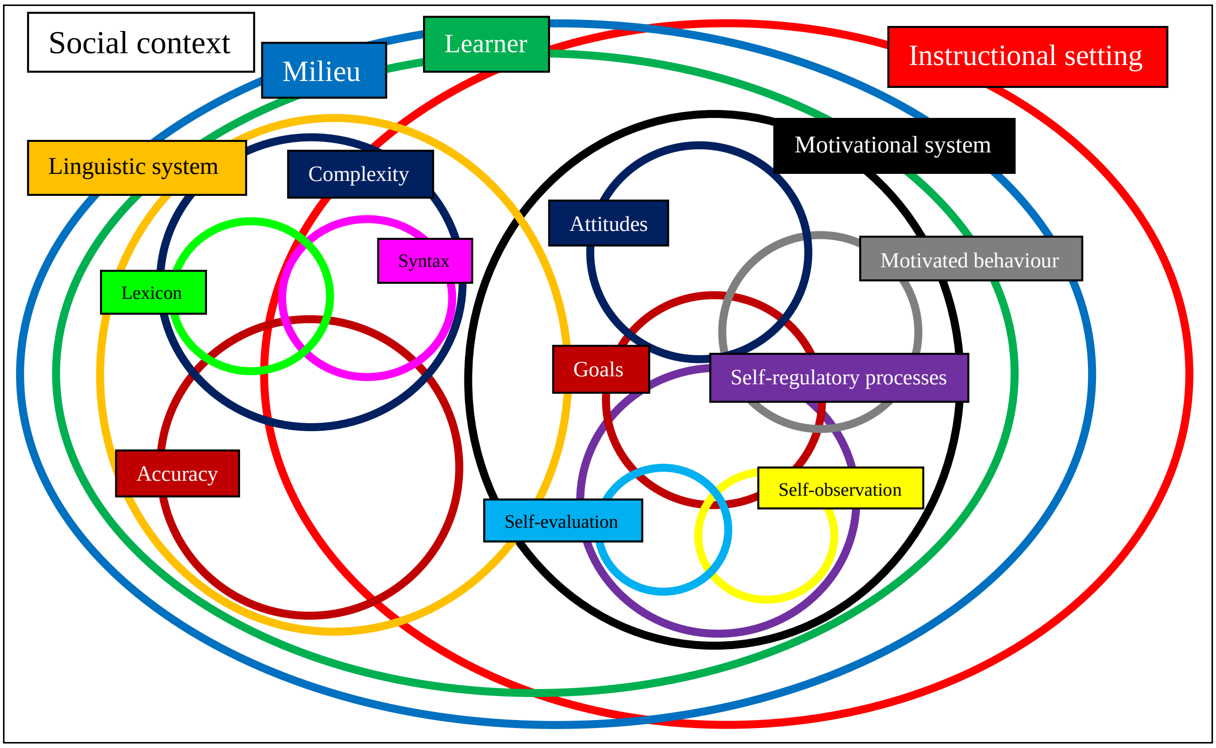


Figure 3.8 Model of the linguistic and motivational systems

Previous research on the interaction of linguistic development and motivation focused on oral tasks or adopted a cross-sectional research design. For example, Kormos and Dörnyei (2004) investigated the role of motivation and some other closely related individual difference variables (anxiety, and willingness to communicate) in oral task performance in a cross-sectional study of 44 intermediate Hungarian secondary school students. The authors found that motivation influenced the quantity of talk more than the quality of the content produced. They concluded that motivation functioned as a driving force that made learners actively engage in a specific task. However, motivation played a limited role determining the quality of the outcome.

Recent research on the relationship between linguistic development and motivation adopted a CDST approach. For example, Zheng (2012) examined the dynamic interplay between four Chinese EFL students' motivation and vocabulary development over a 10-month period. Zheng (2012) found that productive vocabulary development was mediated by multiple factors and processes within the pedagogical and sociocultural context. The students' motivation concerning their ought-to self and their ideal L2 self guided their utilization of available learning resources and mainly mediated their productive vocabulary development.

Likewise, Nitta and Baba (2015) investigated how the ideal L2 self develops and evolves through repeated engagements in language learning tasks over an academic year. Self-regulation was adopted to represent the third component (L2 learning experience) in the L2 Motivational Self System. The analysis of the changes in the participants' self-regulatory processes and L2 writing revealed that the ideal L2 self evolved co-adaptively with the development of micro-level functioning of self-regulation. Furthermore, it was found that the self-evaluation process played a more important role than the self-observation processes in facilitating the development of second language writing. Nitta and Baba (2015) argued that co-adaptation or co-evolution might occur between the ideal L2 self and self-regulation processes in a sense that the "ideal L2 self functions as an evaluative standard for self-regulation, while self-regulatory processes influence the formation and revision of the ideal L2 self" (p. 372).

3.8 Research Gap

Research on L2 writing development to date has several theoretical and methodological implications. Previous research suggested that L2 writing development might be best captured by a more holistic and organic approach such as the Complex Dynamic Systems Theory (Norris & Ortega, 2009). Since L2 writing development is both a cognitive and situated activity (Polio & Friedman, 2017), theories that reduce their focus to either one or the other aspect of L2 writing (such as Goal Theory, Rhetorical Genre Theory) may not be able to gain insight into the entire developmental process.

Earlier studies suggested that both lexical and syntactic complexity are multifaceted constructs (Jarvis, 2013b; Norris & Ortega, 2009). Therefore, a multidimensional measurement is indispensable to capture L2 writing development. Furthermore, there was a strong indication that academic writing is characterised by phrasal complexity instead of clausal elaboration (Biber et al., 2011; Kyle & Crossley, 2018) and that the clause is a more informative unit of analysis than the T-unit (Lu, 2011). In addition, complexity, accuracy and fluency are claimed to be measured by developmentally oriented indices (Lambert & Kormos, 2014).

Although previous longitudinal studies adopting a two-wave research design might provide useful information about the general trends in L2 writing, they limit themselves to a three-way outcome: increase, stagnation or decrease (Barkaoui, 2016; Bulté & Housen, 2014; Knoch et al., 2014; Knoch et al., 2015; Mazgutova & Kormos, 2015; Storch, 2009; Storch & Tapper, 2009). It appears that a multi-wave longitudinal research design, usually adopting the CDST framework, may provide a better insight into the dynamic interplay of the linguistic constructs (Caspi 2010; Larsen-Freeman, 2006; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2004, 2008). Although previous longitudinal studies adopting a multi-wave research design have provided useful information about the dynamic nature of language development, most of these studies limited their focus only on the linguistic aspect of development not taking into consideration the socio-cultural context (Caspi, 2010; Rosmawati, 2016).

Past studies on the role of self-regulation in L2 writing development were mainly cross-sectional (Csizér & Tankó, 2017; Kormos & Csizér, 2014; Mehrabi et al., 2016; Nami et al., 2012; Samanian & Roohani, 2018; Teng & Zhang, 2016; Yang & Plakans, 2012; Zimmerman & Bandura, 1994). However, there is a strong recommendation that motivational factors are not stable and monolithic (Csizér et al., 2010; Dörnyei, 2009b, 2009c; Han & Hiver, 2018; Henry, 2015; Nicolás-Conesa et al., 2014; Nitta & Baba, 2015, 2018; Sasaki et al., 2018; Serafini, 2017; Waninge et al., 2014; Waninge, 2015).

Several previous studies indicated that there is an urgent need to incorporate other closely related systems such as the motivational construct into the investigation of L2 writing development. Although there have been a limited number of studies investigating the relationship, or in dynamic parlance co-adaption, between the linguistic and motivational system, these studies also limited their focus to only one aspect of linguistic development (Nitta & Baba, 2015; Zheng, 2012).

This study fills the research gap by investigating the development of lexical and syntactic complexity and accuracy in the written data of four Hungarian EFL learners over a nine-month period. Furthermore, the evolution of self-regulatory processes will be traced in order to explore how the linguistic and motivational systems co-adapt over time.

3.9 Summary

In Chapter 3, I defined the main constructs of this study: complexity and accuracy. I reviewed the main research on second language writing development and the interactions between complexity and accuracy. I defined the construct of self-regulation and reviewed the main research on self-regulation. Finally, I presented a model of co-adaptation.

Chapter 4. Methodology

Chapter Overview

Chapter 4 presents the research design, data sources, and methodological procedures for the study. There are seven main sections: (1) methodology from a CDST perspective, (2) pilot study, (3) the four cases, (4) EAP programme and context, (5) data collection, (6) data coding, and (7) data exploration.

4.1 Methodology From a Complex Dynamic Systems Theory perspective

As reviewed in Chapter 3, there have been numerous recommendations that future studies on L2 writing development should capture the multi-dimensionality, dynamicity, variability and nonlinearity of the constructs of complexity and accuracy (e.g. Norris & Ortega, 2009). As demonstrated by previous longitudinal studies on L2 writing development (e.g. Casanave, 1994; Ishikawa, 1995; Knoch et al., 2015; Mazgutova & Kormos, 2015), a two-wave research design - collecting data at only two points in time - might not be able to capture the dynamicity of L2 writing development (Ortega & Iberri-Shea, 2005). Therefore, this study adopted a more holistic, multi-wave approach, drawing on Complex Dynamic Systems Theory, which is believed to better capture the dynamic and nonlinear nature of L2 writing development than traditional longitudinal studies adopting a two-wave research design.

Second language writing development has been generally investigated by longitudinal or panel studies adopting a two-wave research design (e.g. Barkaoui, 2016; Knoch et al., 2014, 2015; Mazgutova & Kormos, 2015; Storch, 2009). Development was claimed if changes were detected by inferential statistics in the measures of complexity and accuracy between two points in time. However, some studies used rating scales to assess writing development and these studies typically included a broader set of criteria (e.g. Assiri, 2015; Plakans, 2013). Although longitudinal studies adopting a two-wave research design can reveal general tendencies in L2 writing development by sampling a large population of L2 writers, the dynamicity, variability, and nonlinearity of L2 writing cannot be fully captured. Two-wave studies cannot capture the fluctuations and the nonlinear patters of development over time. As suggested by several recent studies on L2 development (Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Verspoor et al., 2008), CDST provides such a framework which can effectively capture the nonlinear development of complexity and accuracy.

Adopting CDST as a framework provides a set of theoretical and methodological underpinnings (Hiver & Al-Hoorie, 2016). Studies adopting a CDST framework typically involve the following seven design features: (1) they are longitudinal, (2) they are single or multiple case studies, (3) they make use of time-series or multi-wave research designs, (4) they draw on naturalistic data, (5) they include the research environment, (6) they include minimal control, and (7) they utilise mixed methods research designs. Figure 4.1 illustrates a typical CDST research design.

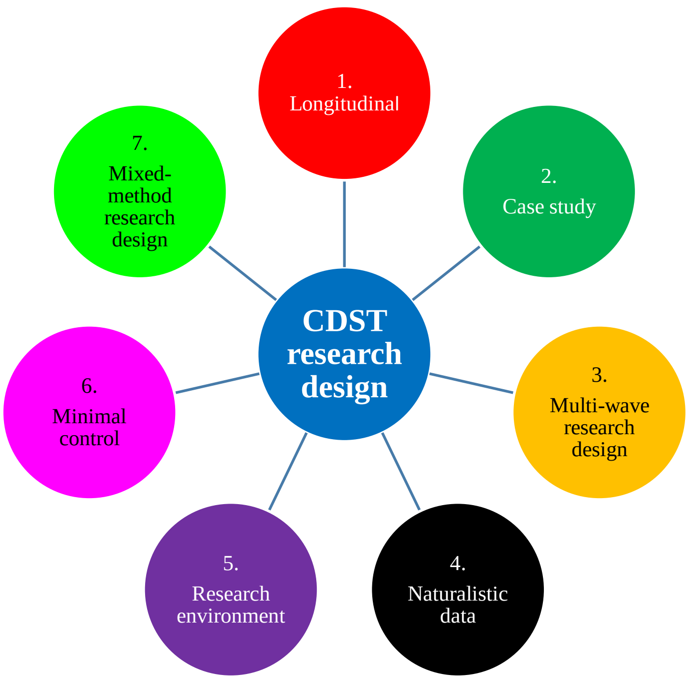


Figure 4.1 CDST research design

Although some of the abovementioned components of the CDST research design are inherently connected with each other, they are presented separately in the following sub-sections below here.

4.1.1 Longitudinal Research

Change over time is a fundamental element of a study adopting the CDST framework. Consequently, CDST studies are generally designed as longitudinal observations on individual cases. The longitudinal observation makes it possible to trace changes in a dynamic system over time. Mellow, Reeder, and Foster (1996) also claimed that language development can be most fruitfully investigated with longitudinal research methods such as time-series design. However, there is no agreement on the appropriate length of a longitudinal research design (Ortega & Iberri-Shea, 2005). Therefore, CDST studies vary in the length of observation. For instance, Verspoor et al. (2004) collected data for six weeks, Verspoor et al. (2008), and Larsen-Freeman (2006) for six months, Rosmawati (2016) for one academic year, while Penris & Verspoor (2017) for 13 years.

4.1.2 Case Study

Longitudinal data collection is closely related to the second major component of CDST research design: the case study approach. It is generally very difficult to collect longitudinal data with a sample size comparable to traditional two-wave panel studies because the level of prolonged engagement required for a high-quality longitudinal study is often not feasible due to resource constraints. This is particularly the case when longitudinal studies are conducted by sole researchers. Therefore, CDST studies generally adopt a case study approach focusing only on the language development of a small number of participants. Yin (1984) defined the case study approach “as an empirical inquiry that investigates a contemporary phenomenon within its real-life context when the boundaries between phenomenon and context are not clearly evident and in which multiple sources of evidence are used" (p. 23). Stake (2005) distinguished three types of case study: (1) the intrinsic case study, (2) the instrumental case study, and (3) the multiple or collective case study. The intrinsic case study is generally conducted to understand the intriguing nature of a particular case. The instrumental case study is undertaken to gain insight into a broader phenomenon while the actual case is of secondary interest. In multiple case studies there is even less interest in one particular case. One of the disadvantages of case studies is that findings cannot be generalised to a larger population. Therefore, Gaddis (2002) recommended "particular generalizations" and not universal ones (p. 62). Likewise, Larsen-Freeman and Cameron (2008) pointed out that CDST researchers "might acknowledge tendencies or patterns, but resist claiming applicability" for the findings beyond specific times and places (p. 235).

Dörnyei (2007) pointed out that a multiple case study is in fact an instrumental case study extended to a number of cases. According to Yin (2003), multiple case studies are better than single-case studies because they can provide compelling evidence of a phenomenon. However, Duff (2008) pointed out that this putative advantage of multiple-case studies "must be weighed against the time and other resources needed to include additional cases and the resulting trade-offs in depth of analysis between study of one and a study of two or more" (p. 113). Norris and Manchón (2012) also pointed out that multiple case studies "explore in depth the variety of factors and outcomes associated with L2 writing development" (p. 231). Previous CDST studies have focused on one single case (Spoelman & Verspoor, 2010; Penris & Verspoor, 2017) or on multiple cases, including four participants (Caspi, 2010; Chan, 2015; Larsen-Freeman, 2006; Rosmawati, 2016).

4.1.3 Multi-Wave Data Collection

The longitudinal component of the CDST research design makes it possible to collect data at more than two points in time. Consequently, studies adopting the CDST framework tend to collect multi-wave data over time. Ortega and Iberri-Shea (2005) pointed out that longitudinal studies of L2 development are characterised by “multiwave data collection from the same individuals over a relatively long period of time, which typically spans anywhere between four months and four years” (p. 29).

In order to perform CDST-related analyses, such as correlations, the minimum number of observation points is dictated by statistical constraints. When designing a study to estimate a Pearson, Kendall or Spearman correlation, the sample size required to obtain a Fisher confidence interval with the desired width is a major concern. Bonett and Wright (2000) claimed that an accurate sample size approximation can be obtained using a sample of five. CDST studies generally collect data at a minimum of eight times over a half-year period (Verspoor et al., 2004; Verspoor et al, 2008, Zheng, 2016).

Furthermore, in order to detect developmental peaks in learners’ data by applying data resampling and Monte Carlo analysis, the minimum number of observation points is again dictated by statistical constraints. Smith and Wells (2006) found that a sample size of 30 was enough to perform Monte Carlo analysis. Previous CDST studies performed data resampling and Monte Carlo analysis on a small sample size. For example, Piniel and Csizér (2015) performed Monte Carlo analysis on a multi-wave data of six measurement points collected from 21 participants, while Zheng (2016) applied the Monte Carlo analysis on a multi-wave data of eight measurement points collected from 15 participants.

4.1.4 Interaction with the Environment

According to CDST, change is caused by interaction with the environment in a complex dynamic system: “context is not separate from the system but part of it and of its complexity” (Larsen-Freeman & Cameron, 2008). Consequently, CDST studies usually exert minimal control on data collection in order to preserve authenticity and naturalness of the data. For instance, in Verspoor et al.’s (2004) study, the writing task that participants completed was “left completely open” and the participants could write about anything they wanted (p. 9). Likewise, Caspi (2010) collected “freely written essays” on both expository and narrative topics. Furthermore, Penris and Verspoor (2017) collected 49 written samples on many different topics over 13 years.

4.1.5 Mixed Methods Research Design

A mixed methods study involves the combination of quantitative and qualitative research methods (Dörnyei, 2007). However, there is a dearth of studies applying mixed methods research designs in the field of applied linguistics (Duff, 2008). Dörnyei (2007) claimed that “combining methods can open up fruitful new avenues for research in the social sciences” (p. 163). Onwuegbuzie and Leech (2005) were even more critical by claiming that “monomethod research is the biggest threat to the advancement of the social sciences” (p. 375). Nevertheless, the vast majority of previous CDST studies on second language development have applied quantitative research methods (e.g. Caspi, 2010; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2004, 2008) and only a few studies have adopted the mixed methods research design (e.g. Nitta & Baba, 2015). However, several previous studies indicated that a mixed methods research design might provide a deeper insight into the behaviour of dynamic systems. For example, Rosmawati (2016) recommended that interview data might “enrich the data and open the ground for further analysis and discussions” (p. 202).

In sum, there have been several recommendations that the dynamic nature of L2 writing development can best be captured by a longitudinal case study research design. It has also been indicated that data should be collected at more than two points in time, and at more than five points in time if statistical analysis is to be implemented. Furthermore, a mixed methods research design can reveal valuable information on the dynamicity of L2 writing development by triangulating quantitative and qualitative data.

4.2 Pilot Study

Prior to the main study, it was deemed necessary to conduct a pilot study[2] to test the reliability of data collection instruments and to perform a preliminary data exploration. Two Hungarian advanced EFL participants, aged 18 and 19, were recruited from a private language institution where they were attending a C1 CEFR level language course in English. Four written samples were collected over four months, each a response to an IELTS-style writing prompt. The participants received the writing prompts by email, and they had approximately one week to compose the written sample (See Appendix 1). The four IELTS-style writing prompts were taken from the course book, *IELTS Testbuilder* (McCarter & Ash, 2003). After the completion of every second written sample, the participants were interviewed in a classroom of the language institution.

The written data were coded for lexical and syntactic complexity by the computational tools Coh-Metrix 3.0 (Graesser, McNamara, & Kulikovich, 2011; Graesser, McNamara, Louwerse, & Cai, 2004) and L2 Syntactic Complexity Analyzer (L2SCA) (Ai & Lu, 2013; Lu, 2010, 2011). The data analysis included plotting the complexity indices, and the calculation of z-scores. Table 4.1 illustrates the lexical and syntactic complexity indices used in the pilot study.

[2] This pilot study was published in Wind (2014).

Table 4.1

Summary of the Measures Used in the Pilot Study

Construct	Sub-construct	Index
Lexical complexity	Variability	Measure of textual lexical diversity (MTLD)
	Sophistication	Log frequency of content words (WRDFRQc)
Syntactic complexity	Length of production unit	Mean length of T-unit (MLTU)
	Sentence complexity	Sentence complexity ratio (C/S)
	Subordination	Dependent clause per T-unit (DC/T)
	Coordination	Coordinate phrases per T-unit (CP/T)
	Particular structures	Complex nominal per T-unit (CN/T)

The main findings of the pilot study are reported in Wind (2014). The pilot study demonstrated that both lexical and syntactic indices showed inter-individual and intra-individual variability. The WRDFRQc index showed a gradual decline which suggested that both participants started to use less frequent lexical items in their writing over the four-month period. The participants' writing also indicated that lexical complexity increased faster than syntactic complexity when the MTLD was plotted against MLTU and MTLD against DC/T. The improvement of lexical complexity was also confirmed by the interview data. Furthermore, the largest rate change occurred for coordinate phrases per T-unit (CP/T) for both participants.

There were several methodological implications of the pilot study. First, more data points are necessary to exploit the methods and techniques provided by current CDST practices (Van Geert & van Dijk, 2002; Verspoor et al., 2011). Second, lexical complexity should be measured multi-dimensionally by following, for example, Jarvis's (2013b) model. In the pilot study only lexical variability and lexical rarity were measured. The main study should measure other sub-constructs of lexical complexity such as lexical disparity. Furthermore, the main study should measure to what extent writers use genre-specific vocabulary. Third, the T-unit based syntactic complexity indices should be exchanged for clause-based indices. For example, the dependent clause per T-unit (DC/T) index should be exchanged for the dependent clause per clause (DC/C) index, as suggested by Lu (2011) and outlined in Chapter 3. Fourth, the semi-structured interviews should not only elicit information concerning the actions taken by the participants during writing. Instead, it was decided that the interviews of the main study should explore the development of "individual differences" (Manchón, 2009b, p. 245) such as language learning goals, self-efficacy beliefs, and self-regulatory capacities as indicated by Kormos (2012).

4.3 Research Questions

Adopting a mixed methods case study approach, the main study explored four Hungarian EFL learners' linguistic development over a nine-month investigation. In addition, the evolution of the four participants' self-regulatory processes were traced. The following research questions informed the main study.

Research question 1.	How do lexical and syntactic complexity and accuracy indices change in the four Hungarian EFL learners' argumentative essays[3] over the nine-month investigation?
Research question 2.	Where do developmental peaks emerge in lexical and syntactic complexity and accuracy indices in the four Hungarian EFL learners' argumentative essays over the nine-month investigation?
Research question 3.	How do lexical and syntactic complexity and accuracy indices interact in the four Hungarian EFL learners' argumentative essays over the nine-month investigation?
Research question 4.	How do the four Hungarian EFL learners' self-regulatory processes change over the nine-month investigation?

[3] In this study the four participants composed IELTS-type argumentative essays which were required for the academic module of the IELTS exam.

4.4 The Four Cases

Duff (2008) pointed out that case selection and sampling are one of the most critical considerations in case study research. Dörnyei (2007) distinguished three interrelated sampling strategies: (1) homogeneous sampling, (2) typical sampling, and (3) criterion sampling. This study applied a combination of the homogeneous and criterion sampling strategies. First, the four EFL learners participated in the same EAP programme (homogeneity). Second, the participants had to be at CEFR B2 level (criterion sampling).

The participants in this study were four Hungarian EFL learners aged between 19 and 22. The participants were all studying at postgraduate level at a university in Budapest. The four participants were all enrolled in an EAP programme offered by their university. The programme lasted one academic year and the data were collected from September 2014 to May 2015. Initially, data were collected from six participants. However, during the data coding stage it was found that the analysis of the four cases allowed for a very in-depth and detailed exploration of development within confines of the thesis. Including six cases would have meant insufficient space to present the cases to the required level of depth. Therefore, those two participants were excluded from the data analysis who did not provide the entire 23-sample written data set. A summary of details regarding the four cases who will be presented in this thesis is provided in Table 4.2.

Table 4.2

The Four Cases

	Participant	Gender	Age	L1	Proficiency level (CEFR)
1.	Dalma	Female	22	Hungarian	B2
2.	Emese	Female	19	Hungarian	C1
3.	Levente	Male	22	Hungarian	B2
4.	Avarka	Female	21	Hungarian	B2

The minimum requirement to participate in the EAP programme was to have a CEFR B2 level language exam certificate. Therefore, it could be presumed that the level of language proficiency of the four participants was around B2 level. The recruitment of the participants was also dictated by the genre of the writing tasks of this study. It can be presumed that students at a B2 or C1 CEFR level might have had some experience in writing argumentative essays. However, on the negative side, advanced level students might have reached their developmental plateau which might result in the stabilization of the constructs of complexity and accuracy.

4.5 EAP Programme and Context

The four participants of this study attended the EAP programme offered by their university from September 2014 to May 2015. The EAP class consisted of 10 students including the 4 participants of this study. The classes were held every Friday afternoon and were taught by a Hungarian teacher of English with 10 years of experience in language education. The students were required to pay for the EAP programme. However, the participants were entitled to a partial refund, if they obtained an overall score of 6.5 on the IELTS language test.

During the 180-minute long classes the participants completed IELTS-type language tests from the *Cambridge Practice Tests for IELTS* (Jakeman & McDowell, 1997, 2000, 2002, 2005, 2006, 2007, 2009) on a weekly basis. In addition, the participants completed vocabulary tasks from *Check your English vocabulary for IELTS* (Wyatt, 2001) during the classes. The students completed one entire sample IELTS test each week. The listening, the reading and the speaking modules were completed on location. However, the writing modules were completed as home assignments. Furthermore, the students were required to take four vocabulary tests over the nine months which were based on the words from the *Check your English vocabulary for IELTS* course book. The first vocabulary test was completed in January 2015, while the remaining three tests in the following three months.

4.6 Data Collection

Both quantitative and qualitative data were collected from the four participants of this study over the nine-month period. Each participant composed 23 argumentative essays, and each was interviewed seven times throughout the duration of the study on their self-regulatory processes. Consequently, there was an intrinsic imbalance between the quantitative and qualitative data. Therefore, this study applied a quantitatively dominant mixed methods research design. Johnson, Onwuegbuzie and Turner (2007) defined quantitatively dominant research methods as a "type of mixed research in which one relies on a quantitative, postpositivist view of the research process, while concurrently recognizing that the addition of qualitative data and approaches are likely to benefit most research projects" (p. 124). By collecting both quantitative and qualitative data it was also possible to triangulate the two types of data. Dörnyei (2007) pointed out that triangulation is "an effective strategy to ensure research validity" (p. 165). Figure 4.2 illustrates the mixed methods research design of this study.

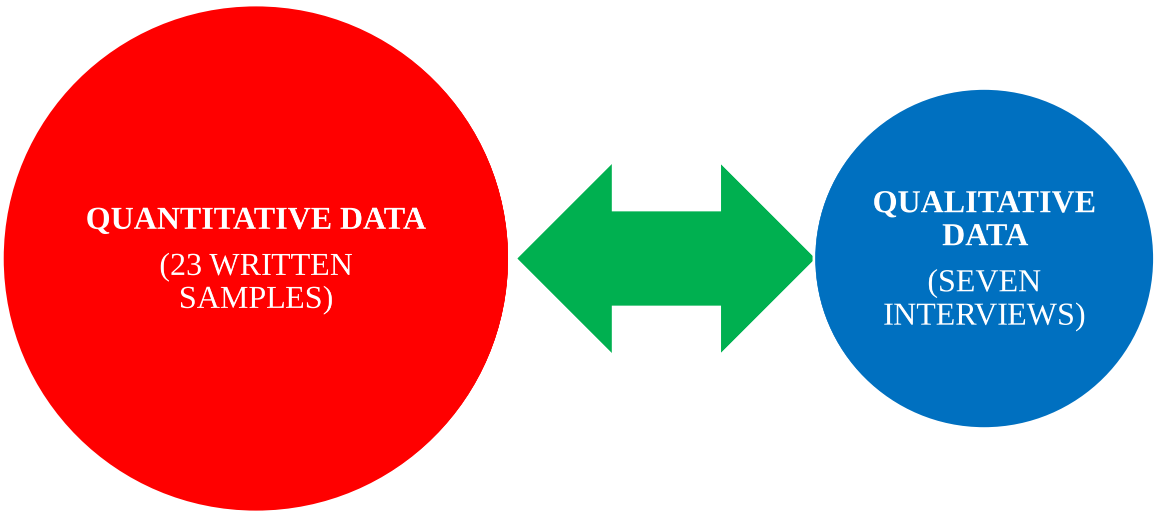


Figure 4.2 Mixed methods research design and triangulation

4.6.1 Written Samples

Each participant composed a total of 23 argumentative essays over a nine-month period. 16 argumentative essays were written as home assignments, while seven argumentative essays were composed during the EAP classes. In other words, 16 essays were written in naturalistic settings, in line with the CDST framework, while seven essays were composed under more controlled environments. A total of 92 written samples were collected for this study and they made up a corpus of approximately 25,000 words.

The writing prompts for the 16 argumentative essays (see Appendix 2) were chosen by the teacher of the EAP class from the primary course book, *Cambridge Practice Tests for IELTS*. The order of the completion of the 16 argumentative essays was harmonised with the order of the topics of the supplementary course book, *Check your English vocabulary for IELTS* (Wyatt, 2001). Therefore, the students of the EAP class could practice the newly acquired topic-specific vocabulary while completing the writing modules at home. The four participants had access to unlimited resources while composing the 16 argumentative essays. The use of dictionaries, auto-correction tools, the internet, electronic editor tools was allowed. The participants were required to send their home assignments to their teacher electronically. Consequently, the 16 argumentative essays were composed by the participants in their natural settings in line with the research designs of previous CDST studies (Caspi, 2010; Rosmawati, 2016; Verspoor et al., 2004). I acknowledge that a large array of variables might shape the written output of the second language learners such as task type, topic, task design, condition, goal, mode, target audience and genre. Nonetheless, according to CDST, these variables are the “soft-assembled” (Thelen & Smith, 1994) elements of L2 academic writing performance and are intrinsic to any language development. The participants received feedback on the 16 written samples in the form of an overall IELTS score on writing by the teacher of the EAP class. Nevertheless, the influence of feedback on the participants’ writing development was not considered in this study. In addition, the participants were free to report on the effectiveness of any particular feedback during the regular interviews throughout the study.

In order to balance between the naturalistic and experimental research designs, more controlled or experimental data were also collected. This took the form of seven argumentative essays which were embedded within the longitudinal design, counter-balanced across the participants to control for task effects. To make these tasks as comparable as possible, the writing prompts for the seven argumentative essays (see Appendix 2) were all related to the topic of foreign language learning, a topic considered relevant and familiar to all participants. The four participants were required to write the seven argumentative essays, one in each month, from October 2014 to March 2015. However, they were also asked to repeat the first writing prompt in order to compare performance of the same task over time. The participants wrote the essays by hand in their classroom. The use of word-processing software, dictionaries, and reference materials was not permitted. The participants were asked to work individually and to produce a written sample (at least 250 words) in approximately 40 minutes. The order of the tasks was counterbalanced by employing a balanced Latin square design (see Table 4.3).

Table 4.3

The Order of Writing Prompts

Participant	Data point						
	3	6	9	12	15	18	21
Flóra	1	2	6	3	5	4	1
Emese	2	3	1	4	6	5	2
Avarka	3	4	2	5	1	6	3
Levente	4	5	3	6	2	1	4
Dalma	5	6	4	1	3	2	5
Orsolya	6	1	5	2	4	3	6

Therefore, none of the participants completed the writing tasks in the same order and none of the participants completed the same writing task at the same time. The researcher did not give feedback on the participants' written samples during the data collection procedure. Table 4.3 also includes the two participants (Flóra and Orsolya) who were excluded in the data exploration stage.

4.6.2 Interviews

The four participants were interviewed on their general goals for writing improvement during the initial interviews. The aim of the initial interview was to become acquainted with the participants and to create a motivational profile (see section 5.6.1). The initial interviews with the four participants lasted approximately five minutes on average. At the end of the data collection, the four participants took part in the final interview which lasted approximately 30 minutes on average. The purpose of the final interviews was to explore any external factors that might have contributed to the four participants' L2 writing development. The interview questions of the initial and final interviews were slightly based on Zhou, Busch, Gentil, Eouanzoui and Cumming's (2006) framework to describe goals to improve writing (see Appendix 3 and Appendix 5).

Furthermore, the four participants were interviewed after the composition of the controlled written samples on seven different occasions over a seven-month period. Each semi-structured retrospective interviews lasted from approximately three to ten minutes and was conducted with each participant individually in a separate room where the EAP programme was held. In total this created two hours of interview data with each participant. Although the initial plan was to conduct lengthier interviews which provided deep insights into each of the seven more controlled pieces of writing, the reality of conducting interviews with busy students on the topic of academic writing was challenging. While participants were engaged in each retrospective interview, they mostly did not seek to prolong the interview. The interviews were conducted by the researcher himself and were recorded with an mp3-player. Table 4.4 is a summary of the data collected from the four participants.

Table 4.4

Timeline of the Data Collection

Month	Data point	Naturalistic data	Controlled data	Retrospective Interview	Interview
1	1	Essay 1			
	2	Essay 2			
2	3		Essay 3	Interview 1	Initial
	4	Essay 4			
	5	Essay 5			
3	6		Essay 6	Interview 2	
	7	Essay 7			
	8	Essay 8			
4	9		Essay 9	Interview 3	
	10	Essay 10			
	11	Essay 11			
5	12		Essay 12	Interview 4	
	13	Essay 13			
	14	Essay 14			
6	15		Essay 15	Interview 5	
	16	Essay 16			
	17	Essay 17			
7	18		Essay 18	Interview 6	
	19	Essay 19			
	20	Essay 20			
8	21		Essay 21	Interview 7	Final
	22	Essay 22			
9	23	Essay 23			

The main focus of the retrospective interviews was to explore the participants' self-regulatory process, namely self-observation, self-evaluation and goal-setting (Appendix 4). During the retrospective interviews the participants commented on the day's essay (interview question 1-9), compared the day's essay with previous essays (interview question 12), and talked about goals for future writing (interview question 13-14). In order to limit the influence of the interview questions on the possible evolution of self-regulatory processes, leading questions were avoided especially questions on goals for writing (interview question 12-13). I acknowledge that the use of implicit questions might have not elicited the goals for writing. Since the structure of the interviews was flexible, the researcher could ask any further questions if interesting topics emerged. The questions of the semi-structured retrospective interviews were loosely based on Zhou et al.'s (2006) framework of goals for second language writing improvement.

4.7 Data Coding

Prior to data exploration, the written samples and the interview data were coded. However, before the coding procedure, the seven hand-written samples were digitalised and the mp3 recorded interviews were transcribed by the researcher. The 23 written samples were submitted to three coders, including the researcher and two English teachers (see section 4.7.5 for a discussion). The interview data were coded by the researcher and an English teacher with 30 years of experience (See Appendix 7-9).

4.7.1 Lexical Complexity

As reviewed in Chapter 3, lexical complexity is a multidimensional construct (Bulté & Housen, 2012; Jarvis, 2013b; Schmitt, 2010). Therefore, five different indices were carefully selected for a fine-grained analysis. To measure general lexical complexity, the average word length (AveWL) index, calculated by dividing the total number of characters by the total number of words in a written sample, was employed. Verspoor et al. (2017) indicated that the AveWL index was a reliable indicator of general lexical complexity in their study. The measurement of lexical diversity, composed of six properties, was based on Jarvis's (2013b) model. However, Jarvis (2013b) indicated that three properties - volume, evenness, and dispersion - suffer from multicollinearity. Therefore, only the three other properties were included: variability, rarity and disparity. Lexical variability was measured by the type-token ratio content word lemmas (TTR) index, lexical rarity was measured by the CELEX log minimum frequency of content words (WRDFRQc) index, while lexical disparity was computed by latent semantic analysis (LSA) overlap - all sentences in paragraph - mean index. All of the three lexical diversity indices were calculated by Coh-Metrix 3.0 (Graesser et al 2004, 2011). Table 4.5 is a summary of the lexical complexity indices used in the main study.

Table 4.5

Lexical Complexity Measures

Sub-construct	Index	Code	Calculation
General	Average word length	AveWL	Manual
Variability	Type-token ratio, content word lemmas	TTR	Coh-Metrix 3.0
Rarity	CELEX log minimum frequency for content words	WRDFRQc	Coh-Metrix 3.0
Disparity	Latent semantic analysis overlap, all sentences in paragraph, mean	LSA	Coh-Metrix 3.0
Academic	Academic Word List	AWL	Vocabprofiler

I acknowledge that the type-token ratio (TTR) might not be the best indicator of lexical variability because it has been found to be sensitive to text length (McCarthy & Jarvis, 2007). Furthermore, some more reliable indices (MTLD, vocD) have been recommended to measure lexical variability. Nevertheless, the type-token ratio of content words was selected for two reasons. First and most importantly, in the data exploration stage, three indices were selected to represent the constructs of lexical and syntactic complexity and accuracy. All three indices were ratios which made the mathematical modelling procedure more reliable and feasible. If the MTLD or the vocD indices had been selected to measure lexical variability, there would not have been a ratio-based index among the five lexical indices. Secondly, the text length of the argumentative essays was controlled in this study excluding the sensitivity to varying degrees of text length of the TTR index.

Finally, in order to explore whether the argumentative essays contain genre-specific lexicon, the Academic Word List (AWL) index was included in the multidimensional measurement. The AWL index was calculated by the Vocabprofiler BNC computer programme (Cobb, 1994; Heatley & Nation, 1994).

4.7.2 Syntactic Complexity

As reviewed in Chapter 3, there have been four recommendations concerning the measurement of L2 syntactic complexity development:

1. syntactic complexity needs to be measured multi-dimensionally (Norris & Ortega, 2009),
2. academic writing is characterised by phrasal complexity (Biber et al., 2011)
3. the clause is a more informative unit of analysis than the T-unit (Lu, 2011)
4. complexity and accuracy need to be measured by developmentally oriented indices (Lambert & Kormos, 2014).

In line with the four recommendations of previous studies on syntactic development, five different syntactic complexity indices were carefully selected to trace the syntactic development of the four participants.

In the first stage of the manual analysis, the essays were coded for sentence types: simple, compound, complex and compound–complex. The coding of sentence types followed Verspoor and Sauter's (2000) definitions of sentence types and the results of the coding were tallied (see Appendix 7). Then, the essays were coded for finite verbs in order to calculate the finite verb ratio (see Appendix 7), which is the total number of words per finite verb index. Verspoor et al. (2017) found that the FVR index is a good indicator of general syntactic complexity. In the second stage of the manual analysis, the essays were coded for clauses by the researcher and by a native speaker of English. According to Lu (2011), there are two approaches to counting clauses. Most studies considered clauses as structures with a subject and a predicate (finite verb) including independent, adjective, adverbial, and nominal clauses (Hunt, 1965; Polio, 1997; Lu, 2011). However, some studies also counted non-finite verb phrases as clauses (Bardovi-Harlig & Bofman, 1989). In this study a clause was defined as "a group of words that expresses a whole event or situation, containing subject and a predicate" (Verspoor & Sauter, 2000; p. 151). In order to measure subordination, the dependent clause per clause index (DC/C) was calculated manually (see Appendix 7). The essays were coded for dependent clauses and the total number of dependent clauses was divided by the total number of clauses in an essay. Later, the DC/C index was also used to simulate the participants' data. To measure syntactic coordination, the T-unit per sentence (T/S) and the coordination phrase per clauses (CP/C) indices were calculated by the computational tool, L2 Syntactic Complexity Analyzer (L2SCA) (Ai & Lu, 2013; Lu, 2010, 2011; Lu & Ai, 2015). To measure phrasal complexity, the complex nominal per clauses (CN/C) index was calculated by the L2SCA, while the mean number of modifiers per noun phrase (SYNNP) index was calculated by the Coh-Metrix 3.0 computational tool.

The global measures of syntactic complexity were complemented with measures specific to the academic genre. The selection of the genre-specific syntactic structures was guided by the findings of two corpus-based studies and one longitudinal study on L2 writing development. Biber, Johansson, Leech, Conrad and Finegan (1999) provided a detailed description of clausal and phrasal level constructions that occur more frequently in academic genres than in conversation, fiction, and news. Moreover, Biber et al. (2011) found that academic writing is characterised by phrasal complexity. Mazgutova and Kormos (2015) found statistically significant increases in the use of conditional and relative clauses in the written data of intermediate EFL learners and statistically significant decreases in the use of infinitive clauses in the written data of upper-intermediate EFL writers. Moreover, Mazgutova and

Kormos (2015) found decreases in the use of prepositional phrases in both intermediate and upper-intermediate EFL learners' written data. Table 4.6 is a summary of the syntactic complexity indices used in the main study.

Table 4.6

Syntactic Complexity Indices

	Sub-construct	Index	Code	Calculation
Global	General	Finite verb ratio	FVR	Manual
	Subordination	Dependent clause per clause	DC/C	Manual
	Clausal coordination	T-unit per sentence	T/S	L2SCA
	Phrasal coordination	Coordinate phrases per clauses	CP/C	L2SCA
	Phrasal complexity	Complex nominals per clauses	CN/C	L2SCA
		Modifiers per noun phrase	SYNNP	Coh-Metrix 3.0
Academic		Conditional clauses	ConC	Manual
		Infinitive clauses	InfC	Manual
		Relative clauses	RelC	Manual
		Prepositional phrases	DRPP	Coh-Metrix 3.0

Warchal (2010) claimed that conditional clauses are especially important in argumentative writing because they can perform a large array of functions. Infinitive clauses can also be used as post-modifiers, and they are more common in writing than in speaking. Byrnes and Sinicrope (2008) found that relative clauses occur frequently in academic writing, although they are less common than prepositional phrases. Furthermore, the frequency of relative clauses is generally used as an indicator of syntactic complexity since they are one of the most explicit types of noun modification (Jucker, 1992). Biber et al. (1999) found that prepositional phrases are the most frequent type of noun post-modifiers in academic discourse. Consequently, the normed rate of occurrence of the conditional, infinitive, and relative clauses was calculated. The normed rate of occurrence was calculated by dividing the raw frequency of clause type in a specific text by the total number of words in the specific text multiplied by the norming number (Gray, 2015). In this study, the norming number was 1000. In addition, the incidence score of the prepositional phrases was used, calculated by the Coh-Metrix 3.0.

4.7.3 Accuracy

Since accuracy is a unidimensional construct, as reviewed in Chapter 3, it was measured by only one index, the error-free clauses per clauses (EFC/C) index. First, the clauses were coded based on the Louvain error-tagset (Dagneaux et al., 1998). As reviewed in Chapter 3, the Louvain error-tagging system is hierarchical and includes seven main error domains. Second, the EFC/C index was calculated by dividing the total number error-free clauses by the total number of clauses in an essay. Table 4.7 is a summary of the accuracy indices used in this study.

Table 4.7

Accuracy Indices

Sub-construct	Index	Code	Calculation
General	Error-free clauses per clauses	EFC/C	Manual
Error-types	Formal	F	Manual
	Grammatical	G	Manual
	Lexico-grammatical	X	Manual
	Lexical	L	Manual
	Word redundant, word missing, word order	W	Manual
	Punctuation	P	Manual

4.7.4 Indices for Variability Analyses and Mathematical Modelling

For the mathematical modelling, three indices were selected to represent (1) lexical complexity, (2) syntactic complexity, and (3) accuracy. During the initial mathematical modelling procedures, the operationalization of the three constructs by using the general measures of lexical complexity (AveWL), syntactic complexity (FVR) and accuracy (EFC/C) was found unreasonable for two main reasons. First, the AveWL, FVR, and EFC/C indices are not represented on the same scale. Therefore, the three indices should have been normalised which would have made the model unstable. Second, an analogous maximal value of the three indices would make the modelling procedure feasible. Consequently, those three indices were selected for the mathematical modelling which were analogous indices mathematically. In this case all three indices were ratios. The type-token for content words (TTR) was chosen to represent lexical complexity, while syntactic complexity was operationalized by the dependent clause per clause (DC/C) index. Accuracy was represented by the error-free clause ratio (EFC/C).

4.7.5 Inter- and Intra-Coder Reliability

The 23 written samples were submitted to three coders. Coder no. 1 was the researcher, coder no. 2 was a native English teacher with more than 30 years of experience, and coder no. 3 was a non-native English teacher with more than 10 years of experience.

The positive overlap ratio (POR) (van Geert & van Dijk, 2003) was calculated to measure inter-coder reliability. The POR reflects the percentage of overlapping positive cases confirmed by the two coders. The calculation of the POR index included two steps: (1) the identification of the cases confirmed by the two coders, (2) the identification of the number of the overlapping cases. The calculation of the POR index is illustrated in Equation 4.1.

Equation 4.1 Positive overlap ratio

$$POR = \frac{2Y}{X_1 + X_2}$$

In Equation 4.1, X_1 denotes the number of cases confirmed by coder 1, while X_2 refers to the number of cases confirmed by coder 2. Y denotes the number of mutually confirmed cases.

In this study, the POR value reached 95% for the complexity measure, 75% for the accuracy measure, and 88% for the coding the interview data for key self-regulatory processes.

Since accuracy is a more ambiguous concept in nature than complexity, a lower percentage was expected. Van Geert and van Dijk (2003) also pointed out that "high agreement would be an indicator of low-quality rating, for instance rating based on common errors and shared biases" (p. 273). In this study, the disagreement between the coders was resolved by a discussion. Table 4.8 provides the summary of the data coding procedures and the final POR values.

Table 4.8

Inter-Coder Reliability

	Complexity	Accuracy	Self-regulation
Coder 1 and Coder 2	POR = 95%	-	POR = 88%
Coder 1 and Coder 3	-	POR = 75%	-

4.8 Quantitative Data Exploration

The data collected for this study were analysed by applying both quantitative and qualitative analyses. The quantitative data exploration included descriptive statistics, statistical inference (resampling), correlation, regression and time series analyses. The qualitative analysis included the quantification of self-regulatory processes, and identification of attractor states and CDST abstractions in the interview transcripts. Figure 4.3 illustrates the multi-methodological data exploration applied in this study.

The first stage of the quantitative analysis involved the visualisation of the development of complexity and accuracy. A smoothing technique, simple moving averages, was employed to look at the general trend in the constructs (see section 4.8.1 for a discussion).

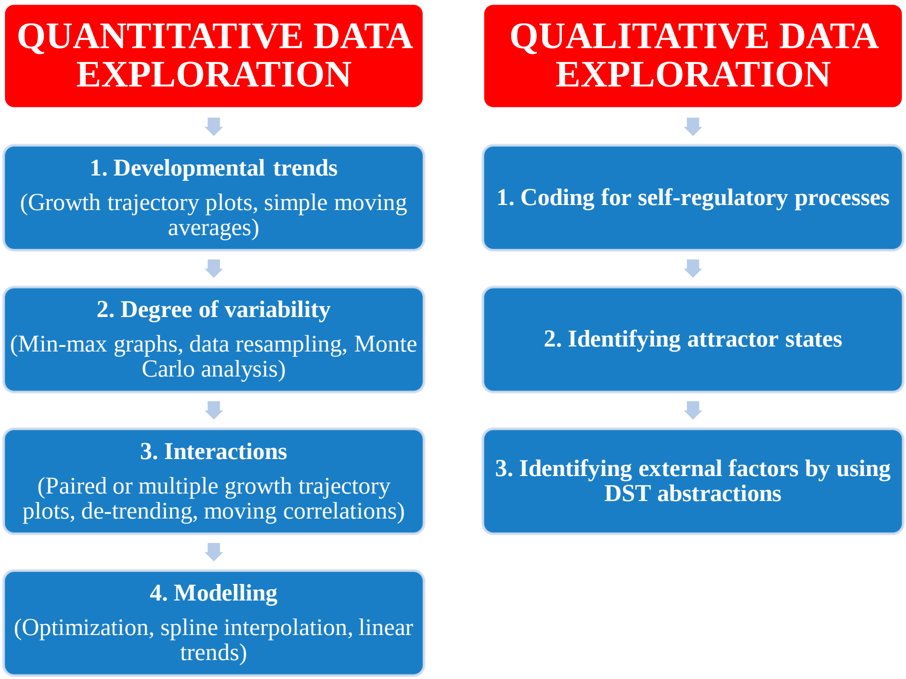


Figure 4.3 Multi-methodological data exploration

The second stage of the quantitative analysis included the exploration of the degree of variability in the constructs of complexity and accuracy. First, the raw data were visualised by min-max graphs and then the raw data were resampled and the developmental peaks in the data set were tested against chance by running Monte Carlo analyses (see 4.8.2 for a discussion). The third stage of the quantitative analysis involved the visualisation of the interactions of the constructs. The raw data were de-trended and moving windows of correlations were calculated in order to visualise the interactions. Finally, an overall correlation coefficient was calculated between the constructs (see 4.8.3 for a discussion). The final stage of the quantitative analysis involved the mathematical modelling of the development of three specific constructs: lexical variability, subordination, and accuracy (see 4.8.4 for a discussion). Table 4.9 provides a summary of the methods and techniques used in this study.

Table 4.9

Quantitative Data Exploration Methods and Techniques

	Method and technique	Programme
Developmental trend	data smoothing by simple moving averages	MS Excel
Degree of variability	data visualisation by min-max graphs	MS Excel
	data resampling	MS Excel PopTools
	testing for significance by Monte Carlo analysis	MS Excel PopTools
Interaction	data visualisation	MS Excel
	de-trending or residual plots	MS Excel
	moving correlations (Spearman's rank)	MS Excel
Modelling	model programming	MS Excel
	data smoothing by spline interpolation	R
	linear trends	MS Excel
	Correlations (Spearman's rank)	R

4.8.1 Developmental Trends

One of the most frequently used technique in CDST data exploration is the growth trajectory plot. A trajectory plot shows the change as a series of values plotted along the x-axis, which represents time. Important information about the nature of development can be gained by plotting the growth trajectories. Regression lines can also be added to the raw data to visualize the general trend of the development. There are three types of trend: upward, downward and sideways (horizontal) (Barros, 2007; Little, 2011; Little & Farley, 2012).

In individual raw data it is sometimes difficult to see the general trend due to fluctuations. By using smoothing techniques, it is possible to observe the general trend in the raw data. According to van Dijk, Verspoor and Lowie (2011), "the purpose of a smoother is to 'sketch' the general trend of the data and leave out many of the irregularities of the actual data" (p. 72).

One of the most popular and most extensively used smoothing techniques, is simple moving averages (SMA). A simple moving average is the unweighted mean of the preceding *n* data. The data points included in the calculation of the average is called: the window. In this study, a window of five observations was used. The bigger the size of the window, the greater the extent of the smoothing procedure. First, the values of the first five data points (data point 1, 2, 3, 4, and 5) were added. Second, the sum of the first five data points was divided by five in agreement with the size of the window. The same procedure was repeated for the next five data points (data point 2, 3, 4, 5, and 6). Finally, the growth trajectories and the simple moving averages were plotted in one graph to inspect the raw and the smoothed data simultaneously. The purpose of the simple moving averages is to display the general trend of a specific index. However, more refined smoothing techniques such as cubic spline interpolation can also be employed as in this study in the data simulation procedures (see 4.8.4 for a discussion).

If the trend cannot be discerned from the smoothing algorithm (simple moving averages), a linear trend is added to show the direction of the development in this study. In this study the linear trend lines were calculated in the MS Excel.

4.8.2 Degree of Variability

As outlined in Chapter 2, complex dynamic systems show a certain degree of variability over time, even if the systems are rather stable (Verspoor et al., 2008). The developmental process can be explored by distinguishing random variability from "developmental variability" (van Dijk et al., 2011, p. 75). The developmental variability can be explored by following three steps: (1) min-max graph, (2) resampling, and (3) Monte Carlo analysis. The use of min-max graphs, data resampling and Monte Carlo analysis are frequently employed in studies on second language writing development (Verspoor et al., 2008; van Dijk et al., 2011), second language vocabulary development (Zheng, 2016), and second language motivation (Piniel & Csizér, 2015).

Van Geert and van Dijk (2002) developed a technique, called min-max graphs, to facilitate the visualisation of the degree of variability. The min-max graph enables us to see the general patterns of variability along with the raw data by using a moving window. The size of the window depends on the number of the data points available for the researcher. Therefore, the smaller the dataset, the smaller the window. In this study, a window of five data points was employed as suggested by Verspoor et al. (2011). The moving window is a time frame that moves one position every time. The first window includes data point 1, 2, 3, 4, and 5, while the second window includes data points 2, 3, 4, 5, and 6 and so on. For each window, the maximum and minimum values are calculated and then displayed in the min-max graphs. The min-max graph displays the bandwidth of the observed data. Van Dijk et al. (2011) pointed out that the min-max graph shows “the amount of variation in relation to developmental jumps” and “the wider the bandwidth, the greater the amount of variation” (p. 76). The min-max graphs were calculated and displayed in the MS Excel in this study.

The min-max graph is a descriptive technique; however, the observations can be tested against chance by data resampling (van Geert & van Dijk, 2002). In the data resampling (Efron & Tibshirani, 1993; Good, 1999) method a huge number of subsamples are randomly drawn from the original sample in order to detect developmental peaks in the dataset. Van Dijk et al. (2011) cautioned the development peak “should not be just one isolated jump” (p. 80). As a first step, a simple moving average over two data points is calculated in the original dataset and repeated until the last data point in the original dataset. As a second step, the difference between data points is calculated in the averaged dataset. Data point 2 is subtracted from data point 1 in the 2-step difference values, while data point 3 is subtracted from data point 1 in the 3-step difference values. The same calculation is repeated until the 6-step difference. As a third step, the maximal distances between the data points are calculated for each step (from step 2 to step 6) and the maximal positive distance of every step is used as the testing criterion in the resampling model. As a fourth step, the original data are reshuffled with replacement by using MS Excel Poptools. As a final step, a Monte Carlo analysis is run with 5,000 simulation steps by selecting the resampled criterion for the dependent range, and by selecting the original criterion for the test values. The likelihood that the peak is produced by the random model is calculated. If the likelihood is small (below 5%), the peak is not considered to have been caused by a random fluctuation.

4.8.3 Interactions

As outlined in Chapter 2, complex dynamic systems are completely interconnected. The interconnectedness of complex dynamic systems is usually explored by several different techniques (Caspi, 2010; Verspoor & van Dijk, 2011). These techniques include data normalisation, data smoothing, de-trending, the calculation of moving correlations, and overall correlation coefficients. However, previous CDST studies differ in the application of these techniques. For example, Verspoor and van Dijk (2011) applied data smoothing (Lowess) to compare more global interactions. Conversely, Caspi (2010) did not smooth the raw data but applied data de-trending techniques. In this study, the latter approach was applied because it was found that data smoothing, to a great extent, distorted the information about the interactions. In conclusion, the interactions between complexity and accuracy were explored in five different steps in this study: (1) data normalisation, (2) paired or multiple growth trajectory plots, (3) data de-trending, (4) moving correlation plots, and (5) the calculation of an overall correlation coefficient.

In order to explore the interactions between general lexical complexity (AveWL), general syntactic complexity (FVR) and accuracy (EFC/C), the raw data were normalised. The normalisation of the raw data enables us to compare trajectories with data that are not presented on the same scale. In the normalisation procedure, the raw data values were recalculated to vales from 0 to 1. The normalised values were plotted and trajectories of the general lexical complexity (AveWL), the general syntactic complexity (FVR) and the accuracy (EFC/C) indices were inspected. However, for the interaction between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C), the raw data were not normalised because the TTR, DC/C, and the EFC/C indices are ratios, and they are presented on the same scale.

In addition to inspecting the developmental trends, the growth trajectories also provide valuable information about the interactions between variables over time. Caspi (2010) pointed out that “plots of two or more trajectories can expose their interactions over time” (p. 28). In the growth trajectory plots, there are two types of shifts: (1) parallel and (2) alternating. When a shift between data point x and y in one trajectory coincides with a shift between data point x and y in another trajectory, it is called a parallel shift. Conversely, when a shift between data point x and y in one trajectory differs from a shift between data point x and y in another trajectory, it is called an alternating shift. Parallel shifts are usually interpreted as supportive relationships, while alternating shifts are generally interpreted as competitive relationships between two variables. Nevertheless, the interpretations are speculative and descriptive in nature and require further exploration by mathematical modelling as suggested by Caspi (2010) and Verspoor & van Dijk (2011).

To inspect the variability independently of the linear trend, the raw data were de-trended. Verspoor et al. (2011) pointed out that “detrended data are useful when you want to make sure the degree of variability and interaction between variables is not distorted by the incline of the slope” (p. 181). The residuals, or de-trended values, are calculated by subtracting the linear trend from the raw data. First, the intercept and the slope of the raw dataset are calculated. Second, the sum of the intercept and the slope is multiplied by the number of data points. For example, at data point 1 the sum of the intercept and the slope is multiplied by 1, at data point 2 by 2 and so on. In this study, data de-trending was performed in MS Excel.

An overall correlation coefficient was calculated in order to explore the surface interactions between variables. In this study, a Spearman’s rank correlation coefficient was calculated because it was assumed that developmental data might not be normally distributed.

For each pair, the Spearman's rho was calculated in R. However, the overall correlation coefficient is a static value which can be supplemented with a moving correlation. Caspi (2010) pointed out that moving correlation "shows temporal changes in the coefficient values in a moving window of several observations" (p. 30). In this study, a moving window of five Spearman's correlation values was used as in previous CDST studies (Caspi, 2010; Rosmawati, 2016). The moving correlations were calculated in MS Excel.

4.8.4 Data Modelling

Visualising interactions, calculating overall correlation coefficients, and moving correlations are only descriptive techniques. Verspoor and van Dijk (2011) recommended data simulation to confirm the findings of interactions between variables. Despite Verspoor and van Dijk's (2011) recommendation, there is a scarcity of studies applying mathematical modelling to confirm the interactions between variables. Exceptions are Caspi's (2010) study in which a precursor model was applied to simulate both L2 vocabulary and writing data and Chan's (2015) study in which the Hidden Markov Model was used.

In this study, the interactions between lexical and syntactic complexity and accuracy were simulated. Due to data modelling constraints, three mathematically analogous indices were selected: lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C). All three indices are ratios which facilitate data simulation. I acknowledge that lexical variability and subordination are rather specific indices and might not be able to fully represent the constructs of lexical and syntactic complexity respectively. Although the general lexical complexity (AveWL) and general syntactic complexity (FVR) and accuracy (EFC/C) indices might have been better candidates for representing each construct, they are not on the same scale. Therefore, data normalisation should have been performed which might have rendered the data simulation pointless.

Alternating growth and variability patterns across lexicon and syntax are usually associated with the precursor model (Caspi, 2010). The precursor model was first used to model data of early L1 vocabulary development (van Geert, 1991). According to the precursor model, a certain amount of vocabulary is prerequisite for the emergence of syntax (Caspi, 2010). In this study, the precursor model was abandoned after initial data simulation procedures for two reasons. First, the participants were at upper-intermediate or advanced levels where it was not assumed that the development of lexicon precedes the emergence of syntax. Second, it was found that a simplified model, which omits the precursor relationships, might more accurately confirm the hypothesised interactions between the three growers.

Consequently, a model of three connected variables, or growers in dynamic parlance, were simulated to confirm the hypothesised interactions between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) in this study. Growers might influence each other positively (support) or negatively (competition). Furthermore, the interactions between the growers might change over time. The model iterates coupled logistic equations that configure interactions between each grower as shown in Equation 4.2, 4.3 and 4.4.

Equation 4.2 Coupled logistic equation (Grower A)

$$A_{n+1} = A_n \times \left[1 + r_A - \frac{r_A \times A_n}{K_A} + S_{BtoA} \times B_n + S_{CtoA} \times C_n\right]$$

Equation 4.3 Coupled logistic equation (Grower B)

$$B_{n+1} = B_n \times \left[1 + r_B - \frac{r_B \times B_n}{K_B} + S_{AtoB} \times A_n + S_{CtoB} \times C_n\right]$$

Equation 4.4 Coupled logistic equation (Grower C)

$$C_{n+1} = C_n \times \left[1 + r_C - \frac{r_C \times C_n}{K_C} + S_{AtoC} \times A_n + S_{BtoC} \times B_n\right]$$

Equations 4.2-4.4 illustrate the models of three connected growers. The equations include both the property and relational parameters. The property parameters are the initial value (ini), the growth rate (r), and the carrying capacity (K). In the equations the initial values are A_{n+1}, B_{n+1}, and C_{n-+1} where n denotes time. The initial values are configured as the corresponding data onset values. For example, the initial value of Grower A (lexical variability) was set 0.551 in the model in Dalma's data (see Table 5.5), since the corresponding data onset value was 0.551. The growth rates were set manually for each variable. The carrying capacity was set to the default value of 1 for each grower because the biggest attainable value was 1 for each ratio. For example, a value of 1 for the TTR index would indicate that there is no repetition in a given text. The relational parameters were denoted as support (S), although it can also be a negative value reflecting a competitive relationship between the growers. Table 4.10 shows the model parameters of Equation 4.2-4.4.

Table 4.10

Model Parameters

Parameter	in Equation 4.2-4.4	Definition
rate_A	r_A	Growth rate of Grower A (lexical variability - TTR)
rate_B	r_B	Growth rate of Grower B (subordination - DC/C)
rate_C	r_C	Growth rate of Grower C (accuracy- EFC/C)
support_AtoB	S_{AtoB}	Level of support from A to B
support_BtoA	S_{BtoA}	Level of support from B to A
support_AtoC	S_{AtoC}	Level of support from A to C
support_CtoA	S_{CtoA}	Level of support from C to A
support_BtoC	S_{BtoC}	Level of support from B to C
support_CtoB	S_{CtoB}	Level of support from C to B
ini_A	A_n	Initial value for Grower A (data based)
ini_B	B_n	Initial value for Grower B (data based)
ini_C	C_n	Initial value for Grower C (data based)
K_A	K_A	Carrying capacity for Grower A (set to 1)
K_B	K_B	Carrying capacity for Grower B (set to 1)
K_C	K_C	Carrying capacity for Grower C (set to 1)

The mathematical models can be optimized in order to achieve the best-fit between the data and the model. In this study, the relational control parameters and the growth rates were converged via the Simplex algorithm (Nelder & Mead, 1965). The purpose of the optimization process is to find a local solution to a problem including several parameters by minimizing the outcome of a function. In this study the outcome of a function is the sum of squared differences between the model and the data. In order to reach an optimal solution, the matrix of squared differences was recalculated up to 10,000 times with a tolerance level of 0.00001. The optimization procedure was performed by the Poptools add-in in MS Excel.

The final step in the data simulation procedure included the evaluation of the model outcome. According to Caspi (2010), a model outcome can be evaluated in two different ways: (1) correlating the model values with the data, linear trend and spline values, and (2) comparing correlation matrixes across data, linear trends, spline and model (Caspi, 2010). In order to perform the evaluation, the linear trends of the raw data were calculated in MS Excel. Furthermore, spline interpolations were calculated in R Project for Statistical Computing 3.4. Data smoothing can be performed by calculating moving averages as discussed in section 4.8.1. However, in a moving average the previous values are more influential than the subsequent values. In addition, moving averages are also sensitive to the effect of extreme values. Therefore, another type of data smoothing procedure is used in the last stage of analysis, cubic spline interpolation. The spline interpolation method (Green & Silverman, 1994) locally fits a cubic smoothing spline to the data sets and it is more dynamic than linear trend lines. The raw data is minimized by a wide class of piecewise polynomial functions during spline interpolation. The spline interpolation eliminates high fluctuations in the data since they are segmented. In this study, the splines are displayed in combination with the raw data, the linear trends and the optimized model outcome (see Figure 5.26 for an example).

4.9 Qualitative Data Exploration

Although this study is primarily a quantitatively-driven exploration of second language writing development, the qualitative data also provides valuable information about the behaviour of the participants. One of the most important merits of this study is the application of mixed methods research by integrating interview data in the data exploration process filling the gap in the CDST literature as indicated by previous research (Rosmawati, 2016).

As shown in Figure 4.3, the qualitative data exploration included three different stages: (1) coding for self-regulatory processes, (2) identifying attractor states, and (3) identifying external factors by using CDST abstractions. Although the qualitative analysis initially included only the first and the third stages, it was found important to include an intercalating stage (second stage), namely the identification of attractor states. The identification of attractor states was motivated by the results of the coding for self-regulatory processes. As it will be shown in Chapter 5-8, the self-regulatory processes did not evolve substantially in the four participants' interview data as it was expected. Therefore, the interview data were coded for possible attractor states in order to explore why the self-regulatory processes did not evolve substantially.

4.9.1 Self-Regulatory Processes

The first stage of the qualitative data exploration involved the coding of the interview data according to three key self-regulatory processes: (1) self-observation, (2) self-evaluation, and (3) goal-setting. The coding of the interview data was performed by the researcher of this study and an English teacher. The coding scheme was based on Nitta and Baba's (2015) study. Self-observation in this study concerned the participants' impressionistic descriptions of their thoughts and actions, usually containing shallow descriptions of their own writing. Conversely, self-evaluation required evaluative comments that involved several cognitive processes such as specifying, reasoning, analysing or comparing essays. In other words, the participants sometimes analysed their own writing or compared the essays written for this study. Furthermore, the participants sometimes specified and gave reasons for their actions concerning writing. These actions were coded as self-evaluation processes. The third category, goal-setting, included comments which concerned future specific actions concerning improving their own writing. Table 4.11 presents and illustrates the main categories, broken down to further subcategories, that were generated though repeated reading of the retrospective interviews.

Table 4.11

Self-Regulatory Processes

Self-regulatory process	Type	Participant	Example
Self-observation	Language	Dalma	I wrote in an easier way so… so this is my main vocabulary.
	Composing processes	Levente	I collected my thoughts and I wrote it down.
	Content	Emese	I just thought about the ideas. I decided to write about the native language teacher's positive aspects and then the negative.
	Quality	Levente	I think it was a good representation
Self-evaluation	Language	Dalma	I'm not satisfied with my vocabulary because I feel I repeat myself too much.
	Composing processes	Avarka	I was not strictly writing the positive things first and then the negative ones. I started saying what I agree and what I don't agree with and then again. So I wasn't really consistent because I always had a different idea.
	Content	Avarka	It was easier than in other essays because I have personal experience.
	Quality	Avarka	I think it's not an improvement because these ones are longer and I think the sentences are better-structured.
Goal-setting	Language	Dalma	I want to read more in English and just improve my language and my grammar so just deal with English.

4.9.2 Content Analysis

As reviewed in Chapter 2, complex dynamic systems might settle into attractor states. During the data exploration it was discovered that the four participants did not develop their self-regulatory processes. Therefore, the interview data were reread to identify possible attractor states. The coding scheme of the interview data at this stage was adopted from Waninge's (2015) study. It was expected that dominant attractor states surface the data and these attractor states can be identified by the content analysis of the interview data. According to Dörnyei (2007), qualitative content analysis is a latent level analysis since "it concerns second-level, interpretive analysis of the underlying deeper meaning of the data" (p. 246). Dörnyei (2007) also pointed out that "the qualitative categories used in content analysis are not predetermined but are derived inductively from the data analysed" (p. 245). Waninge (2015) kept the interviews deliberately short since she assumed that dominant attractor states might emerge in the first few minutes of an interview. Therefore, Waninge's (2015) study provides evidence that the shortness of the interviews is not a problem to identify attractor states in interview data. Table 4.12 illustrates examples of two different attractor states found in the participants' interview data.

Table 4.12

Content Analysis

		Participant	Example
Attractor state	boredom	Emese	There were one or two topic which was… not that interesting
	anxiety	Dalma	I think that was the problem with my motivation because that was my last semester at the university and… it was really nervous for me because my teacher was horrible… and I was really stressed about my thesis… so I think my motivation could have been better

As reviewed in Chapter 2, co-adaptation refers to the interactions of two or more closely related complex dynamic systems. Furthermore, events (referred to as perturbations) sometimes disrupt the stability of a complex dynamic system. The last stage of the qualitative data exploration involved three minor steps. The three steps were based on Henry's (2015) qualitative analysis of interview data. In the first step, the interview data were read several times for each participant, making notes for interesting features. In the second step, the interview transcripts were re-read, this time using the abstractions of CDST (co-adaptations and perturbations) as a compass. The initial notes and ideas were transformed into theoretically resonant themes. In the third step, connections were searched, and the themes were grouped together in categories. The themes included indications of changes in attractor states, perturbations and co-adaption between the linguistic and motivational systems. Figure 4.4 shows the co-adaptation of the motivational and linguistic systems in Dalma' data.

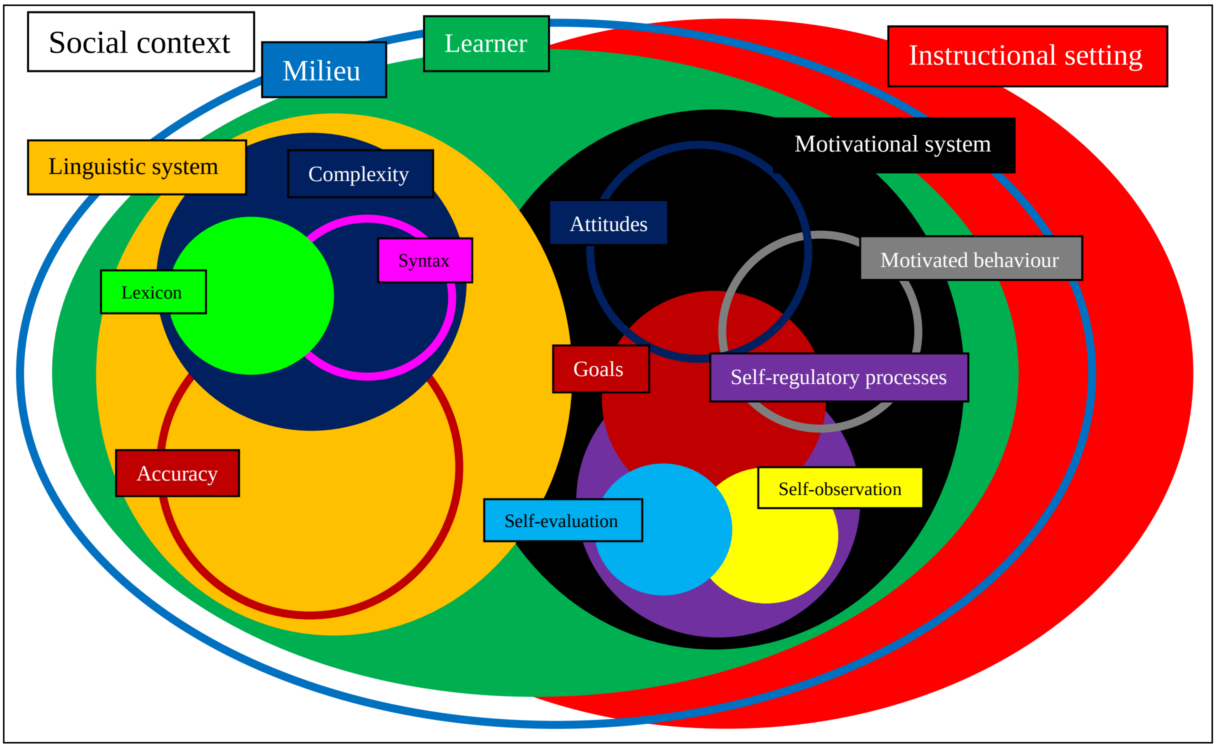


Figure 4.4 Co-adaptation in Dalma's data (based on Csizér, Kormos & Sarkadi, 2010)

Figure 4.4 illustrates the co-adaptation of the linguistic and motivational systems. The activated systems and subsystems are filled, while inactivated systems remained unfilled in Figure 4.4. A change in Dalma's instructional setting (successful completion of her state examination, time to learn), displayed in red filled ellipsis, triggered a change in the learner's system, shown with a green filled ellipsis. Subsequently, a change in the learner's system caused a change in her motivational system demonstrated by goal-setting. A change in the self-regulation subsystem, shown with a purple filled ellipsis, triggered a change in Dalma's linguistic system shown by a statistically significant developmental peak in her lexical system.

As reviewed in Chapter 2 (section 2.3.10), perturbations refer to events that disrupt the stability of a complex dynamic system. Examples of perturbations in the language classroom include teacher interventions (Larsen-Freeman & Cameron, 2008). Hiver (2015) exemplified perturbations with the sudden news of an unexpected language test, the announcement of poor test results and the use of an extrinsic reward or prize. In this study perturbations referred to the vocabulary tests written during the EAP course.

4.10 Summary

This study is a longitudinal investigation of second language writing development over a nine-month period. The participants of this study were four Hungarian EFL learners enrolled in an EAP course offered by a Hungarian university. The data were collected by both the researcher and the teacher of the EAP course. The full dataset included 92 written samples and 28 retrospective interviews (23 written samples and seven retrospective interviews per participant) collected over a nine-month period. Although both quantitative and qualitative data were collected, this study is characterised as quantitatively dominant mixed methods research due to the quantity and the quality of the collected data. The quantitative data were explored at four different stages: developmental trends, the degree of variability, interactions, and modelling. The qualitative data were explored at three different stages: coding the interviews for self-regulatory processes, identifying attractor states in the interview data, and using the CDST abstractions.

Chapter 5. Case Study One

Chapter 5 presents the changes in the lexical and syntactic complexity and accuracy indices in Dalma's written data. Furthermore, the changes in her self-regulatory processes are presented. There are six main sections: (1) the participant and the context, (2) the developmental profile, (3) the degree of variability, (4) interactions, (5) modelling, and (6) motivation.

5.1 The Participant and the Context

Dalma (pseudonym), a 22-year-old female Bachelor of Engineering student, was the first participant of this study. Dalma's English language education started at the age of fifteen at a secondary school in Budapest. Previously, she had learned only German as an L2 at elementary school for eight years. At secondary school she attended three classes on a weekly basis over four years. At the end of secondary school Dalma took The European Language Certificates (TELC) language test which placed her at a B2 CEFR level.

During her Bachelor studies she did not engage in formal language education apart from occasional private English language classes. Her main aspiration was to continue her studies in a master's programme at a foreign university, preferably in Germany. Since the master's programme required the applicants to take either the IELTS or the Test of English as a Foreign Language (TOEFL) language tests, Dalma decided to enrol in the EAP course offered by a university in Budapest, Hungary. She was planning to take the IELTS language test in the summer of 2015.

The data for this case study were collected using the procedures outlined in Chapter 4. Table 5.1 shows Dalma's written corpus, interviews, and timeline.

Table 5.1

Written Corpus, Interviews, and Timeline

		Written sample			Interview	
Data point	Code	Writing prompt	Date of submission	Word count	Type	Duration
1	Essay 1	WP1	19/09/2014	300		
2	Essay 2	WP2	03/10/2014	253		
3	Essay 3	CWP5	17/10/2014	316	Initial	5:51
					Retrospective 1	3:27
4	Essay 4	WP3	31/10/2014	299		
5	Essay 5	WP4	07/11/2014	245		
6	Essay 6	CWP6	21/11/2014	238	Retrospective 2	4:42
7	Essay 7	WP5	28/11/2014	266		
8	Essay 8	WP6	05/12/2014	271		
9	Essay 9	CWP4	19/12/2014	271	Retrospective 3	11:02
10	Essay 10	WP7	19/12/2015	297		
11	Essay 11	WP8	09/01/2015	239		
12	Essay 12	CWP1	23/01/2015	357	Retrospective 4	9:48
13	Essay 13	WP9	30/01/2015	302		
14	Essay 14	WP10	06/02/2015	277		
15	Essay 15	CWP3	27/02/2015	329	Retrospective 5	9:37
16	Essay 16	WP11	27/02/2015	254		
17	Essay 17	WP12	06/03/2015	239		
18	Essay 18	CWP2	20/03/2015	216	Retrospective 6	10:49
19	Essay 19	WP13	27/03/2015	229		
20	Essay 20	WP14	03/04/2015	232		
21	Essay 21	CWP5	17/04/2015	293	Retrospective 7	11:23
					Final	30:22
22	Essay 22	WP15	24/04/2015	222		
23	Essay 23	WP16	08/05/2015	242		
		Total word count		6187	Total duration	1:51:49

Note. Controlled data points are highlighted. WP = Writing prompt; CWP = Controlled writing prompt.

5.2 The Developmental Profile

Lexical and syntactic complexity and accuracy indices were traced as a function of time (represented on the x-axis as the number of data points from the onset of the study). These growth trajectories were supplemented with a smoothing technique, simple moving averages (SMA), to explore global trends of the development. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity and (3) accuracy.

5.2.1 Lexical Complexity

Figure 5.1 shows that all five indices developed nonlinearly and fluctuated to different degrees over the nine-month investigation. However, the normalized indices of lexical complexity also show that three measures, namely the AveWL, TTR, and AWL indices showed similar patterns. For example, all three indices exhibited a peak at data points 16 and 17.

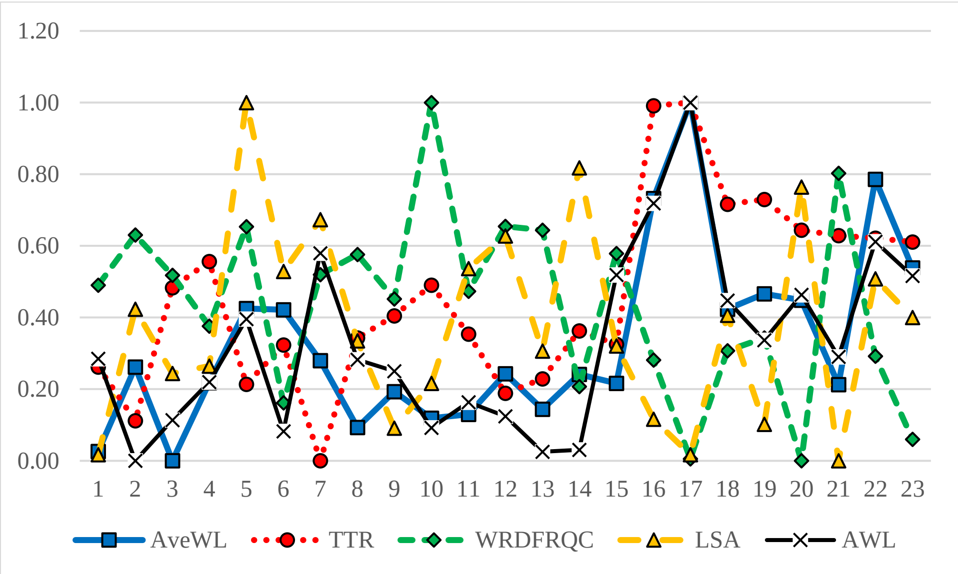


Figure 5.1 Lexical complexity indices

Average word length (AveWL) was employed to measure general lexical complexity. The higher the AveWL index, the longer the words are in a given text. Verspoor et al. (2017) claimed that longer words tend to be more academic. Figure 5.2 shows that the AveWL index was relatively stable between data points 1 and 15 ranging from 4 to 5 characters. However, at data point 16, the AveWL index increased and exceeded 5 characters. The AveWL index reached its zenith at data point 17 (AveWL = 5.67) and remained relatively high between data points 18 and 23. The smoothing algorithm (3-period simple moving averages) of the general lexical complexity (AveWL) index displays an upward trend over the nine-month investigation.

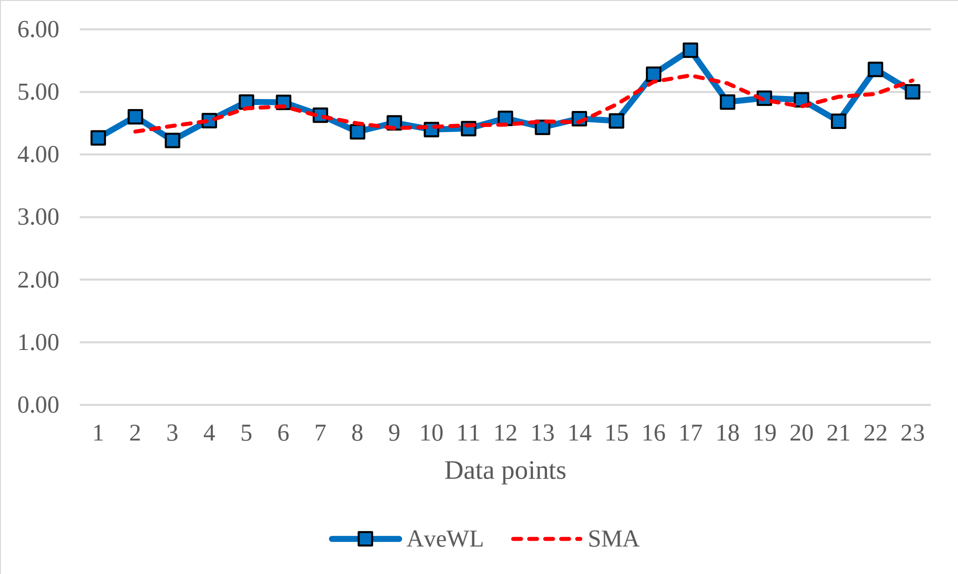


Figure 5.2 General lexical complexity (AveWL)

Lexical diversity is used as an umbrella term for variability, rarity and disparity (Jarvis, 2013b). Therefore, the first property of lexical diversity, variability, was measured by the type token ratio (TTR) of content words (Figure 5.3).

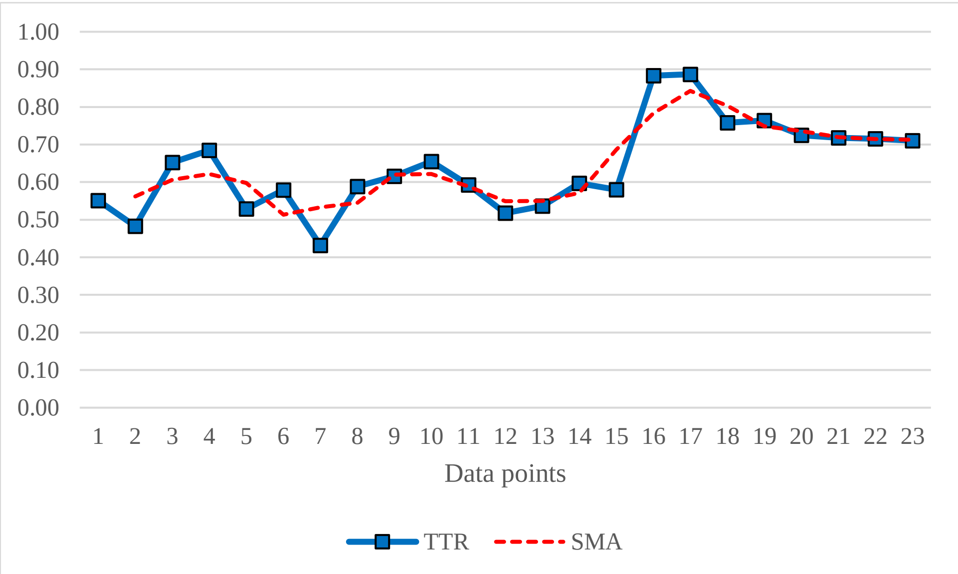


Figure 5.3 Lexical variability (TTR)

The higher the TTR index, the more variability the text exhibits. Figure 5.3 shows that the TTR index fluctuated between 0.43 and 0.69 between data points 1 and 15. The lowest TTR value (TTR = 0.43) was also measured at data point 7. The TTR index also showed a peak at data point 16 (TTR = 0.88) and at data point 17 (TTR = 0.89) as the AveWL index. In addition, the TTR index, as well as the AveWL index, remained relatively stable and high between data points 18 and 23. The smoothing algorithm of the lexical variability (TTR) index displays an upward trend over the nine-month investigation.

The second property of lexical diversity, rarity, was measured by the CELEX log minimum frequency of content words index (WRDFRQc) and shown in Figure 5.4. The lower the WRDFRQc value, the less frequent words are in a text.

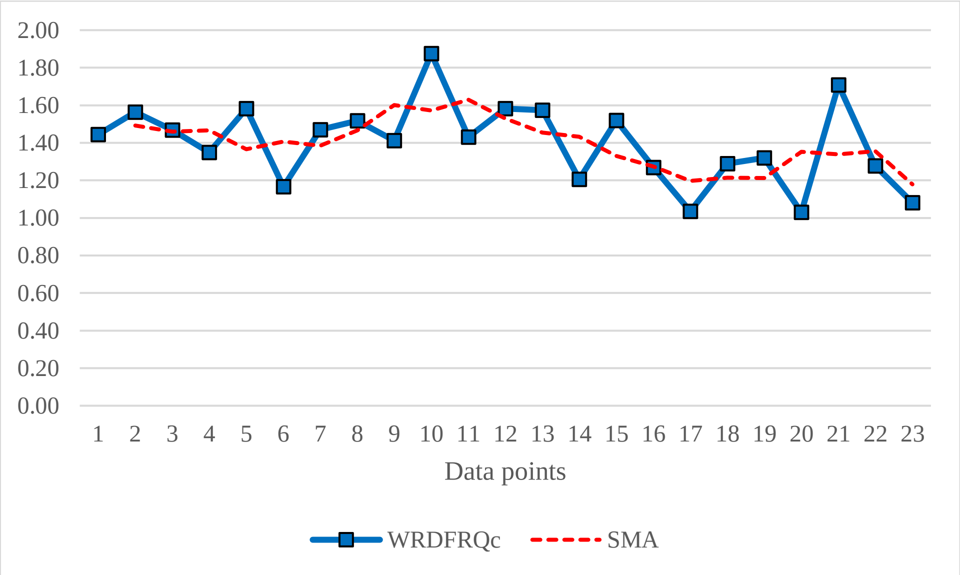


Figure 5.4 Lexical rarity (WRDFRQc)

The WRDFRQc index showed slight fluctuations over the nine months. The WRDFRQc reached its highest point at data point 10 (WRDFRQc = 1.88) which suggested the most frequent words were used at data point 10. The lowest values were measured at data point 17 (WRDFRQc = 1.04) and 20 (WRDFRQc = 1.03) which implied that the least frequent words were used at data points 17 and 20. This finding was not surprising since both the AveWL and the TTR indices were the highest at data point 17. The smoothing algorithm of the lexical rarity (WRDFRQc) index displays a downward trend over the nine-month investigation.

Verspoor et al. (2017) pointed out that "unique words, academic words and average word length all correlate highly" since they partially tap into the same construct (p. 9). Therefore, it can be assumed that the trajectory of the WRDFRQc index curve might be the inverse of the trajectory of the AveWL index. Indeed, if we observe Figure 5.2 and Figure 5.4, the smoothing lines were the inverse of each other. In addition, the highest value of the AveWL index was measured at data point 17 (AveWL = 5.67), while the second lowest value of the WRDFRQc index was measured at data point 17 (WRDFRQc = 1.04).

The third property of lexical diversity, disparity, was measured by the latent semantic analysis index (LSA) and shown in Figure 5.5. Lexical disparity measures how evenly the tokens are distributed across types in a text.

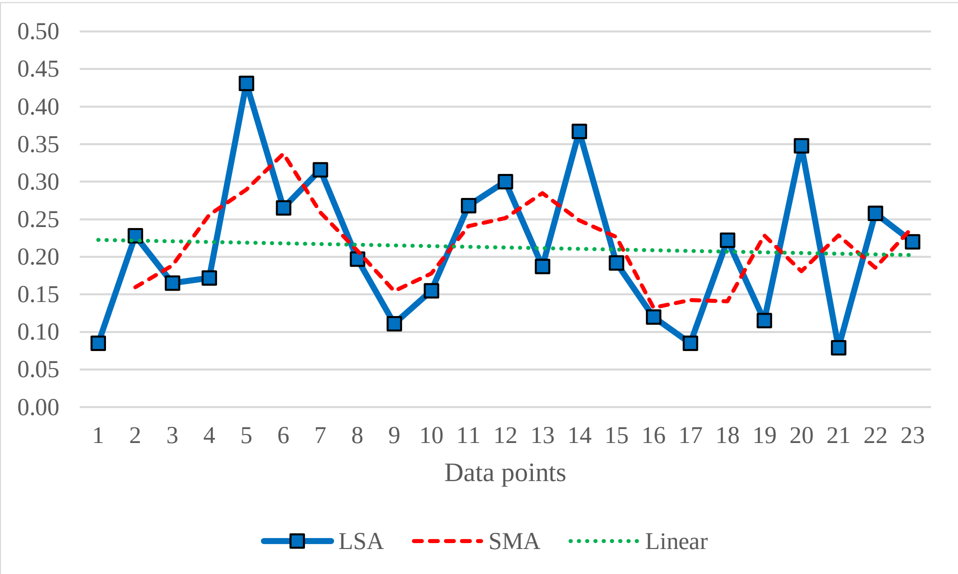
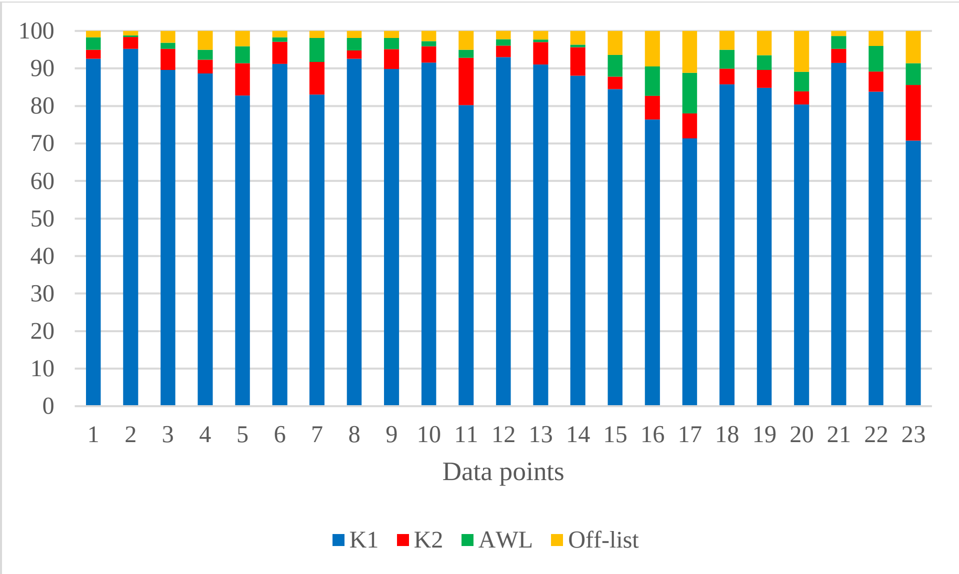


Figure 5.5 Lexical disparity (LSA)

The higher the LSA index value, the more cohesive the text is (Crossley & McNamara, 2013). The LSA index (Figure 5.5) fluctuated wildly over the course of the investigation without a discernible pattern. The highest value was measured at data point 5 (LSA = 0.43), while the lowest value was gauged at data point 21 (LSA = 0.08). The smoothing algorithm of the lexical disparity (LSA) index displays a sideways (horizontal) trend over the nine-month investigation.

However, the linear trend line of the LSA index shows a slightly downward trend over the nine months.

To investigate the extent of genre-relevant lexical choice in Dalma's written data, the percentage of academic words in her essays was estimated by the Academic Word List (AWL) measure (Figure 5.6).



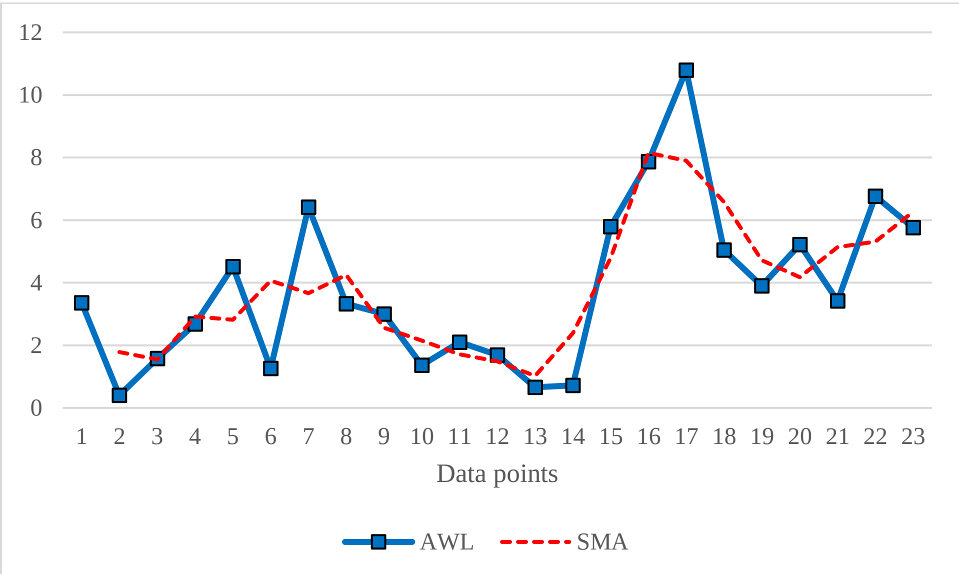


Figure 5.6 Lexical sophistication (top), and Academic Word List (bottom)

The higher the AWL index, the more words from the Academic Word List in a specific text are used. Figure 5.6 (top) shows the four frequency bands: K1 is the list of the most frequent 1000-

word families, K2 is the list of the second most frequent 1000 word families, the Academic Word List, and off-list words. Figure 5.6 (top) shows that the highest percentage of K1 words was measured at data point 2 (K1 words = 95.24%), while the lowest percentage of K1 words was detected at data point 17 (K1 words = 71.37%). The low percentage of K1 words at data point 17 was attained by the increase of the percentages of the Academic Word List and off-list words. Indeed, both the percentages of the Academic Word List and the off-list words were the highest at data point 17. In order to gain a better understanding of the use of the words from the Academic Word List, a separate graph was plotted in Figure 5.6 (bottom). The AWL index shows fluctuations over the nine months. The lowest AWL index was measured at data point 2 (AWL = 0.4) which suggests that Essay 2 contained the lowest percentage of words from the Academic Word List. Conversely, the highest AWL index was measured at data point 17 (AWL = 10.79) which implied that Essay 17 included the highest percentage of words from the Academic Word List. The smoothing algorithm of the Academic Word List (AWL) index displays an upward trend over the nine-month investigation.

5.2.2 Syntactic Complexity

Figure 5.7 displays the five syntactic indices used in this study. It can be observed that the FVR, CP/C and the CN/C indices showed similar trajectories. This finding was not surprising since all three indices partially tap into similar sub-constructs of syntax, phrasal complexity. Figure 5.7 also displays the DC/C index as showing an inverse trend from the trends of the FVR, CP/C and CN/C indices. This result suggested that Dalma made her writing more complex by employing more noun phrases at the expense of employing more dependent clauses.

General syntactic complexity was measured by the finite verb ratio (FVR) in Dalma's written data (Figure 5.8). The higher the FVR index, the more words divided by the finite verbs are. A high FVR index implies a higher level of internal sentence complexity (Verspoor et al., 2017).

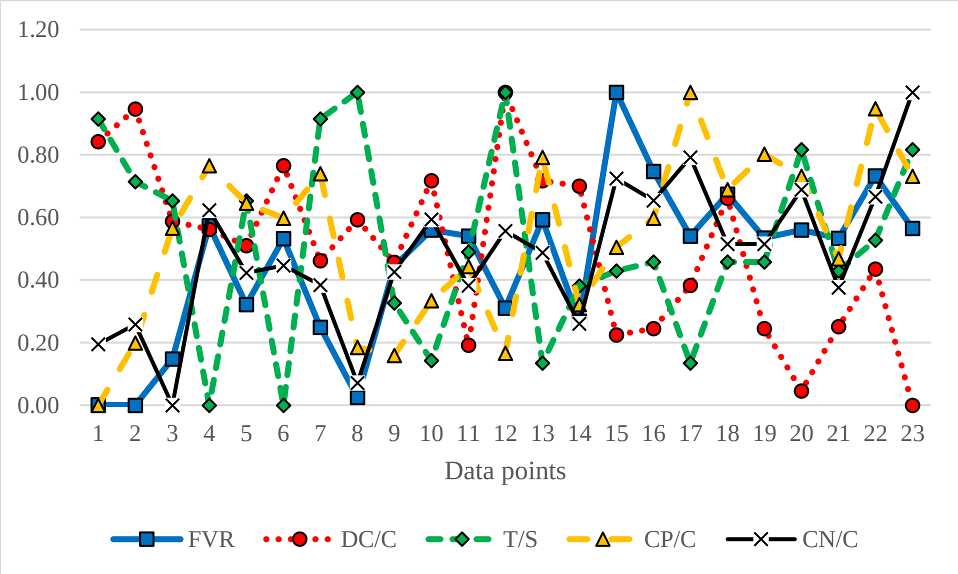


Figure 5.7 Syntactic complexity indices

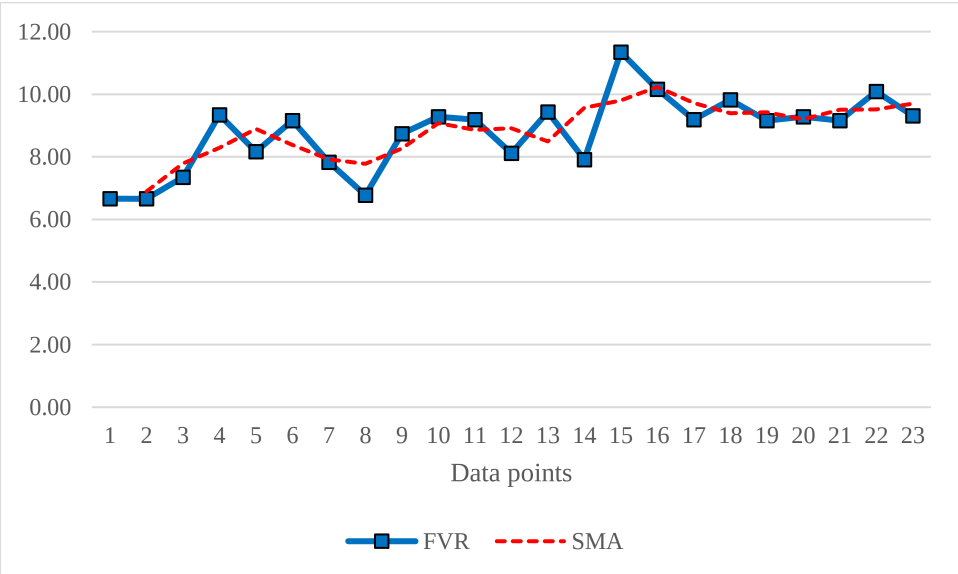


Figure 5.8 General syntactic complexity (FVR)

The FVR index, shown in Figure 5.8, started from its lowest point (FVR = 6.67) at data point 1 and increased gradually until data point 4. Between data points 4 and 14 the FVR index fluctuated between 8 and 10 FVR index value, except for data points 8 and 14. However, the FVR index increased dramatically at data point 15 and reached its zenith (FVR = 11.34). The

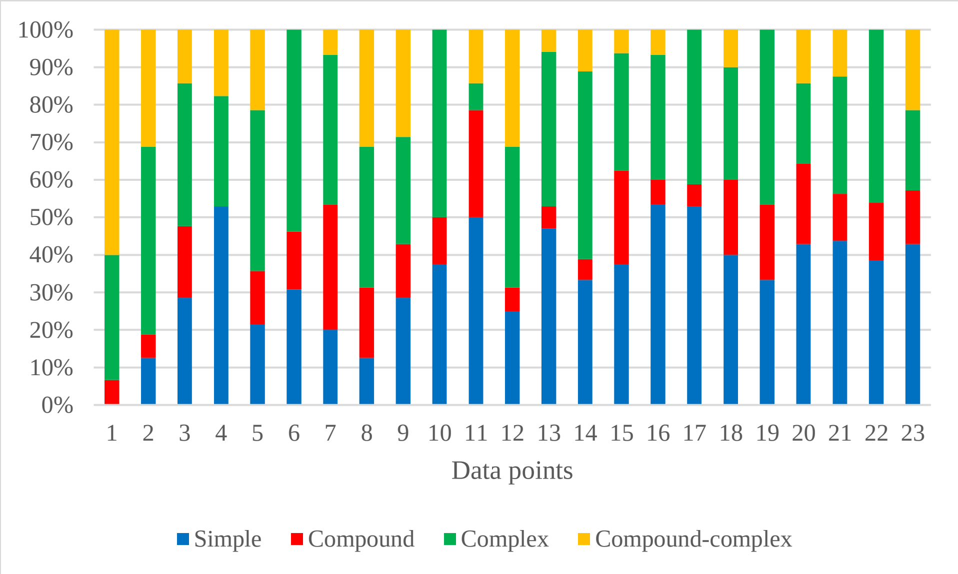
FVR index remained relatively high between data points 15 and 23. The smoothing algorithm of the general syntactic complexity (FVR) index displays an upward trend over the nine-month investigation. The following examples, taken from Dalma's Essay 1 and Essay 15, illustrate the differences in the general measure of syntactic complexity. Essay excerpt 5.1 was extracted from Essay 1 when the FVR was the lowest (6.67). In contrast, Essay excerpt 5.2 was taken from Essay 15 when the FVR index was the highest (11.34). The finite verbs are underlined, and the example sentences are left uncorrected.

Essay excerpt 5.1 (Essay 1)
People also [1]<u>could</u> spend more time with their hobby which [2]<u>was</u> also good.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 13; finite verb count = 2; FVR = 6.5)

Essay excerpt 5.2 (Essay 15)
On the other hand, nowadays most of ventures and companies [1]<u>insist</u> on minimum an intermediate exam, when people [2]<u>would</u> like to apply to the company.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 25; finite verb count = 2; FVR = 12.5)

Although Essay excerpt 5.1 and Essay excerpt 5.2 had the same number of finite verbs and were the same type of sentence (complex), they differed in their level of syntactic complexity. Essay excerpt 5.2 was undoubtedly more complex than Essay excerpt 5.1 since the FVR index of Essay excerpt 5.2 (12.5) was higher than the FVR index of Essay excerpt 5.1 (6.5).

Dalma's written samples were also coded for sentence types in order to gain a better insight into the syntactic structures she used in her written data. The distribution of the different sentence types is presented in a bar chart in Figure 5.9. Figure 5.9 (top) shows the distribution of sentence types in Dalma's written data, while Figure 5.9 (bottom) shows the linear trend lines of the sentence types. Visual inspection of Figure 5.9 (top) indicated that simple (Si) and complex (Cx) sentences evenly dominated in Dalma's written samples, especially in Essay 9. Nevertheless, Figure 5.9 (bottom) also shows that the proportion of simple sentences (Si) increases over time, whereas the proportion of compound (Co) sentences stagnated over the nine-month investigation.



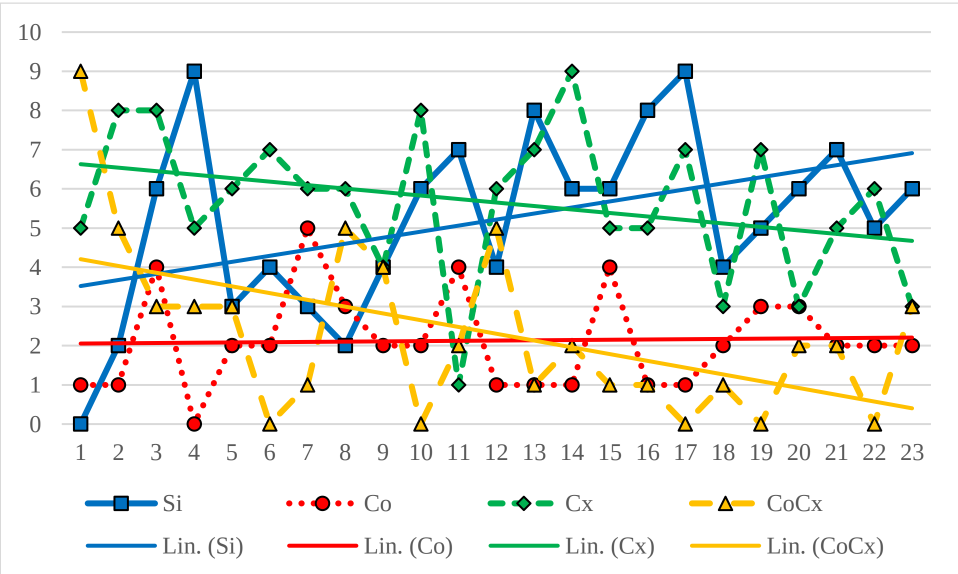


Figure 5.9 Sentence types (top: distribution; bottom: trend)

Figure 5.9 (bottom) also shows that both the proportion of complex (Cx) and compound-complex (CoCx) sentences decreased over the nine-month investigation. Therefore, it can be presumed that the proportion of simple (Si) sentences increased at the expense of the decrease of the proportion of complex (Cx) and compound-complex (CoCx) sentences. The increase in the proportion of simple (Si) sentences was not surprising since the FVR index was relatively high between data points 15 and 23. It might be speculated that Dalma made her writing more complex at the phrasal level rather than at the clausal level in the last third of the investigation. This phenomenon is in line with current findings of academic writing which suggest that advanced users use more complex nominalisations than clausal embeddings in academic writing (Biber & Gray, 2010). Compound (Co) sentences were present in every essay except for Essay 4, and their proportion was relatively low compared to simple (Si) and complex (Cx) sentences. Dalma also employed compound-complex (CoCx) sentences in her written data in almost every essay. However, the proportion of compound-complex (CoCx) sentences gradually decreased over time.

Dalma made her writing more elaborate by employing different syntactic structures such as subordination, coordination, and phrasal complexity. Subordination was gauged in this study by the ratio of dependent clauses to clauses (DC/C) index. Figure 5.10 shows that the DC/C index fluctuated slightly over the course of investigation, ranging between 0.2 and 0.4 with the exception of Essay 1, Essay 2, and Essay 12.

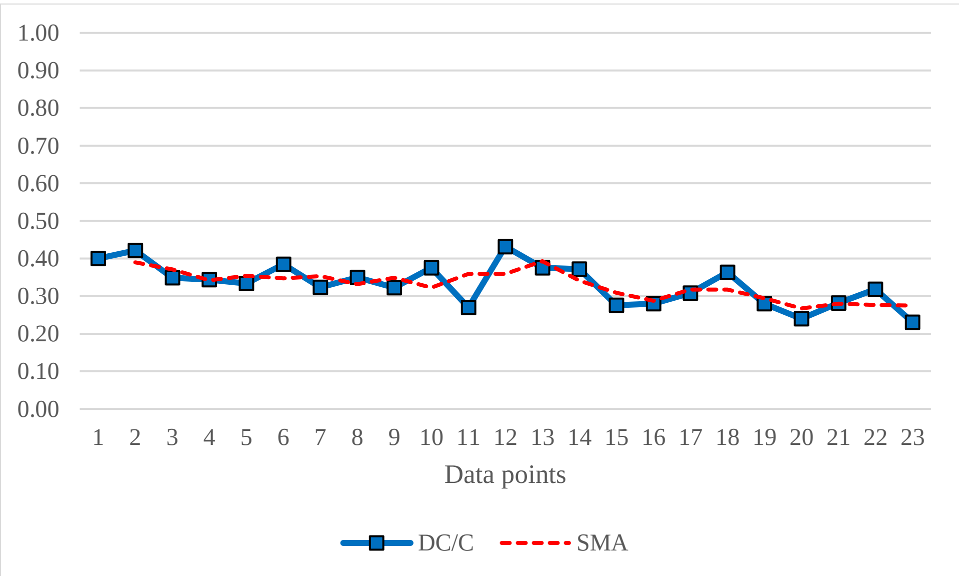


Figure 5.10 Subordination (DC/C)

The smoothing algorithm of the subordination (DC/C) index displays a downward trend over the nine-month investigation. The highest value was measured at data point 12 (DC/C = 0.43). As illustrated in Essay excerpt 5.3, this high value means that approximately every second clause was a dependent clause.

Essay excerpt 5.3 (Essay 12)

All in all, I can imagine the fact [1]<u>that, people can learn a foreign language without visiting the countries</u> [2]<u>where the language is spoken.</u>
(Errors are uncorrected; dependent clauses are underlined and numbered in superscript)
(Dependent clause count = 2; clause count = 3; DC/C = 0.66)

Besides subordination, Dalma also employed both clausal and phrasal coordination structures in her essays (Figure 5.11). The clausal coordination (T/S) index, shown in Figure 5.11, displays slight fluctuations from data point 4 to 13. The T/S index reached its zenith at data point 8 and 12 (1.44) which is to say that there were approximately one and half T-units in every sentence. The T/S index was also relatively high at data points 1, 2, 3, 5, 8, and 12. This finding was anticipated since the proportion of the compound-complex (CoCx) sentences was relatively high in these essays (Figure 5.9).

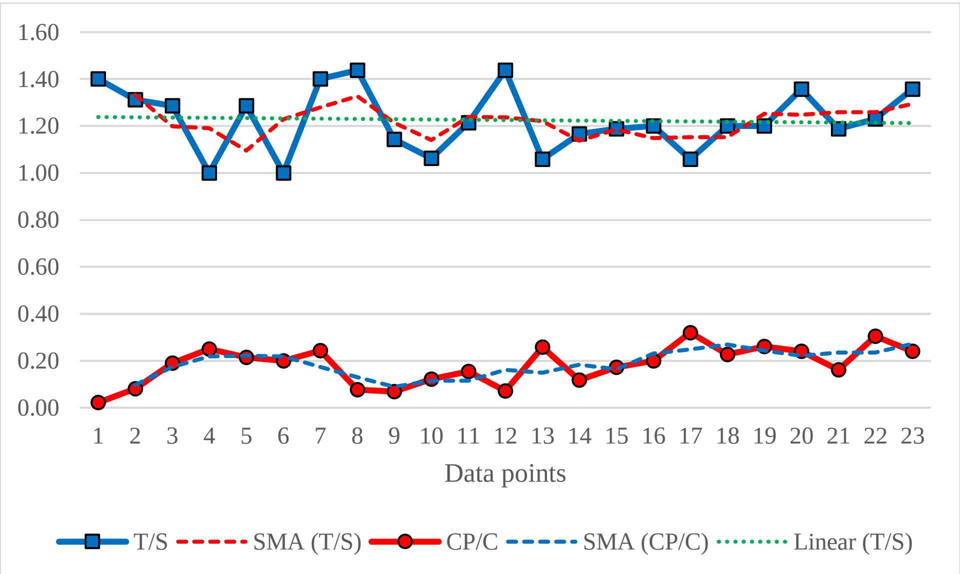


Figure 5.11 Coordination (T/S and CP/C)

The smoothing algorithm of the clausal coordination (T/S) index (red dashed line) displays a sideways trend over the nine-month investigation. However, the linear trend line of the T/S index (green dotted line) shows a slightly downward trend over the nine months.

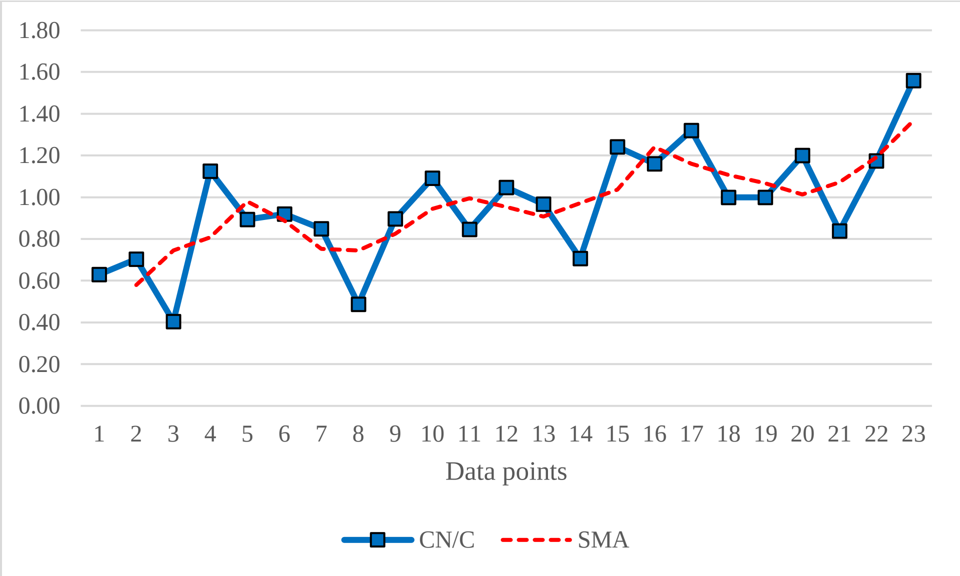
In addition to clausal coordination, Dalma also employed phrasal coordination structures to make her sentences longer and consequently her writing more complex. Figure 5.11 shows that the CP/C index reached its zenith at data point 17 (CP/C = 0.32). This finding suggested that on average one coordinate phrase was used for every three clauses in Essay 17, as illustrated in Essay excerpt 5.4.

Essay excerpt 5.4 (Essay 17)

People are constantly obsessed with their [1]mobile phones, televisions, electronic gadgets and devices. There should be the right balance between [2]the futuristic world and the old-fashioned way of life.
(Errors are uncorrected; coordinate phrases are underlined and numbered in superscript)
(Coordinate phrase count = 2; clause count = 2; CP/C = 1)

The smoothing algorithm of the phrasal coordination (CP/C) index displays a slightly upward trend over the nine-month investigation.

Phrasal complexity was measured by the complex nominals per clause (CN/C) and the mean number of modifiers per noun phrases (SYNNP) indices. Figure 5.12 (top) shows that the CN/C index fluctuated over time. The lowest CN/C index value was detected at data point 3 (CN/C = 0.40), while the highest CN/C index value was measured at data point 23 (CN/C = 1.56). The smoothing algorithm of the phrasal complexity (CN/C) index displays an upward trend over the nine-month investigation. Figure 5.12 (bottom) shows that the lowest SYNNP value (SYNNP=0.40) was measured at data point 8, while SYNNP index reached its highest point (1.07) at data point 20. The smoothing algorithm of the SYNNP index displays an upward trend over the nine-month investigation.



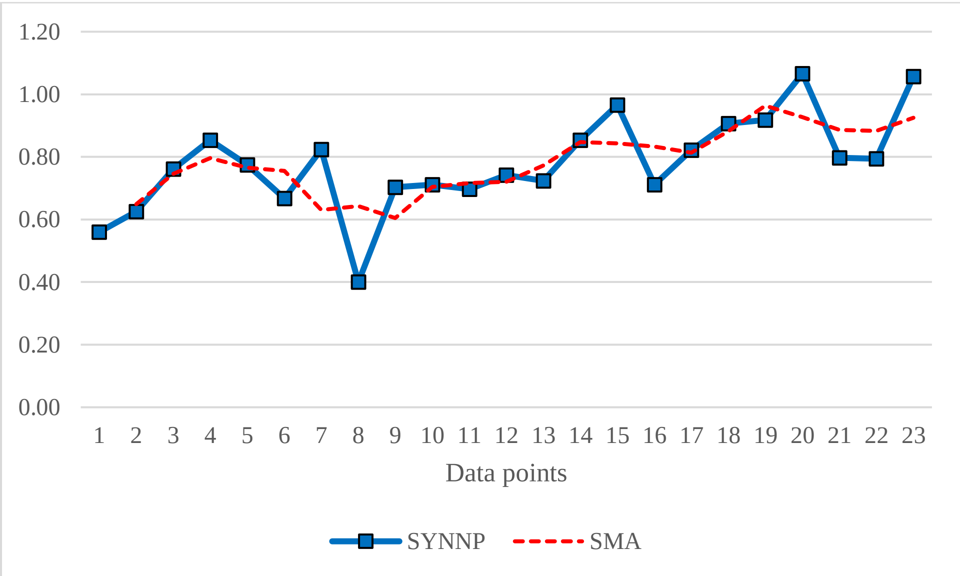


Figure 5.12 Phrasal complexity (CN/C and SYNNP)

Essay excerpt 5.5 exemplifies the high CN/C index value (CN/C = 1.56) at data point 23.

Essay excerpt 5.5 (Essay 23)

The way I see it, [1]popular events like the [2]football World Cup and [3]other international sporting occasions are crucially important in easing [4]international tensions and releasing [5]patriotic emotions in a [6]safe way.
(Errors are uncorrected; complex nouns are underlined and numbered in superscript)
(Complex nominal count = 6; clause count = 2; CN/C = 3)

Figure 5.13 shows the syntactic indices specific to the academic genre: the normed rate of occurrence of conditional clauses (ConC), the normed rate of occurrence of infinitive clauses (InfC), the normed rate of occurrence of relative clauses (RelC), and the incidence score of prepositional phrases (DRPP) indices.

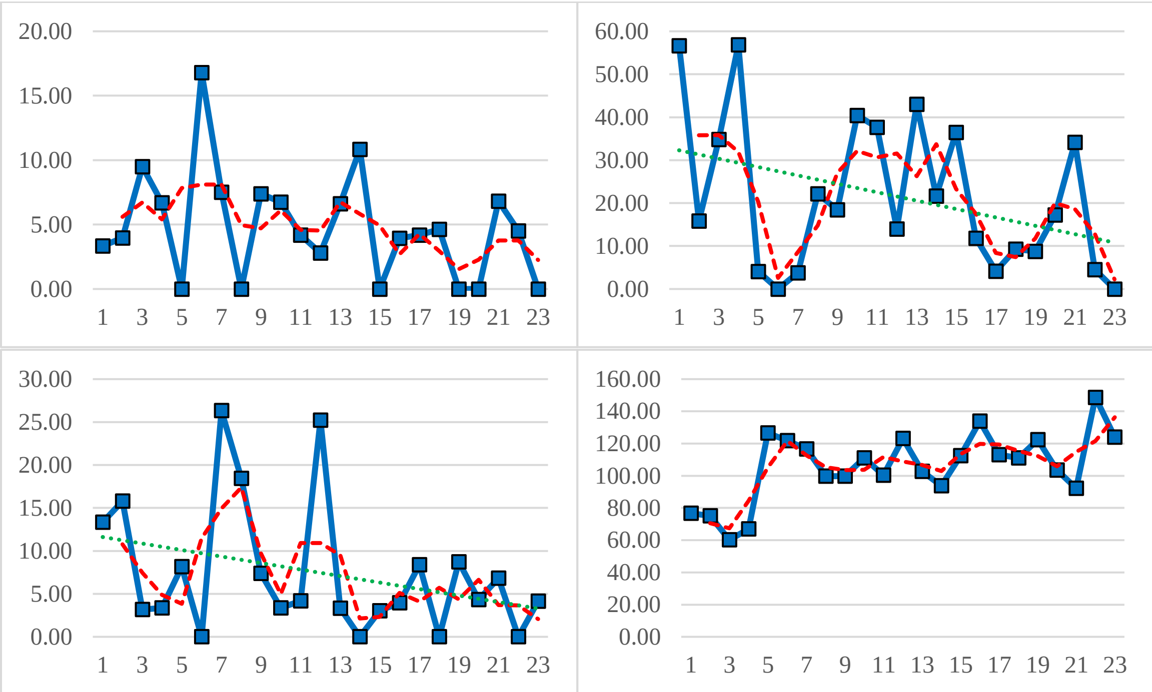


Figure 5.13 Syntactic measures specific to the academic genre (top left: conditional clauses (ConC); top right: infinitive clauses (InfC); bottom left: relative clauses (RelC); bottom right: prepositional phrases (DRPP).

The smoothing algorithm of the prepositional phrases (DRPP) index show a clear upward trend, whereas the smoothing algorithm of the ConC index displays a downward trend over the nine-month investigation. The smoothing algorithms of the InfC and the RelC indices display sideways trends over the nine-month investigation. However, the linear trend lines of the infinitive and the relative clauses indices show clear downward trends over the nine months. These changes in the genre-specific syntactic indices indicate that Dalma tended to use more prepositional phrases in her essays. Conversely, a decrease can be observed in the usage of conditional, infinitive and relative clauses over the nine-month investigation.

5.2.3. Accuracy

In addition to the constructs of syntactic and lexical complexity, accuracy was also measured manually in Dalma’s written data by the ratio of error-free clauses to the total number of clauses (EFC/C) index. The higher the EFC/C index, the more error-free clauses are in a text. Consequently, a value of 1 means that the text contains no errors. As shown in Figure 5.14, the EFC/C index decreased between data points 1 and 5, but then from data point 5, it increased gradually to data point 9. However, from data point 9 the EFC/C index decreased until the end of the investigation with small fluctuations at the end of the data collection.

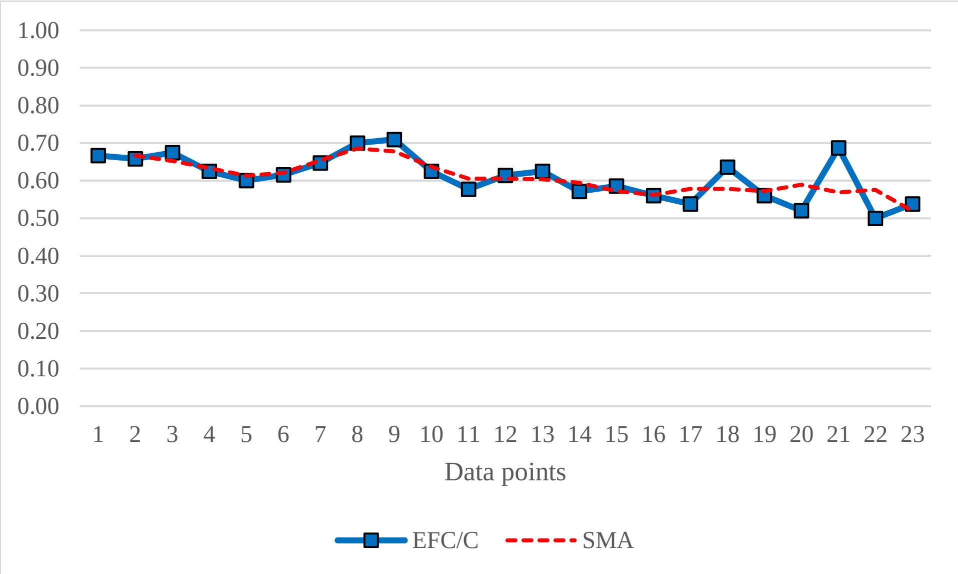


Figure 5.14 Accuracy

The smoothing algorithm of the accuracy (EFC/C) index displays a clear downward trend over the nine-month investigation. The highest EFC/C value (EFC/C = 0.71) was detected at data point 9, while the lowest EFC/C value (EFC/C = 0.50) was measured at data point 22. The lowest EFC/C value meant that there was at least one error in every second clause in Essay 22. Essay excerpt 5.6 illustrates the low accuracy rate in Essay 22.

Essay excerpt 5.6 (Essay 22)

If universities selected their students based on sex, the quality of the education would [1]detereirate significantly.
(Errors are uncorrected, underlined and numbered in superscript)
(Error-free clause count = 1; clause count = 2; EFC/C = 0.5)

In order to gain a deeper insight into the accuracy construct of Dalma's written data, this study also looked at the type of errors she made in her essays. The errors were coded and categorised according to the Louvain Error Tagset, and the results are displayed in Figure 5.15.

Figure 5.15 shows that the number of errors slightly decreased from data point 3 to 9, but then the number of errors started to fluctuate between data points 10 and 12. From data point 13 to 23, almost an even number of errors was detected in Dalma's written data.

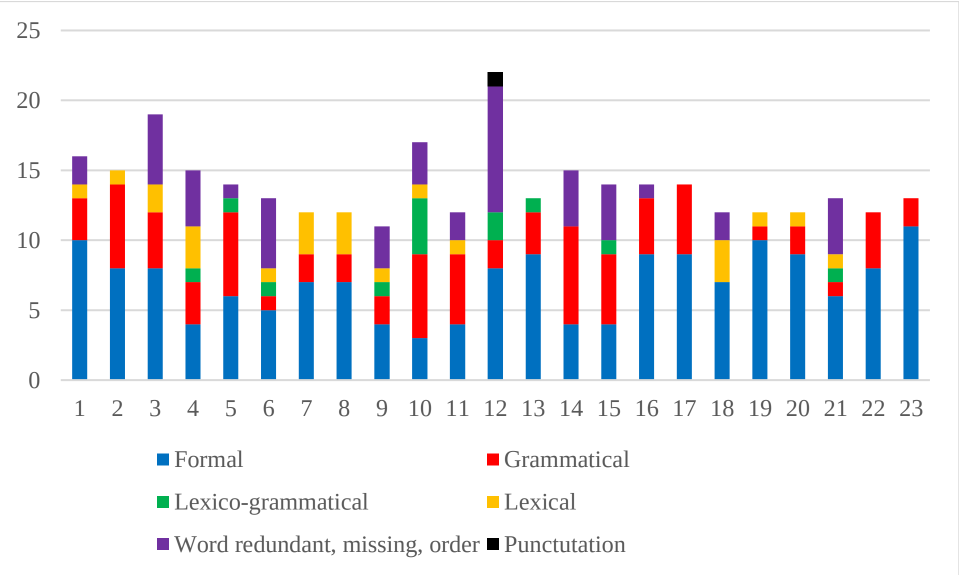


Figure 5.15 Error type

Figure 5.15 shows that Dalma made six different types of errors in her essays: (1) formal, (2) grammatical, (3) lexico-grammatical, (4) lexical, and (5) word redundant, word missing and word order, and (6) punctuation. Among the six different types of errors, formal errors (morphology and spelling) were the most prevalent (50%) in Dalma's written data. In addition, formal errors were the only type of errors which were present in every essay. The highest proportion of formal errors was measured in Essay 23 (85%). The second most common type of errors included grammatical errors (24%) which were present in every essay except for Essay 18. The third most common type of error (14%) belonged to the word redundant, word missing, and word order category.

5.3 The Degree of Variability

Section 5.3 explores the degree of variability in the lexical and syntactic complexity and accuracy indices in Dalma's written data. First, the degree of variability was visualised by a descriptive technique, moving min-max graphs (van Geert & van Dijk, 2002). Second, developmental peaks were detected in the lexical and syntactic complexity and accuracy indices by data resampling and Monte Carlo analyses. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity and (3) accuracy.

5.3.1 Lexical Complexity

The degree of variability was measured in (1) the general lexical complexity (AveWL), (2) the lexical variability (TTR), (3) the lexical rarity (WRDFRQc), (4) the lexical disparity (LSA), and (5) the Academic Word List (AWL) indices. Among the five lexical complexity indices, the lexical variability (TTR) and the Academic Word List (AWL) indices displayed the highest degree of variability over the nine months.

In the min-max graph of the lexical variability (TTR) index (Figure 5.16) we can observe three main stages: (1) a rather wide bandwidth of variability up between data points 1 and 13 (from 0.4 to 0.7), (2) a much wider bandwidth of variability between data points 14 and 19 (with a jump from 0.52 to 0.88) and then (3) a rather narrow bandwidth (0.72 to 0.76) between data points 20 and 23. In other words, lexical items were varied randomly ("free variability") between data points 1 and 13 in Dalma's written data. Between data points 14 and

17, lexical items were varied to a great extent (“overgeneralization”). Finally, lexical items were varied in a regular pattern (in a target-line manner) from data point 18 in Dalma’s written data.

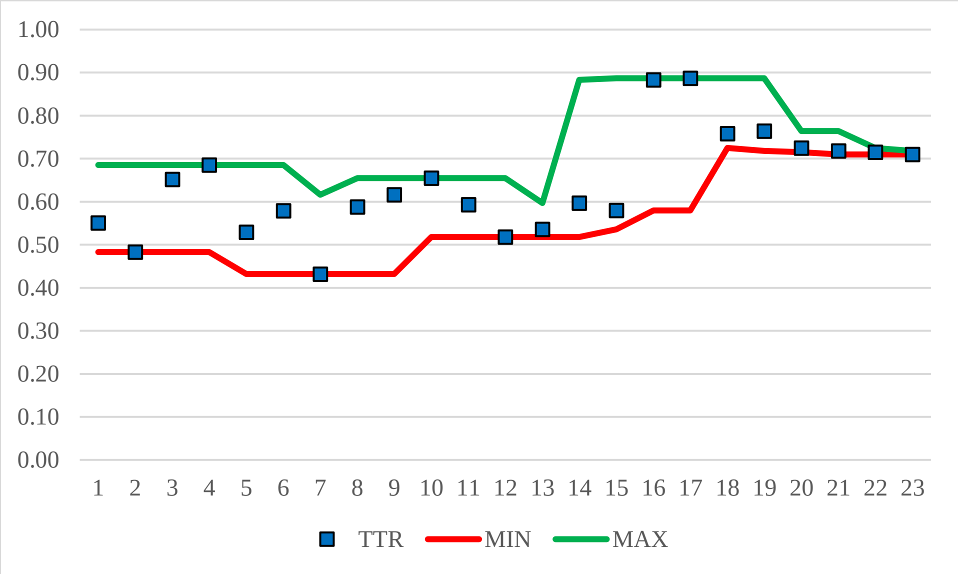


Figure 5.16 Lexical variability (min-max graph)

In the min-max graph of Academic Word List (Figure 5.17) we can see three main stages: (1) a rather wide bandwidth of variability up to data point 11 (from 0.4% to 6.4%), (2) a much wider bandwidth of variability between data points 12 and 19 (with a jump from 0.66% to 10.79%) and then (3) a rather narrow bandwidth (from 3.42% to 6.76%) between data points 20 and 23. In other words, Dalma used lexical items from the Academic Word List randomly (“free variability”) between data points 1 and 11. She overused the words from the AWL (“overgeneralization”) between data points 11 and 19. Finally, lexical items were used from the AWL in a more regular pattern between data points 20 and 23 (in a target-like manner).

The data of the lexical variability (TTR) and the Academic Word List (AWL) indices were resampled and two separate Monte Carlo analyses were run. Table 5.2 shows that the p-value of the lexical variability (TTR) index was 0.04 and the p-value of the Academic Word List (AWL) index was 0.02. In other words, statistically significant developmental peaks were detected in the TTR and the AWL indices at around data point 16.

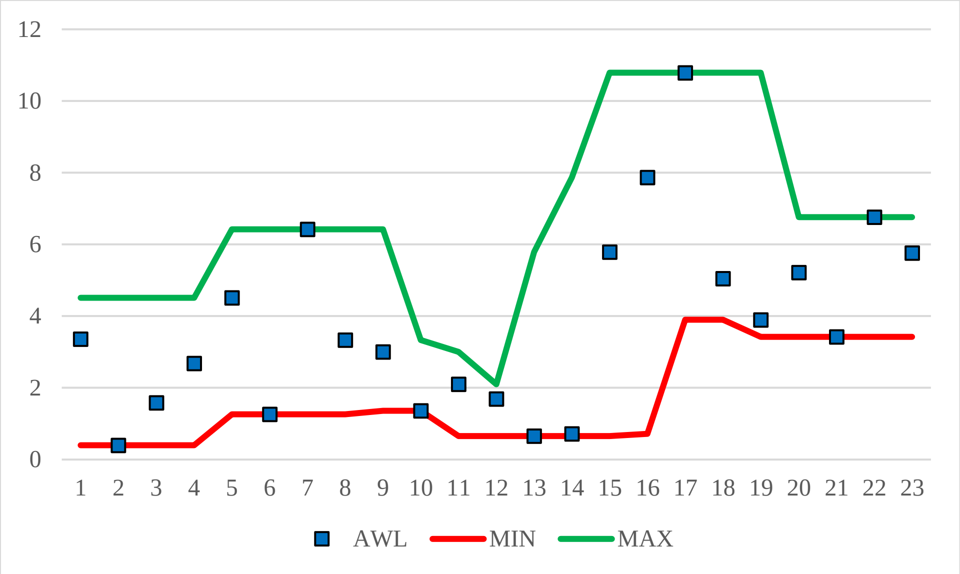


Figure 5.17 Academic Word List (min-max graph)

The same statistical procedures were repeated for the general lexical complexity (AveWL), the lexical rarity (WRDFRQc) and the lexical disparity (LSA) indices. However, the p-values of the AveWL, the WRDFRQc and LSA indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the AveWL, the WRDFRQc, and the LSA indices in Dalma's written data.

Table 5.2

P-values for the Complexity and Accuracy Indices

Index	P-value
AveWL	0.17
TTR	0.04*
WRDFRQc	0.85
LSA	0.61
AWL	0.02*
FVR	0.74
DC/C	0.92
T/S	0.53
CP/C	0.36
CN/C	0.90
EFC/C	0.83

Note. *Statistically significant at 0.05 level.

5.3.2 Syntactic Complexity

The degree of variability was measured in: (1) the general syntactic complexity (FVR), (2) the subordination (DC/C), (3) the clausal coordination (T/S), (4) the phrasal coordination (CP/C), and (5) the phrasal complexity (CN/N) indices. Among the five syntactic complexity indices, the phrasal coordination (CP/C) index demonstrated the highest degree of variability.

In the min-max graph of phrasal coordination (Figure 5.18) we can see some variability. However, the difference in the bandwidth was not as great as in the two lexical indices (TTR and AWL). Nevertheless, the band moved up three times over the nine months. The first move can be observed at data point 6, while the second move occurred at data point 15.

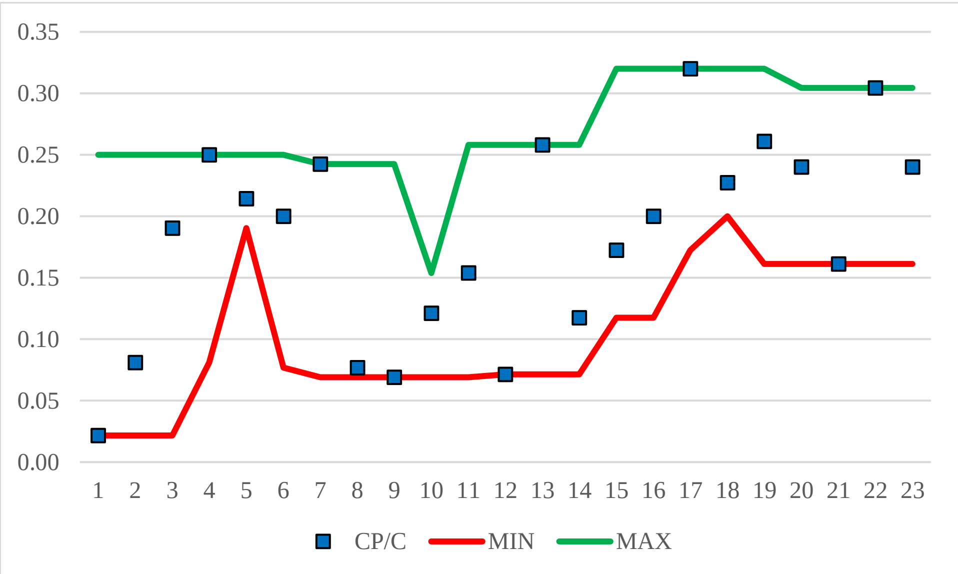


Figure 5.18 Phrasal coordination (min-max graph)

Table 5.2 shows that the p-values of the five syntactic complexity indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the five syntactic complexity indices in Dalma’s written data.

5.3.3 Accuracy

The degree of variability was measured in the error-free clause ratio (EFC/C) index. In the min-max graph (Figure 5.19) of EFC/C index we can observe some variability. However, there were no large differences in the bandwidth over the nine months. The band was quite stable between data points 1 and 12, but then the band itself moved down to data point 23. In addition, the bandwidth became wider from data point 19.

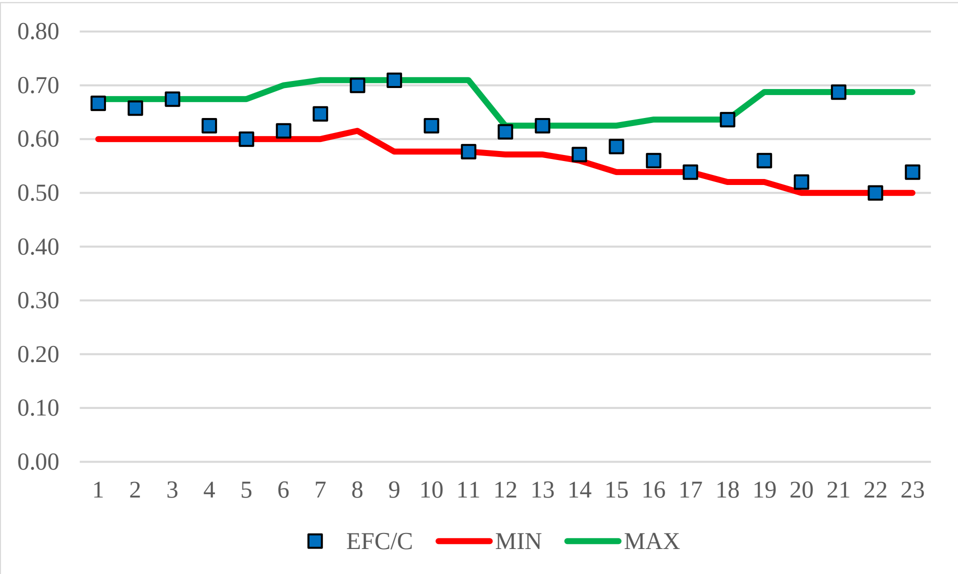


Figure 5.19 Accuracy (min-max graph)

These ranges of variability can be interpreted as follows. Error-free clauses occurred in a more regular pattern between data points 1 and 18, while error-free clauses occurred randomly in Dalma's written data from data point 19. Table 5.2 shows that the p-value of the accuracy (EFC/C) index was 0.83. In other words, statistically significant developmental peak was not detected in the accuracy (EFC/C) index in Dalma's written data.

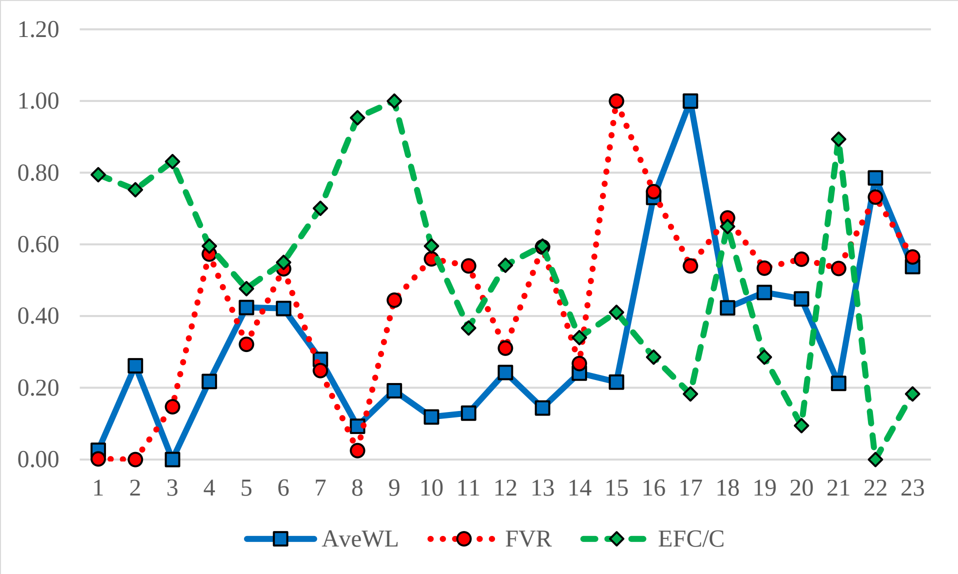
5.4 Interactions

The purpose of this section is to explore the dynamic interactions between lexical and syntactic complexity and accuracy. This section is subdivided into two parts: (1) interactions between general lexical and syntactic complexity and accuracy indices and (2) interactions between lexical variability, subordination, and accuracy indices.

5.4.1 Interactions Between General Lexical and Syntactic Complexity and Accuracy

Figure 5.20 shows (top) the normalised data of the general lexical complexity (AveWL) and general syntactic complexity (FVR) and the accuracy (EFC/C) indices. The trajectories of the general lexical complexity (AveWL) and the general syntactic complexity (FVR) indices exhibit both alternating and parallel patterns. However, between data points 6 to 9 and between 20 and 23, there were clear parallel patterns. The growth trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) indices show mainly alternating shifts for the entire period of investigation. The growth trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) indices demonstrate both parallel and alternating shifts.

The residual plot in Figure 5.20 (bottom) shows that the trajectories of the general lexical and syntactic complexity indices were both parallel and alternating. The clear parallel period between data points 6 and 9 and between 20 and 23 was also confirmed by the residual plot. Figure 5.20 (bottom) also shows that the trajectories of the AveWL and the EFC/C indices were predominantly alternating. In these shifts, a movement above the trend in the AveWL index was accompanied by a movement below it in the EFC/C index and vice versa, for example between data points 3 and 4 and between data points 17 and 18. The residual plot (Figure 5.20) also shows both parallel and alternating variability patterns for the trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) indices. However, there were clear parallel shifts between data points 1 and 3 and between data points 15 and 19.



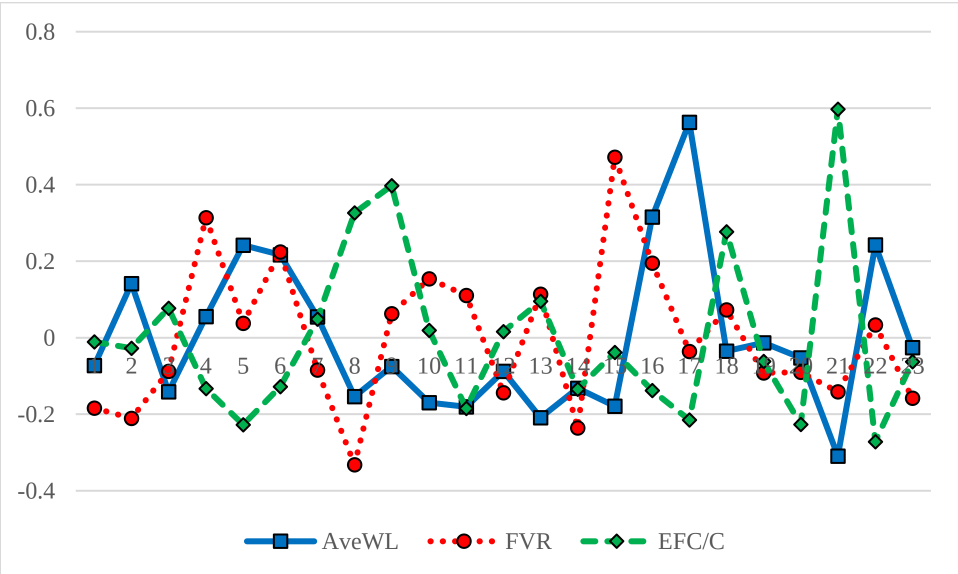


Figure 5.20 General lexical (AveWL) and syntactic complexity (FVR) and accuracy (EFC/C) (top: normalised data; bottom: residual plot)

Figure 5.21 shows the interactions between general lexical (AveWL) and syntactic complexity (FVR) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 5.21 shows that the general lexical complexity (AveWL)-general syntactic complexity (FVR) correlation shifted between moderately positive and moderately negative values, with one peak towards a strongly positive value in the 21st window. The AveWL-EFC/C correlation was, for the most part, negative with strongly negative values between windows 3 and 7. However, the AveWL-EFC/C correlation was positive between windows 8 and 9. The FVR-EFC/C correlation alternated between moderately negative and moderately positive values, with one peak towards a strongly positive value in the 12th window and with one peak towards a strongly negative value in the 21st window.

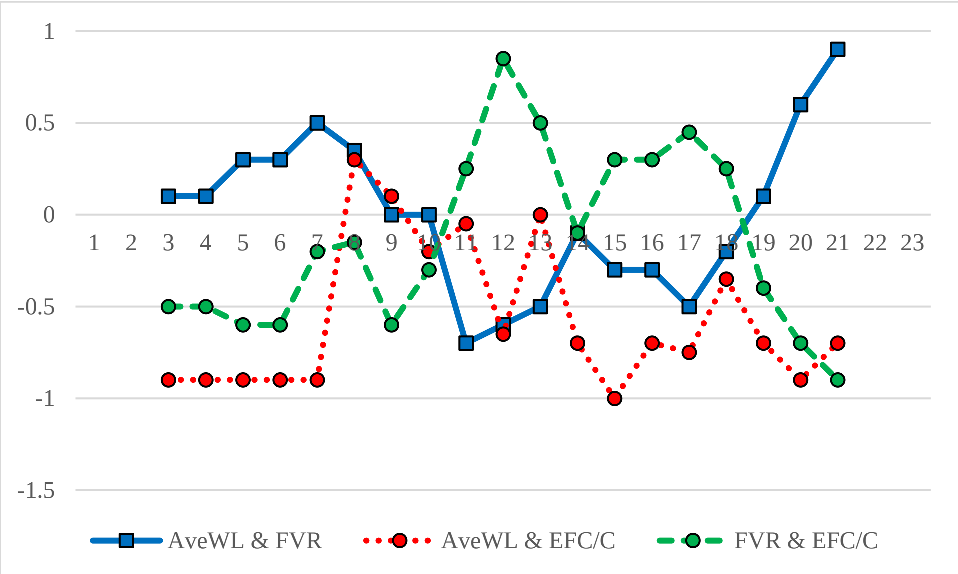


Figure 5.21 Moving correlation plot (General lexical and syntactic complexity and accuracy)

Table 5.3 shows the Spearman's rank correlation coefficient values between the general lexical and syntactic complexity and accuracy indices.

Table 5.3

Correlation Matrix (General Lexical and Syntactic Complexity and Accuracy)

		AveWL	FVR	EFC/C
AveWL	ρ	1.000	.402	-.738**
	Sig. (2-tailed)		.058	.000
	N	23	23	23
FVR	ρ	.402	1.000	-.539**
	Sig. (2-tailed)	.058		.008
	N	23	23	23
EFC/C	ρ	-.738**	-.539**	1.000
	Sig. (2-tailed)	.000	.008	
	N	23	23	23

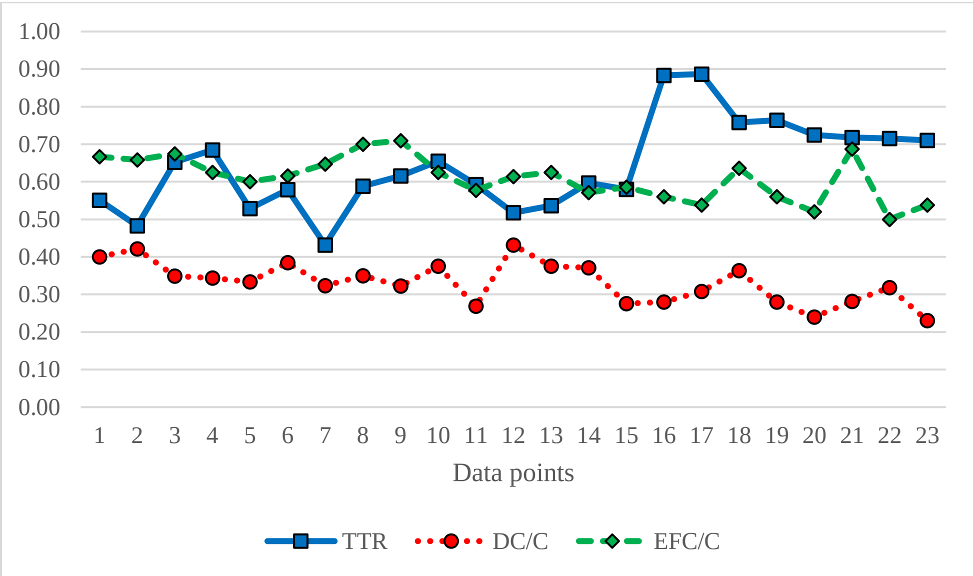
Note. ** Correlation was significant at the 0.01 level (2-tailed).

The correlation between the general lexical complexity (AveWL) and the general syntactic complexity (FVR) was positive but not significant. Conversely, the correlation between the general lexical complexity (AveWL) and accuracy (EFC/C) was highly negative and statistically significant ($p<0.01$). Thus, the mainly alternating variability patterns (in Figure 5.20) were confirmed by a strongly negative correlation coefficient. The correlation between the general syntactic complexity (FVR) and accuracy (EFC/C) was also negative and statistically significant ($p<0.01$).

5.4.2 Interactions Between Lexical Variability, Subordination, and Accuracy

Figure 5.22 (top) shows the growth trajectories of lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices. Figure 5.22 shows that the trajectories of lexical variability (TTR) and subordination (DC/C) were both alternating and parallel. There were periods with clear parallel shifts, for example between data points 4 and 8. Conversely, there were periods with clear alternating shifts, for example between data points 11 and 14. The growth trajectories of lexical variability (TTR) and accuracy (EFC/C) indices demonstrate both parallel and alternating variability patterns. There were periods with clear alternating shifts for example between data points 15 and 18. The growth trajectories of subordination (DC/C) and accuracy (EFC/C) indices were predominantly parallel over the nine months.

Figure 5.22 (bottom) shows the residual plot of the lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices. The trajectories of lexical variability (TTR) and subordination (DC/C) indices display both parallel and alternating growth patterns. The clear parallel growth patterns of the TTR and the DC/C indices between data points 4 and 8 were also confirmed by the residual plot. Likewise, the clear alternating growth patterns of the TTR and the DC/C indices were also corroborated by the residual plot. The trajectories of the lexical variability (TTR) and accuracy (EFC/C) also show both alternating and parallel variability patterns. However, the clear alternating patterns in the TTR and the EFC/C indices observed in the data plot between data points 15 and 18 was not confirmed by the residual plot since during this period the residual plot shows a parallel shift between data points 16 and 17. The residual plot shows both alternating and parallel growth patterns for the trajectories of subordination (DC/C) and accuracy (EFC/C) indices. However, there were clear parallel growth patterns (for example between data points 17 and 21).



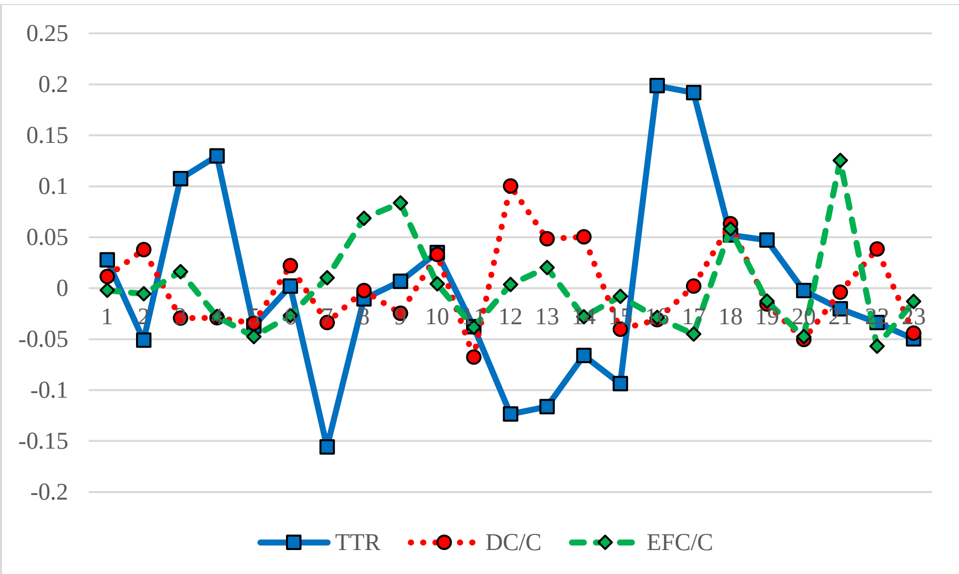


Figure 5.22 Data and residual plot (Lexical variability, subordination & accuracy)

Figure 5.23 shows the interactions between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 5.23 shows that the lexical variability (TTR)-subordination (DC/C) correlation shifted between moderately negative and moderately positive values, with two peaks towards strongly negative values in the 13th and 14th windows. The lexical variability (TTR)-accuracy (EFC/C) correlation varied between moderately positive and strongly negative values. Conversely, the subordination (DC/C)-accuracy (EFC/C) was, for the most part, positive over the entire period of investigation with one peak towards a moderately negative value in the 7th window. There were only three windows (7, 8, and 21) in which the subordination (DC/C)-accuracy (EFC/C) correlation coefficients were negative. The visual inspection of the data, the residual, and the moving correlation plots might indicate that there was a moderately competitive relationship between the TTR and the DC/C indices.

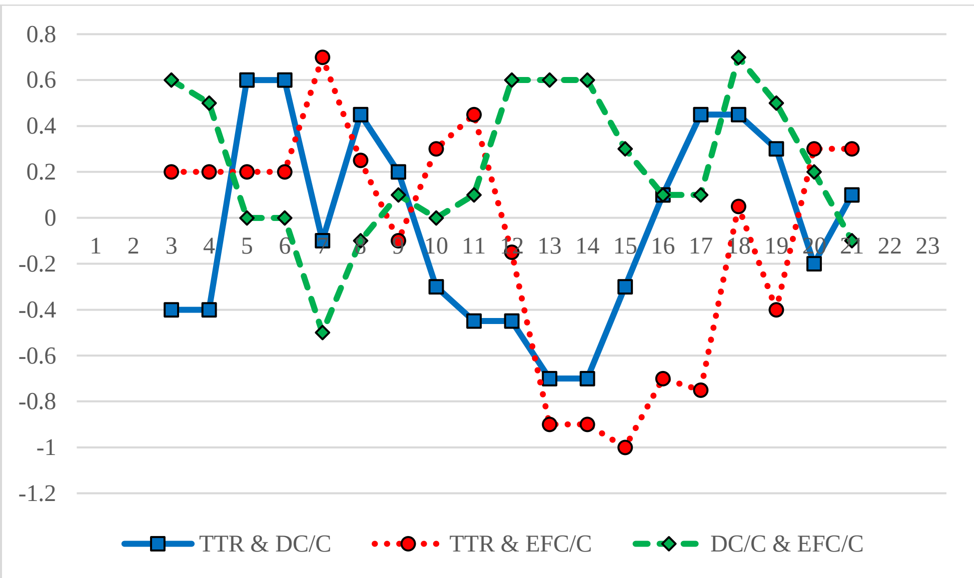


Figure 5.23 Moving correlation (Lexical variability, subordination, and accuracy)

Table 5.4 shows the correlation coefficient values between the lexical variability (TTR), subordination (DC/C) and the accuracy (EFC/C) indices. The Spearman's rank correlation coefficient between the lexical variability (TTR) and subordination (DC/C) was moderately negative (-.562) and statistically significant (*p*<0.01). Likewise, the correlation coefficient between lexical variability (TTR) and accuracy (EFC/C) was negative and statistically significant (*p*<0.05). However, the correlation between subordination (DC/C) and accuracy (EFC/C) was positive (.505) and statistically significant (*p*<0.05).

Table 5.4

Correlation Matrix (Lexical Variability, Subordination, and Accuracy)

		TTR	DC/C	EFC/C
TTR	ρ	1.000	-.562**	-.419*
	Sig. (2-tailed)		.005	.047
	N	23	23	23
DC/C	ρ	-.562**	1.000	.505*
	Sig. (2-tailed)	.005		.014
	N	23	23	23
EFC/C	ρ	-.419*	.505*	1.000
	Sig. (2-tailed)	.047	.014	
	N	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed);** Correlation was significant at the 0.01 level (2-tailed).

To sum up, the data, the residual and the moving correlation plots indicated that the growth trajectories of lexical variability (TTR) and subordination (DC/C) were both alternating and parallel. The correlation coefficient confirmed the competitive relationship between the two sub-constructs of complexity. Likewise, the data, the residual and the moving correlation plots showed that the trajectories of the lexical variability (TTR) and accuracy (EFC/C) were both alternating and parallel. As in the TTR-DC/C interaction, the correlation coefficient suggests that there was a competitive relationship between lexical variability (TTR) and accuracy (EFC/C). The correlation coefficient indicates that there was a supportive relationship between subordination (DC/C) and accuracy (EFC/C), although the data and the residual plot indicated both alternating and parallel variability patterns.

5.5 Modelling

The purpose of this section was to test the hypothesised interactions between lexical variability (TTR) and subordination (DC/C), between lexical variability (TTR) and accuracy (EFC/C), and between subordination (DC/C) and accuracy (EFC/C) by simulating the developmental processes of these three constructs. After data collection, data description, and data exploration, such as variability analyses, the hypotheses can be tested by modelling the development. Based on the interaction analyses (section 5.4) the following hypotheses were formulated: there was

1. a moderate competition between lexical variability (TTR) and subordination (DC/C),
2. a moderate competition between lexical variability (TTR) and accuracy (EFC/C),
3. a moderate support between subordination (DC/C) and accuracy (EFC/C).

To test the abovementioned hypotheses, a model of three connected growers was used. "The coupled-dynamics model specifies only the supportive and competitive relationships, not the conditional relationships" (van Geert, 2008, p. 195). The dynamic model was based on the logistic growth model with an added supportive and competitive relationship (van Geert, 1991; 1994), discussed in section 4.8.4. The coupled-dynamics model was implemented in the form of a program, and it was programmed in MS Excel spreadsheet.

5.5.1 Model Setup

Three growers were specified in the model which corresponded with the three constructs under investigation in section 5.4 that was: (1) lexical variability (TTR), (2) subordination (DC/C), and (3) accuracy (EFC/C). First, the property parameters were specified, namely the initial value (ini), the growth rate (r), and the carrying capacity (K). The initial values were configured as the data onset values, while the carrying capacity values were configured as the default value of 1. Since all three growers were configured as ratios, namely type token ratio, dependent clause ratio, and error-free clause ratio, the maximum attainable value was 1. The developmental delay and the random variation parameters were omitted from the model setup to focus on the relationships among the three growers. Table 5.5 shows the numerical values that were assigned to the parameters in the model. Grower A corresponded to the lexical variability (TTR) index, Grower B referred to the subordination (DC/C) index, while Grower C represented the accuracy (EFC/C) index.

Table 5.5

Property and Relational Parameters

		Grower A	Grower B	Grower C
Property	Initial value	0.551	0.4	0.6666
parameters	Rate	0.05	-0.01	-0.01
	Carrying capacity	1	1	1
Relational		to A	to B	to C
parameters	from A	-	-0.028	0.021
	from B	-0.028	-	0.025
	from C	0.021	0.025	-

The second step included the configuration of the relational parameters. Although there could be precursor interactions among the three growers, this study focused only on the interactions between the variables. Consequently, the precursor values in the model were omitted. The relational parameters were entered in line with the hypotheses of the variability analyses. Table 5.5 also shows the relational parameters used in Dalma's data.

After specifying the property and relational parameters the data were simulated. The outcome of 100 iterations of the configured equations was obtained by the data simulation. Figure 5.24 shows the result of the simulations after 100 iterations.

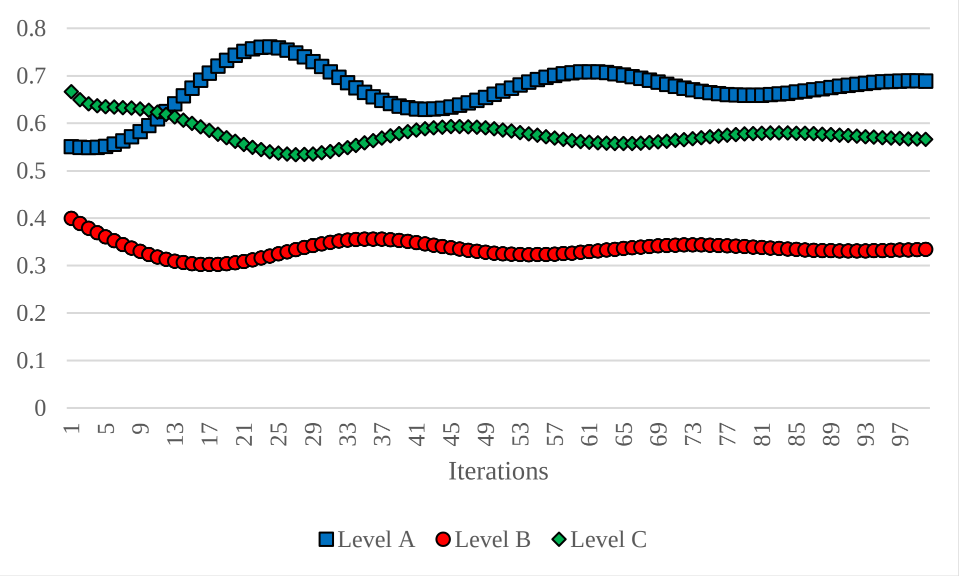


Figure 5.24 Growth model

5.5.2 Optimization

Before evaluating the model by correlating the model with the data, with the linear trends, and the cubic spline interpolations, the relational and property parameters were optimized. Two different steps can be performed to optimize a mathematical model: (1) simplification of the model, and (2) converging the parameter values via the Simplex algorithm (Nelder & Mead, 1965). In this study, only the second step was performed since the model was simple because precursor interactions were omitted. However, the bidirectional interactions could not be eliminated since the constructs of writing performance at advanced levels might include interactions between pre-acquired skills, and not relatively or partly novel knowledge (Caspi, 2010).

The initial values and the carrying capacity values were excluded from the optimization process since the initial values were configured as the corresponding data onset values, while the carrying capacity values were set to the default value of 1. Hence, the optimization focused only on the rate parameters and the relational parameters. Both the rate and the relational parameters were converged via the Simplex algorithm. The purpose of the optimization was to minimize the outcome of the function, in this case the sum of squared differences between the model and the data. Table 5.6 shows the optimized parameter values.

Table 5.6

Optimized Parameter Values

		Grower A	Grower B	Grower C
Property parameters	Rate	0.02063	0.00219	0.61072
Relational		to A	to B	to C
parameters	from A	-	0.11401	-0.3224
	from B	-0.5052	-	-0.1288
	from C	0.28556	-0.1379	-

The optimization of the parameters made it possible to assess the overall fit of the model to the data. First, the sum of squared residuals between the model and the data was compared, which was 0.2667 in Dalma's data. Second, the model outcome was visually contrasted with the data. Figure 5.25 shows that trajectories of the model was highly similar to the trajectories of the raw data.

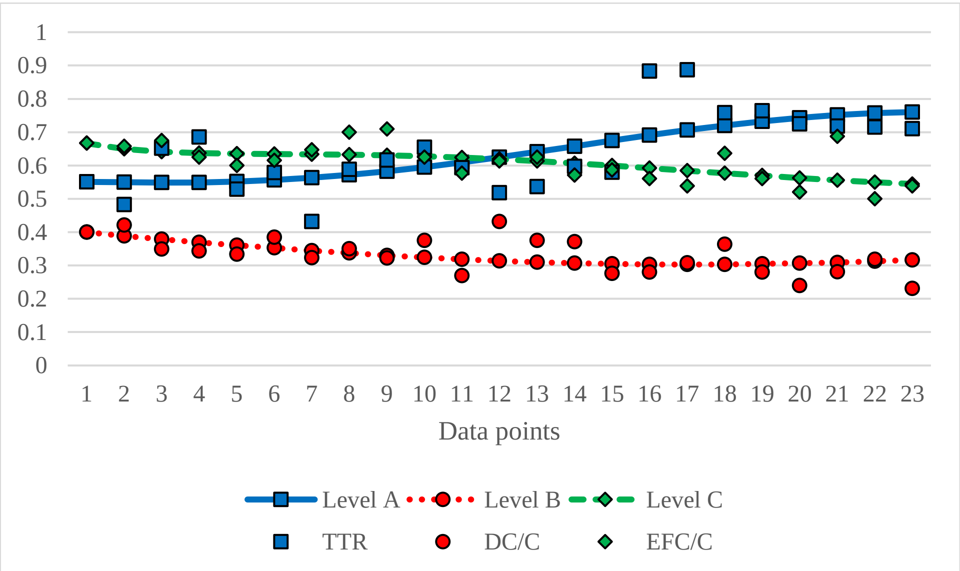


Figure 5.25 Data and model (fit = 0.26679)

The optimized parameter values confirm two of the three hypotheses that informed the model setup. The aggregated negative relational parameter value (Table 5.6) between Level A and B confirmed hypothesized moderately competitive association between the lexical variability (TTR) and subordination (DC/C). Interestingly, the relational parameter value from Level A to Level B was positive (0.11401), while from Level B to Level A it was negative (-0.5052). However, when the relational parameter was aggregated, the result was negative (-0.3911) confirming the negative relationship. The aggregated negative relational parameter value between Level A and Level C was also negative (-0.0368) corroborating the hypothesized negative relationship between lexical variability (TTR) and accuracy (EFC/C). While the relational parameters from Level A to Level C was negative (-0.3224), the relational parameter from Level C to Level A was positive (0.28556). The hypothesised moderate support between subordination (DC/C) and accuracy (EFC/C) was not confirmed by the model since both the relational parameter from Level B to C and vice versa were negative indicating a competitive relationship between subordination (DC/C) and accuracy (EFC/C).

5.5.3 Evaluation of the Model

The goodness-of-fit of the model can be confirmed by (1) correlating the model with the data, the linear trends, and the cubic spline interpolations, and by (2) comparing the correlation matrix between the various categories in the model, the data, the linear trends, and the cubic spline interpolations. Van Gelder and Port (1995) claimed that “if the dynamical model and the observed phenomena agree sufficiently in broad qualitative outline, then insight into the nature of the systems has been gained” (p. 16). Figure 5.26 presents the data, the linear trends, the cubic spline interpolations, and model outcome. Visual inspection of these plots indicates that the Level A, Level B and Level C resemble a combination between the linear trends and the spline interpolations.

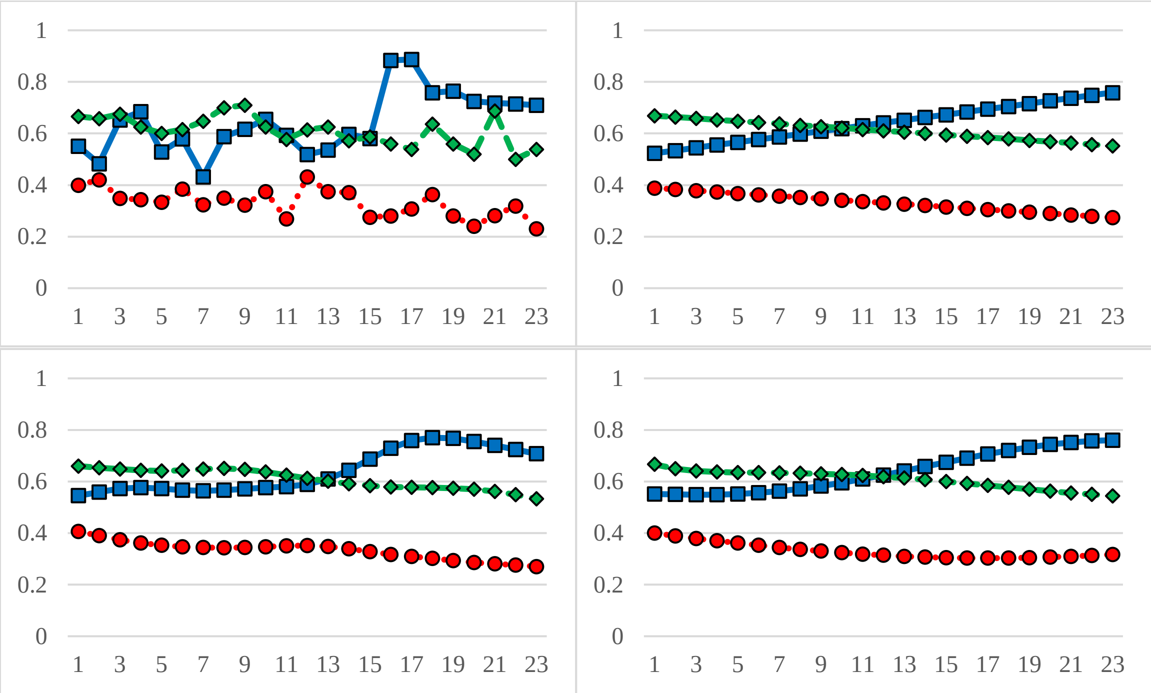


Figure 5.26 Data (top left), linear trends (top right), smoothing splines (bottom left) and model (bottom right)

The model was correlated with the data, the linear trends, and the smoothing splines. Table 5.7 shows that the model correlated positively at all three levels with the original data.

Table 5.7

Correlations Between the Model and the Data, Linear Trends, and Splines

			Model-data	Model-linear	Model-spline
Level A	TTR	ρ	.629**	.991**	.854**
		Sig. (2-tailed)	.002	.000	.000
		N	23	23	23
Level B	DC/C	ρ	.480*	.838**	.753**
		Sig. (2-tailed)	.021	.000	.000
		N	23	23	23
Level C	EFC/C	ρ	.612**	1.000**	.960**
		Sig. (2-tailed)	.002	.000	.000
		N	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed);** Correlation was significant at the 0.01 level (2-tailed).

In addition to the correlation between the model and the data, the linear trends, and the smoothing splines, another type of correlation analysis was performed to confirm the goodness-of-fit of the model. Correlation matrixes between the various categories in the data, the linear trends, the cubic spline interpolations, and the model were compared to see whether the model replicated the surface level interactions of the data. Table 5.8 shows the correlation matrixes for the data, the linear trends, the smoothing splines and the model.

Table 5.8

Correlation Matrixes

			Data	Linear	Spline	Model
TTR	DC/C	ρ	-.562**	-1.000**	-.766**	-.829**
		Sig. (2-tailed)	.005	.000	.000	.000
		N	23	23	23	23
TTR	EFC/C	ρ	-.419*	-1.000**	-.927**	-.991**
		Sig. (2-tailed)	.047	.000	.000	.000
		N	23	23	23	23
DC/C	EFC/C	ρ	.505*	1.000**	.849**	.838**
		Sig. (2-tailed)	.014	.000	.000	.000
		N	23	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed);** Correlation was significant at the 0.01 level (2-tailed).

Table 5.8 presents that correlation matrixes in the data, the linear trends, the smoothing splines and the model outcome show very similar patterns. There was a high similarity between the smoothing splines and the model outcome. The correlation coefficients between the TTR and DC/C indices and between the TTR and the EFC/C were negative in the data, the linear trends, the smoothing splines and the model outcome. In addition, the correlation coefficients between the data, the linear trends, the smoothing splines, and the model outcome were statistically significant ($p<0.05$). Likewise, the correlation coefficient between the DC/C and the EFC/C indices were positive and statistically significant ($p<0.05$) in the data, the linear trends, the smoothing splines and the model outcome.

5.6 Motivation

The purpose of this section is to explore the changes in Dalma’s motivational system. This section is subdivided into four parts: (1) goals for academic writing, (2) self-regulatory processes, (3) attractor states, and (4) external factors.

5.6.1 Goals for Academic Writing

During the initial interview (data point 3), Dalma reported that she would write her dissertation in Hungarian. However, she added that in the future she would like to write her second dissertation in English at a university in Germany. Consequently, she enrolled in the EAP course to take an IELTS language exam. Dalma also said that she would like to write reports in English in her future job. In order to improve her English, she reported that she usually reads in English. Dalma also said that she wanted to improve her skills in letter writing during the EAP course because she might send some e-mails to her prospective university in Germany. Dalma reported that her usual method of writing includes collecting ideas in Hungarian, writing in English, revising her essay and correcting the mistakes. She also said that she sometimes writes to her German friend in English. She also reported that she usually reads her English grammar books. Furthermore, she added that she had started to read Harry Potter in English. Dalma also reported vague goals such as “I would like to know more words”. Dalma also said that she regularly uses the Google translator during writing, and she often feels bored while composing texts.

5.6.2 Self-Regulatory Processes

Figure 5.27 shows the evolution of Dalma’s self-regulatory processes. Each value represents the number of themes (thematic units) identified in the interview data for data points 3, 6, 9, 12, 15, 18, and 21. Figure 5.27 shows that Dalma focused on self-observation processes between data points 9 and 21. Conversely, Dalma consistently utilised self-evaluation processes over the nine-month investigation. In other words, Dalma maintained her focus on self-evaluation processes over time. Figure 5.27 shows a dramatic increase of the self-evaluation processes at data point 9. Dalma set only one goal, at data point 12, over the course of the investigation.

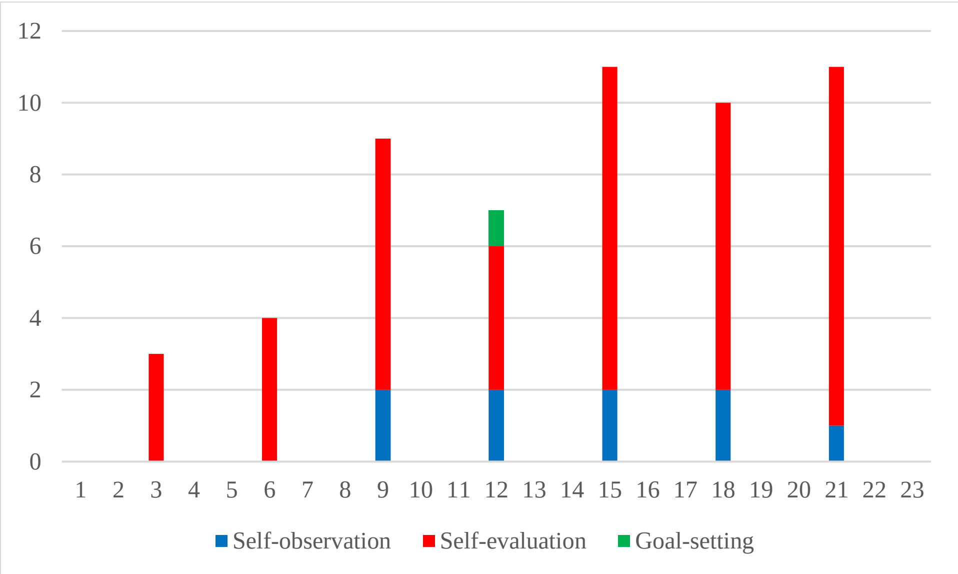


Figure 5.27 Self-regulatory processes

Table 5.9 provides a detailed breakdown of the range of Dalma's self-regulatory processes over the nine months. The number of self-observation thematic units was evenly distributed (ranging from 0 to 2 thematic units) over time. However, the types of self-observation thematic units were not evenly distributed since Dalma mainly focused on her composing processes. The number of self-evaluation thematic units was not evenly distributed. A dramatic increase can be observed in the number of self-evaluation thematic units between data points 15 and 21. Therefore, a shift from self-observation to self-evaluation, observed in Figure 5.27, was confirmed. In addition, the types of self-evaluation thematic units were not evenly dispersed since Dalma mainly focused on language. Table 5.9 also shows that the number of identified goals was 1 at data point 12.

Table 5.9

Self-Regulatory Processes

		Data point							
		3	6	9	12	15	18	21	Total
Self-observation	Language	0	0	0	1	1	0	0	
	Composing processes	0	0	1	1	1	1	1	
	Content	0	0	1	0	0	1	0	
	Quality	0	0	0	0	0	0	0	
	Total	0	0	2	2	2	2	1	9
Self-evaluation	Language	1	1	3	3	4	4	5	
	Composing processes	1	1	2	0	2	1	2	
	Content	1	1	1	1	2	2	2	
	Quality	0	1	1	0	1	1	1	
	Total	3	4	7	4	9	8	10	45
Goal-setting	Language	0	0	0	1	0	0	0	
	Composing processes	0	0	0	0	0	0	0	
	Content	0	0	0	0	0	0	0	
	Quality	0	0	0	0	0	0	0	
	Total	0	0	0	1	0	0	0	1

The following examples (Interview excerpt 5.1-5.4) show that Dalma's focus shifted from self-observation to self-evaluation. Interview excerpt 5.1 is an example of Dalma's impressionistic description of her action containing somewhat superficial description of her own writing at data point 12.

Interview excerpt 5.1 (Self-observation: language; data point 12)

Interviewer: Were there any points during the composition when you were searching for a particular word or phrase that you wanted to use?

Dalma: Not exactly, I wrote in an easier way so… these words I can use in everyday. So this is my main vocabulary.

However, Interview excerpt 5.2-5.4 contain evaluative comments that involved cognitive processes, such as reasoning and comparing. Interview excerpt 5.2-5.4 also show that the specificity of Dalma's descriptions of her L2 learning experience did not shift from a lack of tangibility to more elaborate and detailed explanations over time.

Interview excerpt 5.2 (Self-evaluation: language; data point 12)

Interviewer:	Did you try out any new words or phrases that you haven't used much before?
Dalma:	I think no.
Interviewer:	Is there a reason?
Dalma:	My reason is I didn't have enough time to improve my vocabulary
Interviewer:	Because now you are writing your dissertation.
Dalma:	Yes, every day.

Interview excerpt 5.3 (Self-evaluation: language; data point 12)

Interviewer:	Are you happy with the level of vocabulary?
Dalma:	No. I'm not satisfied with my vocabulary because I feel I repeat myself too much.

Interview excerpt 5.4 (Self-evaluation: language; data point 12)

Interviewer:	Compared to the previous pieces of writing was it (Essay 12) an improvement?
Dalma:	I think no because my last essay (Essay 9) was I think the same level so I'm not satisfied with my vocabulary.

Dalma set only one goal over the course of investigation at data point 12. Interview excerpt 5.5 shows the shift from self-evaluation to goal-setting in Dalma's interview data.

Interview excerpt 5.5 (Goal-setting: language; data point 12)

Interviewer:	Do you think that this piece of writing is a good representation of how well you can write?
Dalma:	Not how well. So this is my level today.
Interviewer:	How is it going to be better? What should you do?
Dalma:	So I want to read more in English and just improve my language and my grammar so just deal with English.

5.6.3 Attractor States

The first dominant state mentioned by Dalma was anxiety. She started attending the EAP course in September 2014. However, at the same time she was also writing her BSc dissertation. The submission deadline of her dissertation was in January 2015 which coincided with data point 12 of the data collection. Indeed, at data point 12 Dalma admitted that she was unable to prepare for the EAP course because she had to invest all her effort into writing and finishing her dissertation. Therefore, she also had to skip some classes. Furthermore, she composed Essay 15 later than the other participants of this study. She also explained the stagnation of her vocabulary development by her limited time to invest into learning the new words. Interview excerpt 5.2 also illustrates her problem at data point 12.

During the final interview Dalma elaborated on the anxiety or stress caused by writing her dissertation. Waninge (2015) claimed that anxiety "has the possibility to be more detrimental for learning" (p. 203). Indeed, at data point 12, one of the lowest lexical variability values was measured (TTR = 0.52). However, it was also interesting that the highest DC/C value (0.43) was measured at data point 12 which might suggest that Dalma employed the highest number of dependent clauses during the data collection. It can be speculated that L2 writers under stressful situations, such as in Dalma's data, prefer to step back in the developmental ladder, as suggested by Biber et al. (2011), and employ subordination instead of phrasal complexity in their academic writing. Interview excerpt 5.6 illustrates Dalma's anxiety she experienced at data point 12.

Interview excerpt 5.6 (data point 21)

Interviewer:	Were you motivated during the course?
Dalma:	Yes, I think that was the problem with my motivation because that was my last semester at the university and… it was really nervous for me because my teacher was horrible… and I was really stressed about my thesis… so I think my motivation could have been better.

The anxiety that Dalma experienced during one part of the EAP course might have had detrimental effects on the development of her self-regulatory processes and consequently her L2 writing development.

5.6.4 External Factors

The purpose of this subsection is to identify external factors that might have affected Dalma’s L2 learning experience. Interview excerpt 5.7 illustrates that Dalma did not have time to learn for the EAP course because she was preparing for her state examination.

Interview excerpt 5.7 (data point 21)

Interviewer: Do you think your level of motivation had an effect on the quality of your essays you were writing? Maybe on one day you were more motivated and

Dalma: Aha, yes because after my state examination I’ve got much more free time than before so I was able to start learn new words and grammar and I think you could see from my essays

Dalma submitted her dissertation and passed her state examination before data point 15. Therefore, she had time to learn for the EAP course. Interview excerpt 5.8 illustrates that she learned the lexical items for the class, and she was able to use them in Essay 15. The Academic Word List (AWL) index also started to rise at data point 15. However, the developmental peak was detected at around data point 16 and 17 (see Figure 5.17).

Interview excerpt 5.8 (data point 15)

Interviewer: So thinking about the language in this particular piece of writing did you try out any new words or phrases that you haven’t used?

Dalma: Yes, ‘hence’, ‘furthermore’, ‘venture’, ‘accentuate’, ‘to sum up briefly, ‘inevitable’.

Furthermore, Interview excerpt 5.9 also demonstrates that she felt a certain type of achievement by using lexical items she learnt during the EAP course.

Interview excerpt 5.9 (data point 15)

Interviewer: Are you happy with the level of vocabulary?
Dalma: Yes, this is really better than the other… I was able to put these new words.

5.6.5 Co-Adaptation

Co-adaptation refers to change in a system that is triggered by change in another closely related system (Larsen-Freeman & Cameron, 2008). There was a strong indication of co-adaptation of linguistic and motivational systems in Dalma's data. Her motivational system might have been settled into an attractor state between data points 1 and 12. Interview excerpt 5.6 show that Dalma felt anxious because of her state examination during this period. However, the successful completion of her state examination ceased the feeling of anxiety. In addition, Dalma had time to prepare for the EAP class and concentrate on learning new vocabulary. She also set a goal at data point 12 as illustrated by Interview excerpt 5.5. In other words, the efficient use of self-regulatory processes contributed to a change in her motivational system. Furthermore, the constant perturbations in the form of vocabulary tests and external factors such as time to learn led to the weakening of the attractor states. It was quite likely that a change in her lexical system, demonstrated by the detection of developmental peaks in lexical variability and Academic Word List at around data point 16, might have been triggered by a change in her motivational system.

5.7 Summary

All three constructs of linguistic complexity exhibited nonlinear developmental trends over the nine-month period in Dalma's written data. It was explored that there were developmental peaks in two of the five lexical complexity indices, namely the lexical variability (TTR) and the Academic Word List (AWL) indices. Nonetheless, in the five syntactic complexity indices and in the accuracy index, developmental peaks were not detected. The visual inspection of the growth trajectories, along with the residual plots, and the moving correlation plots, indicated that the interactions between lexical and syntactic complexity and accuracy were dynamic. Both the magnitude and the polarity of these relationships fluctuated over the nine months. Furthermore, correlation coefficients were also calculated to explore the surface interactions between the constructs. The model confirmed the competitive association between lexical variability (TTR) and subordination (DC/C) and between lexical variability (TTR) and accuracy (EFC/C). However, the model did not confirm the hypothesised moderately positive relationship between subordination (DC/C) and accuracy (EFC/C). The evolution of Dalma's self-regulatory processes was also investigated and it was found that her focus shifted from self-observation to self-evaluation over the course of the investigation. This finding suggests that Dalma developed her self-regulatory processes over time. However, only one goal was identified in her interview data. Therefore, attractor states were identified in her interview data, and it was found that anxiety might have contributed to the stagnation of self-regulatory processes.

Chapter 6. Case Study Two

Chapter 6 presents the changes in the lexical and syntactic complexity and accuracy indices in Emese's written data. Furthermore, the changes in her self-regulatory processes are presented. There are six main sections: (1) the participant and the context, (2) the developmental profile, (3) the degree of variability, (4) interactions, (5) modelling, and (6) motivation.

6.1 The Participant and the Context

Emese (pseudonym), a 19-year-old first-year Bachelor of Economics student, was the second participant of this study. Her mother tongue is Hungarian, and she had been learning English for 12 years at the beginning of the data collection. Emese's English language education started at the age of seven at a primary school in Budapest. She attended three English classes on a weekly basis over six years. At the age of 12 she enrolled in a prestigious secondary school in Budapest. During her secondary school years, she also attended three English classes on a weekly basis. At the end of secondary school Emese successfully took a final exam in English at a medium level. In addition, Emese took the City and Guilds language test which scored her at a C1 CEFR level.

Emese wanted to spend several months at a university in Ireland. In order to improve her chances of obtaining an Erasmus scholarship, Emese wanted to take the IELTS language exam. Therefore, she enrolled in the EAP course offered by a university in Budapest, Hungary. Emese attended two-hour long classes on a weekly basis from September 2014 to May 2015.

The data for this case study were collected using the same procedure as outlined in Chapter 4. Table 6.1 shows Emese's written corpus and timeline.

Table 6.1

Written Corpus and Timeline

	Written sample				Interview	
Data point	Code	Writing prompt	Date of submission	Word count	Type	Duration
1	Essay 1	WP1	19/09/2014	229		
2	Essay 2	WP2	03/10/2014	218		
3	Essay 3	CWP2	17/10/2014	258	Initial	6:34
			17/10/2014		Retrospective 1	3:44
4	Essay 4	WP3	31/10/2014	314		
5	Essay 5	WP4	07/11/2014	291		
6	Essay 6	CWP3	14/11/2014	291	Retrospective 2	3:25
7	Essay 7	WP5	28/11/2014	243		
8	Essay 8	WP6	05/12/2014	232		
9	Essay 9	CWP1	19/12/2014	249	Retrospective 3	7:15
10	Essay 10	WP7	19/12/2014	256		
11	Essay 11	WP8	09/01/2015	211		
12	Essay 12	CWP4	23/01/2015	254	Retrospective 4	6:06
13	Essay 13	WP9	30/01/2015	215		
14	Essay 14	WP10	06/02/2015	212		
15	Essay 15	CWP6	20/02/2015	260	Retrospective 5	6:31
16	Essay 16	WP11	27/02/2015	255		
17	Essay 17	WP12	06/03/2015	236		
18	Essay 18	CWP5	27/03/2015	260	Retrospective 6	6:10
19	Essay 19	WP13	03/04/2015	235		
20	Essay 20	WP14	10/04/2015	263		
21	Essay 21	CWP2	10/04/2015	239	Retrospective 7	5:58
			10/04/2015		Final	18:09
22	Essay 22	WP15	17/04/2015	224		
23	Essay 23	WP16	08/05/2015	274		
		Total word count		5720	Total duration	1:03:52

Note. Controlled data points are highlighted. WP=Writing prompt; CWP=Controlled writing prompt.

6.2 The Developmental Profile

This section is a detailed description of the developmental profiles of Emese's written data pertaining to complexity and accuracy. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity, and (3) accuracy.

6.2.1 Lexical Complexity

Figure 6.1 shows that all five indices developed nonlinearly and fluctuated to different degrees over the nine months. As in Dalma's written data, the general lexical complexity (AveWL) and Academic Word List (AWL) indices show similar patterns, while the lexical variability (TTR), lexical rarity (WRDFRQc) and lexical disparity (LSA) indices exhibit different trajectories over the nine-month investigation.

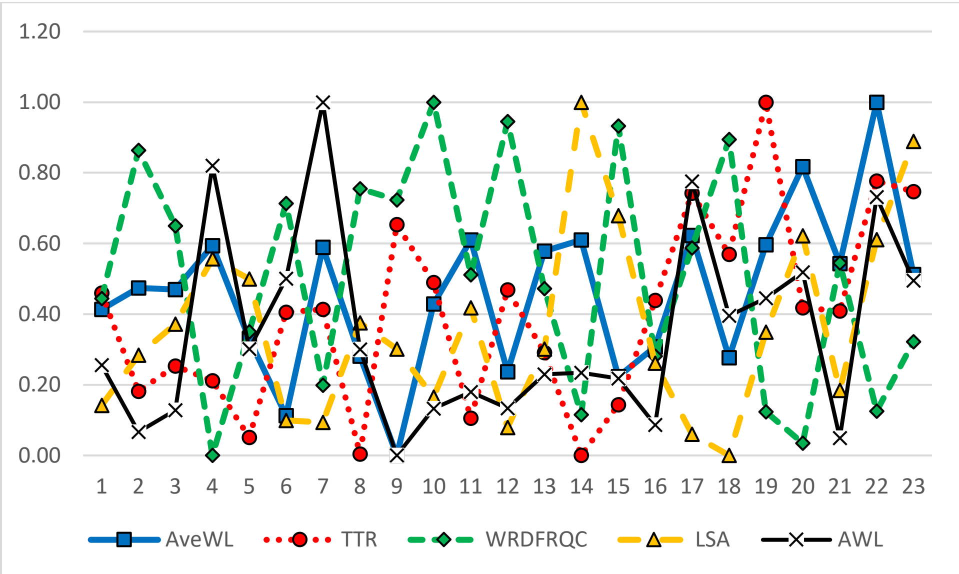


Figure 6.1 General lexical complexity

The average word length (AveWL) index was employed to measure general lexical complexity. Figure 6.2 shows that the AveWL index fluctuated between 4.12 and 5.43 characters. The lowest AveWL index was measured at data point 9, while the highest AveWL value was gauged at data point 22. The AveWL index fluctuated between 4.12 and 4.94 characters between data points 1 and 19. However, between data points 20 and 23 the AveWL index exceeded the five characters at data points 20 and 22. The smoothing algorithm of the general lexical complexity (AveWL) index displays a clear upward trend over the nine-month investigation.

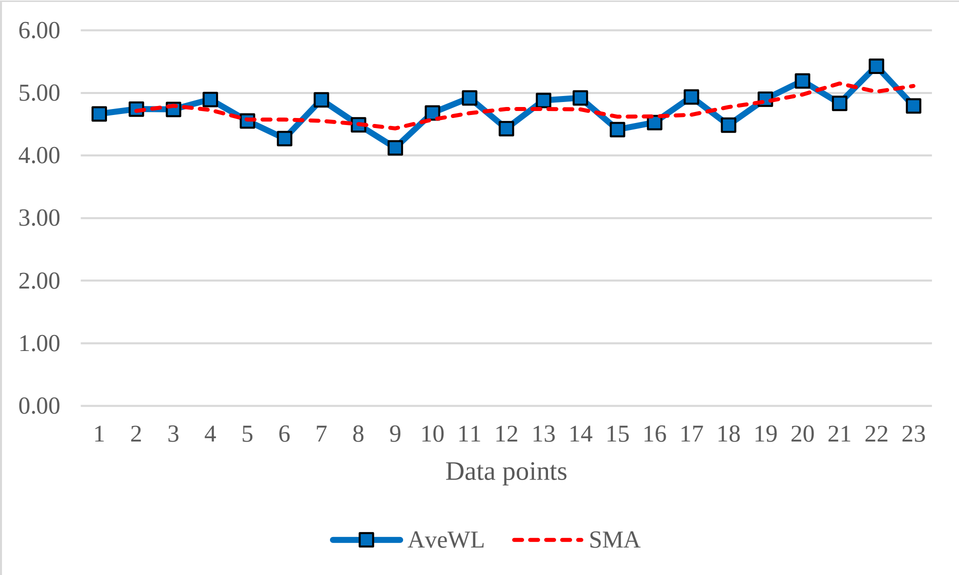


Figure 6.2 Lexical complexity (AveWL)

Lexical diversity is used as an umbrella term for variability, rarity, and disparity (Jarvis, 2013b). Figure 6.3 shows that the TTR index fluctuated between 0.55 and 0.79 over the nine-month investigation.

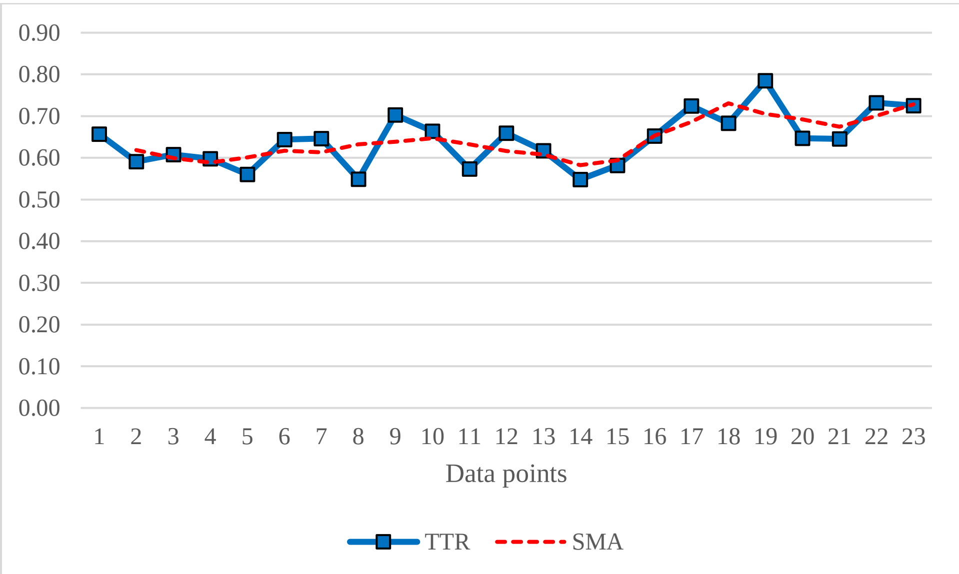


Figure 6.3 Lexical variability (TTR)

Figure 6.3 also shows that the TTR index varied between 0.55 and 0.70 between data points 1 and 16, while the TTR index fluctuated between 0.65 and 0.79 between data points 17 and 23. The TTR index reached its nadir at data point 14 (TTR = 0.55), while it reached its zenith at data point 19 (TTR = 0.78). The smoothing algorithm of the lexical variability (TTR) index displays a clear upward trend over the nine-month investigation.

The second property of lexical diversity, rarity, was measured by the CELEX log minimum frequency of content words index (WRDFRQc) and shown in Figure 6.4. The WRDFRQc index shows more fluctuations than the AveWL and the TTR indices over the nine months.

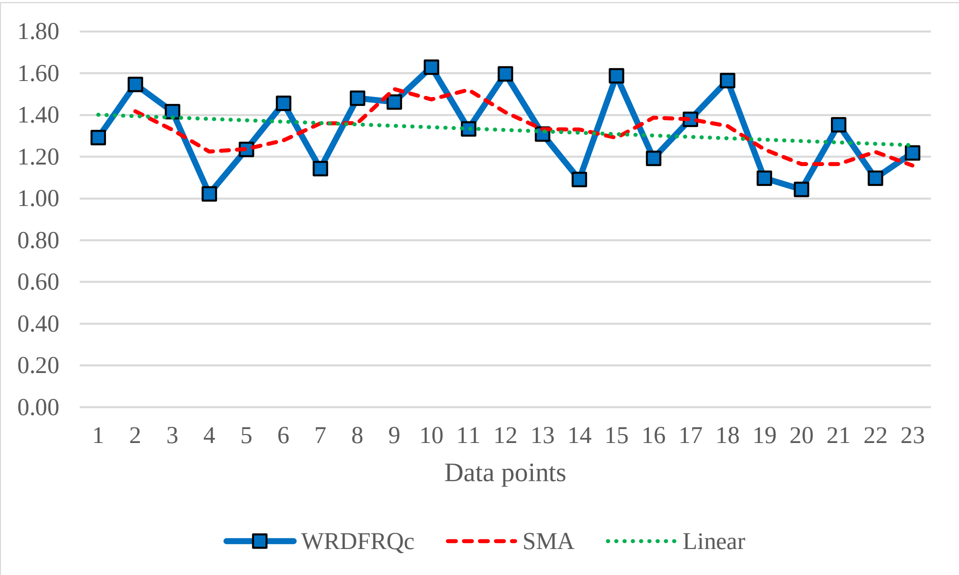


Figure 6.4 Lexical rarity (WRDFRQc)

The highest lexical rarity value (WRDFRQc = 1.63) was measured at data point 10, while the lowest WRDFRQc value (1.02) was detected at data point 4. The lexical rarity (WRDFRQc) index was below 1.2 at data points 4, 7, 14, 16, 19, 20, and 22. The essays written during these data points included less frequent words than the essays written at the rest of the data points. The smoothing algorithm of the lexical rarity (WRDFRQc) index displays a sideways trend over the nine-month investigation. However, the linear trend line of the WRDFRQc index shows a slightly downward trend over the nine months.

The third property of lexical diversity, disparity, was measured by the latent semantic analysis index (LSA). The LSA index (Figure 6.5) fluctuated wildly over the nine months. The lowest LSA value (LSA = 0.1) was measured at data point 18, while the highest LSA value (LSA = 0.45) was detected at data point 14. The smoothing algorithm of the lexical disparity (LSA) index displays a sideways trend over the nine-month investigation. However, the linear trend line of the LSA index shows a clear upward trend over the nine months.

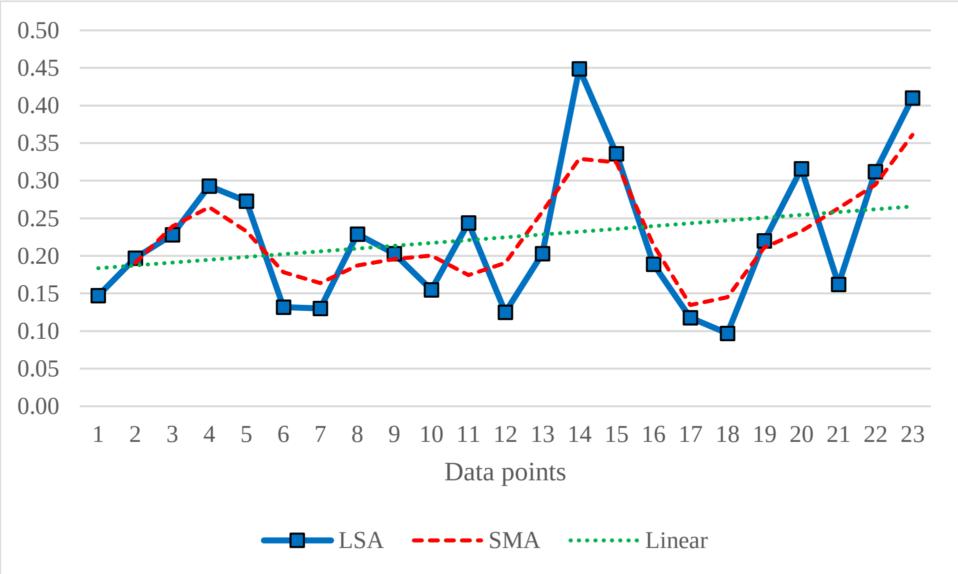
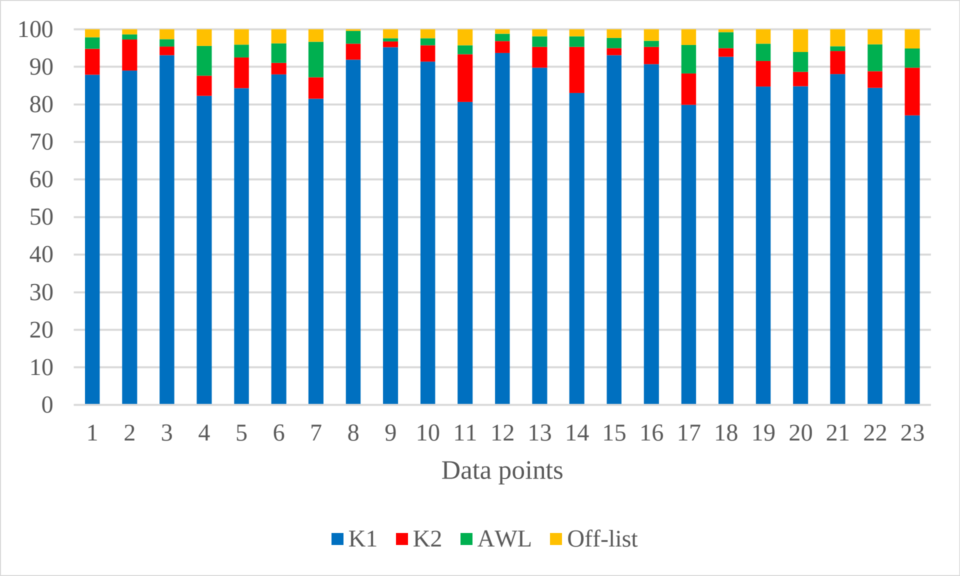


Figure 6.5 Lexical disparity (LSA)

To investigate the extent of genre-relevant lexical choice in Emese's academic writing, the percentage of academic words in her written data was estimated by the Academic Word List (AWL) measure (Figure 6.6). Figure 6.6 (top) shows the four frequency bands: K1 is the list of the most frequent 1000-word families, K2 is the list of the second most frequent 1000-word families, the Academic Word List, and off-list words. Figure 6.6 (top) shows that the highest percentage of K1 words was measured at data point 9 (K1 words = 95.2%), while the lowest percentage of K1 words was detected at data point 23 (K1 words = 77.09%). The low percentage of K1 words at data point 23 was attained by the increase of the percentages of K2 words. Indeed, one of the highest percentages of K2 words was detected at data point 23. In order to gain a better understanding of the use of the words from the Academic Word List, a separate graph was plotted (Figure 6.6, bottom).



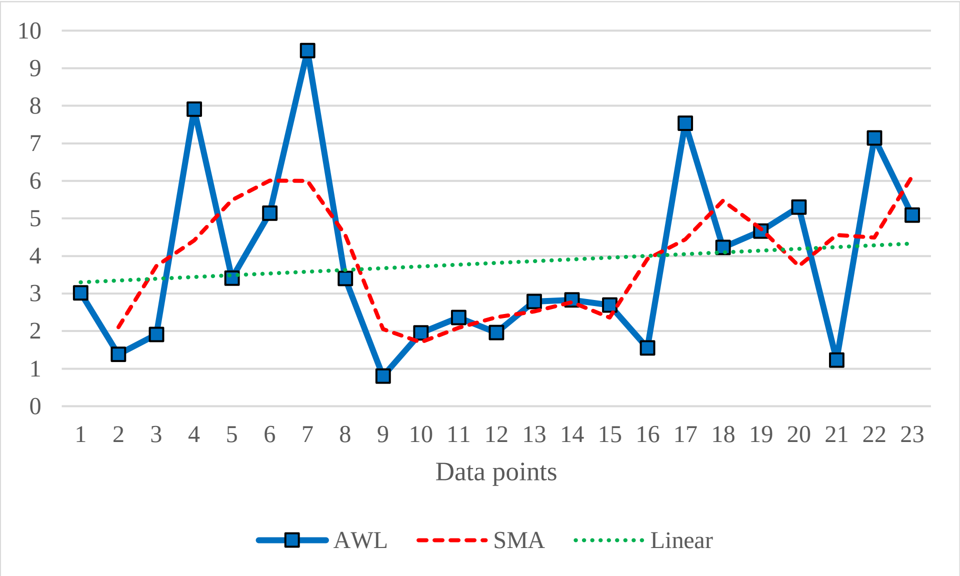


Figure 6.6 Lexical sophistication (top), and Academic Word List (bottom)

The AWL index shows wild fluctuations over the nine months. However, a stable period can be observed between data points 9 and 16. The AWL index reached its zenith (AWL = 9.47%) at data point 7, while the lowest point (AWL = 0.8%) was measured at data point 9. The smoothing algorithm of the Academic Word List (AWL) index displays a sideways trend over the nine-month investigation. However, the linear trend line of the AWL index shows a clear upward trend over the nine months.

6.2.2 Syntactic Complexity

Figure 6.7 displays the five syntactic indices used in this study. It can be observed that the FVR and the CN/C indices display similar trajectories since they tap into a similar construct, phrasal complexity. Whereas we could observe that Dalma made her writing more complex by employing more complex nominals (see Chapter 5), in Emese's data the FVR, the DC/C and the CN/C indices show similar trajectories over the nine months.

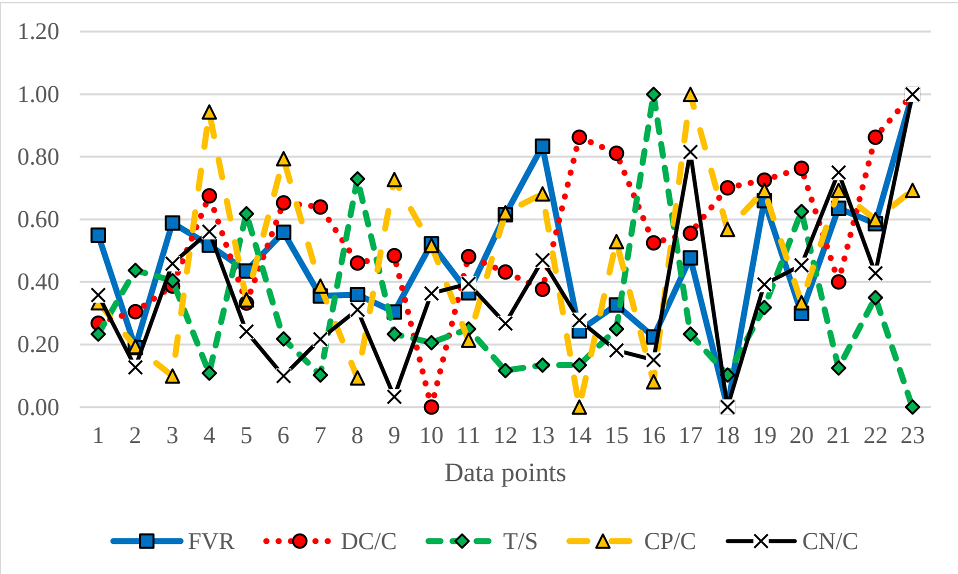


Figure 6.7 Syntactic complexity

General syntactic complexity was measured by the finite verb ratio (FVR) in Emese's written data. The FVR index, shown in Figure 6.8, dropped from data point 1 to 2, but then a relatively stable period can be observed. From data point 11, the FVR started to fluctuate and reached its lowest point at data point 18 (FVR = 5.91) but then it rose and reached its zenith (FVR = 10.65) at data point 23. The smoothing algorithm of the general syntactic complexity (FVR) index displays a sideways trend between data points 2 and 17. However, the linear trend line shows an upward trend over the nine months.

The following examples, taken from Emese's Essay 18 and Essay 23, illustrate the difference in the general syntactic complexity (FVR) index.

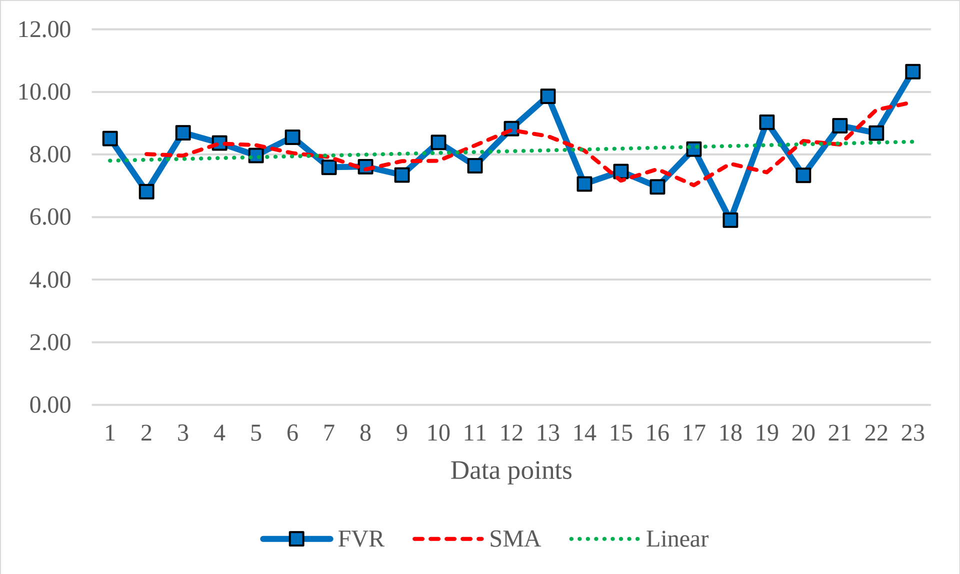


Figure 6.8 General syntactic complexity (FVR)

Essay excerpt 6.1 was extracted from Essay 18 where the FVR was the lowest (5.91). In contrast, Essay excerpt 6.2 was taken from Essay 23 when the FVR index was the highest (10.65). The finite verbs are underlined, and the example sentences are left uncorrected.

Essay excerpt 6.1 (Essay 18)

As people [1]<u>get</u> older they [2]<u>become</u> more difficult.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 8; finite verb count = 2; FVR = 4)

Essay excerpt 6.2 (Essay 23)

I [1]<u>can</u> imagine that the purpose of the organization of this event [2]<u>was</u> to divert people's attention from current war conflicts to sport events.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 24; finite verb count = 2; FVR = 12)

The two sentences exhibit different levels of syntactic complexity, although they are the same type of sentence i.e. complex. The ratio of the words per finite verbs clearly indicates that Essay excerpt 6.2 (FVR = 12) is syntactically more complex than Essay excerpt 6.1 (FVR = 4).

Emese's written samples were also coded for sentence types in order to gain a better insight into the syntactic structures she used in her written data. The distribution of the different sentence types is presented in a bar chart in Figure 6.9.

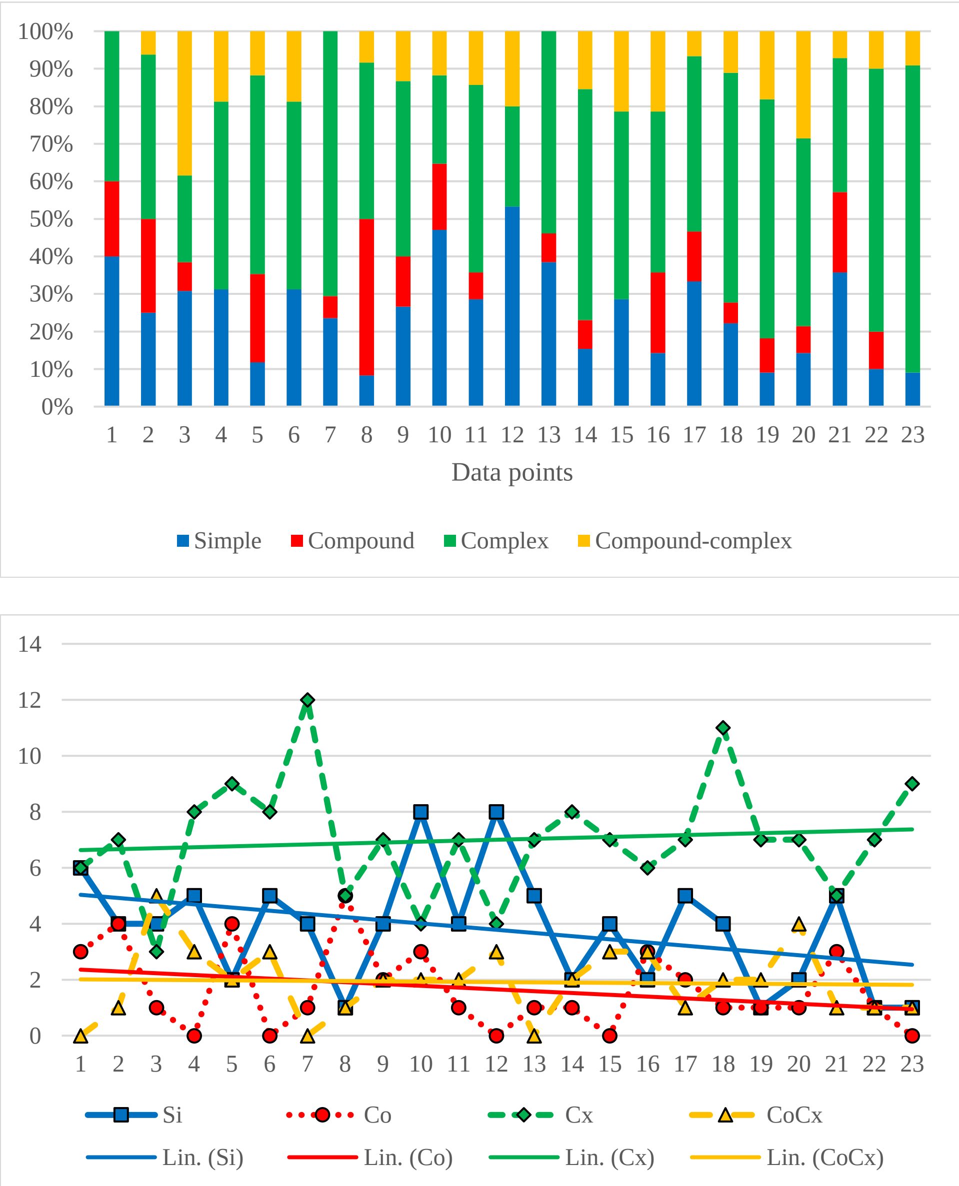


Figure 6.9 Sentence types (top: distribution; bottom: trend)

Figure 6.9 shows the distribution of sentence types in Emese's written data. Visual inspection of Figure 6.9 indicates that complex sentences dominated in Emese's written data. She used almost twice as many complex sentences (161) than simple sentences, and four times more complex sentences than compound (38) and complex-compound sentences (44). An increase can be seen in the use of complex sentences over the nine months, while the use of simple, compound and complex-compound sentences slowly decreased over time in Figure 6.9 (bottom). This phenomenon does not corroborate the current findings of academic writing which suggest that advanced users use more complex nominalisation rather than clausal embeddings in academic writing (Biber & Gray, 2010). Furthermore, the increase of complex sentences in Emese's data is also different from Dalma's data, in which an increase in the use of simple sentences and a decrease in the use of complex sentences can be observed. Simple sentences were present in every essay, while compound sentences were absent in Essay 4, 6, 12, 15, and 23. Not surprisingly, complex sentences were found in each essay, while complex-compound sentences were not used in Essay 1, 7, and 13.

Emese made her writing more elaborate by employing different syntactic structures such as subordination, coordination and phrasal complexity. Subordination was gauged in this study by the ratio of dependent clauses to clauses (DC/C) index (Figure 6.10).

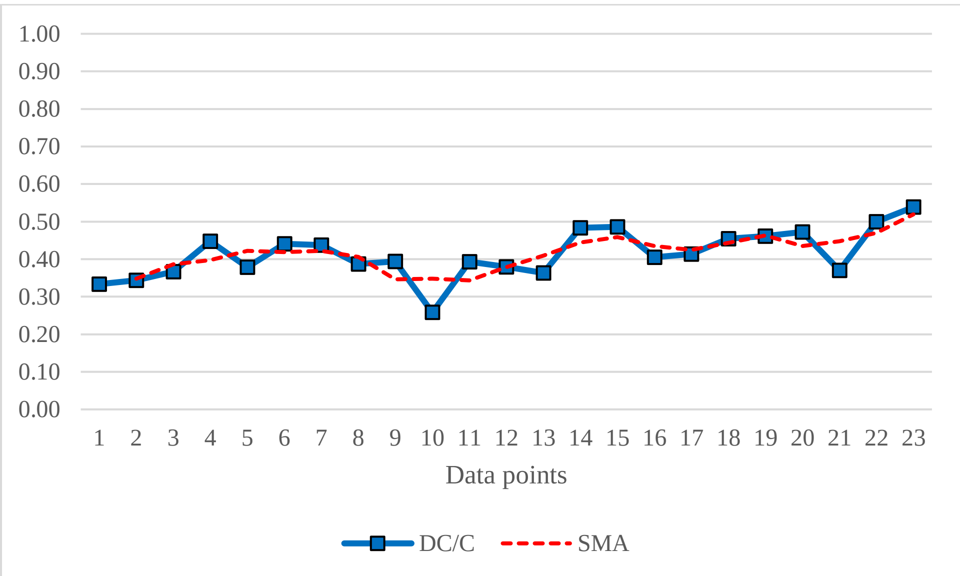


Figure 6.10 Subordination (DC/C)

Figure 6.10 shows that the DC/C index was relatively smooth over the nine months except for a decrease at data point 10, and 21. The smoothing algorithm of the subordination (DC/C) index displays an upward trend over the nine-month investigation. The lowest DC/C index value (DC/C = 0.26) was measured at data point 10, while the highest DC/C value (DC/C = 0.54) was gauged at data point 23. This finding means that every second clause was a dependent clause in Essay 23 as illustrated in the subsequent example.

Essay excerpt 6.3 (Essay 23)

Football, [1]<u>which is played all over the world</u>, is one of the most popular sports in the world [2]<u>because it is able to move the masses</u>.
(Errors are uncorrected; dependent clauses are underlined and numbered in superscript)
(Dependent clause count = 2; clause count = 3; DC/C = 0.66)

In addition to subordination, Emese also employed both clausal and phrasal coordination structures in her essays. The clausal coordination (T/S) index, shown in Figure 6.11, displays slight fluctuations between data points 1 and 9, and between data points 15 and 23.

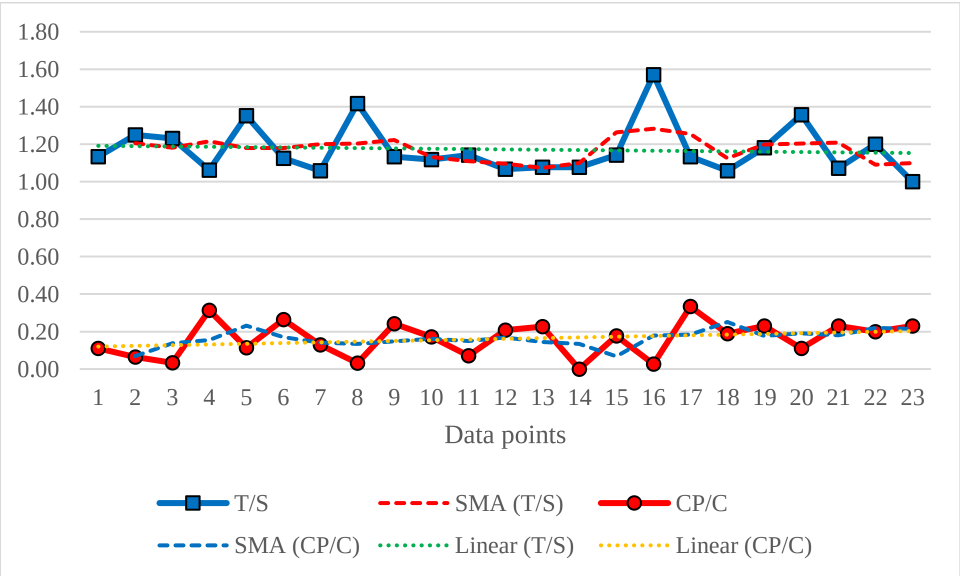


Figure 6.11 Coordination (T/S and CP/C)

However, between data points 9 and 15 a relatively stable period can be observed. The smoothing algorithm of the clausal coordination (T/S) index (red dashed line) shows a sideways trend over the nine-month investigation. However, the linear trend line of the T/S index (green dotted line) shows a slightly downward trend over the nine months.

In addition to, clausal coordination, Emese also employed phrasal coordination to make her writing more complex. Figure 6.11 shows that the CP/C index reached its zenith at data point 17 (CP/C = 0.33). This finding suggests that on average one coordinate phrase was used in every third clause in Essay 17. See Essay excerpt 6.4 for an example.

Essay excerpt 6.4 (Essay 17)

Finding the right balance between [1]traditional skills and technology is an incredibly difficult task.
(Errors are uncorrected; coordinate phrases are underlined and numbered in superscript)
(Coordinate phrase count = 1; clause count = 1; CP/C = 1)

The smoothing algorithm of the phrasal coordination (CP/C) index (blue dashed line) shows a sideways trend over the nine-month investigation. However, the linear trend line of the CP/C index (yellow dotted line) shows a slightly upward trend over the nine-month investigation.

Phrasal complexity was measured by the complex nominal per clause (CN/C) and the mean number of modifiers per noun phrase (SYNNP) indices. Figure 6.12 (top) shows that the CN/C index fluctuated wildly in contrast with the other syntactic indices. The lowest CN/C index value (CN/C = 0.54) was measured at data point 18, while he highest value (CN/C = 1.62) at data point 23. The smoothing algorithm of the phrasal complexity (CN/C) index displays an upward trend over the nine-month investigation. Figure 6.12 (bottom) shows that the lowest SYNNP values were measured at data points 9 and 18 (SYNNP = 0.47), while the highest SYNNP value (SYNNP = 1.03) was gauged at data point 17. The smoothing algorithm of the SYNNP index displays a sideways trend over the nine-month investigation. However, the linear trend of the SYNNP index shows a slightly upward trend over the nine months.

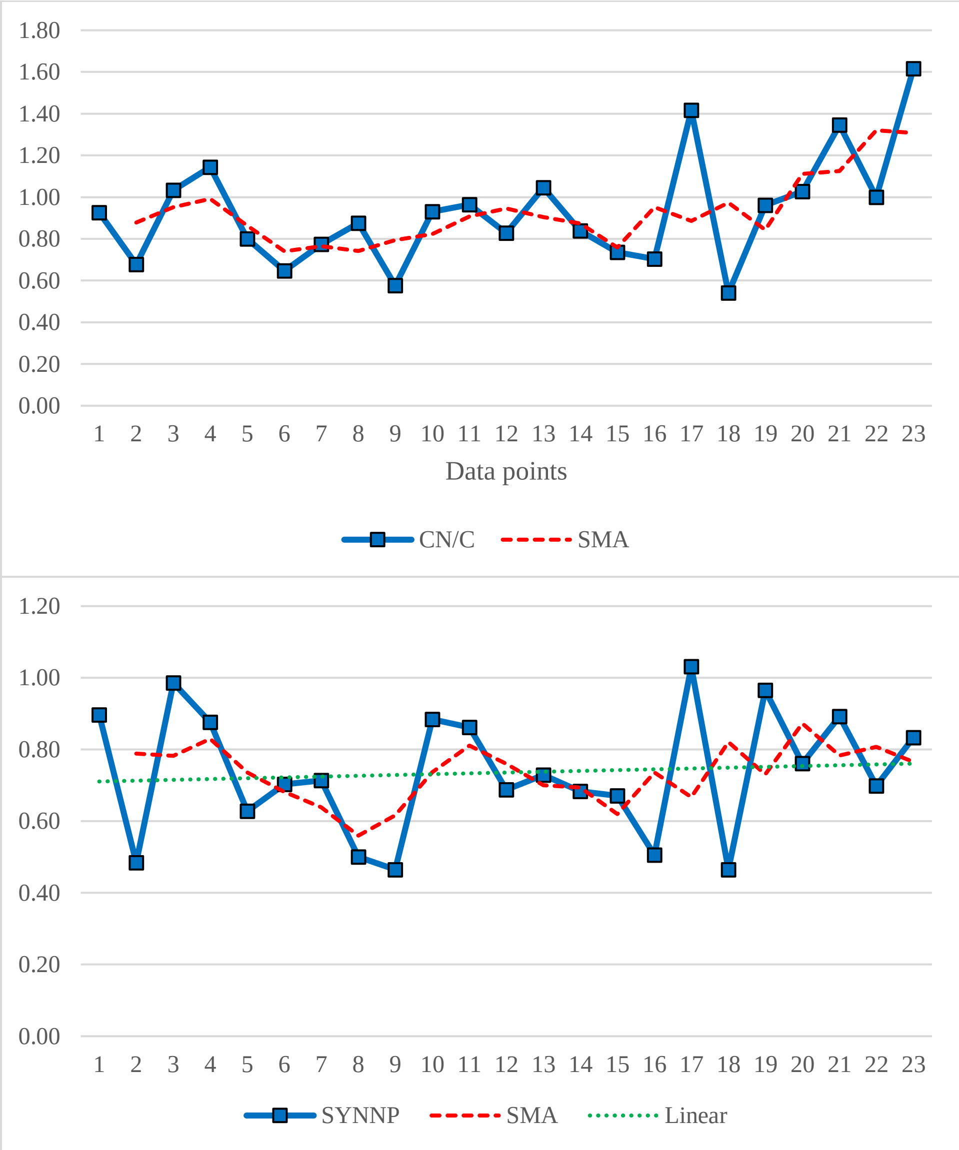


Figure 6.12 Phrasal complexity (CN/C and SYNNP)

Essay excerpt 6.5 exemplifies the high CN/C index value (CN/C = 1.56) at data point 23.

Essay excerpt 6.5 (Essay 23)

I can imagine that [1]the purpose of the organization [2]of this event was to divert [3]people's attention from [4]current war conflicts to [5]sport events.
(Errors are uncorrected; complex nominals are underlined and numbered in superscript)
(Complex nominal count = 5; clause count = 2; CN/C = 2.5)

Figure 6.13 shows the syntactic indices specific to the academic genre: the normed rate of occurrence of conditional clauses (ConC), the normed rate of occurrence of infinitive clauses (InfC), the normed rate of occurrence of relative clauses (RelC), and the incidence score of prepositional phrases (DRPP) indices.

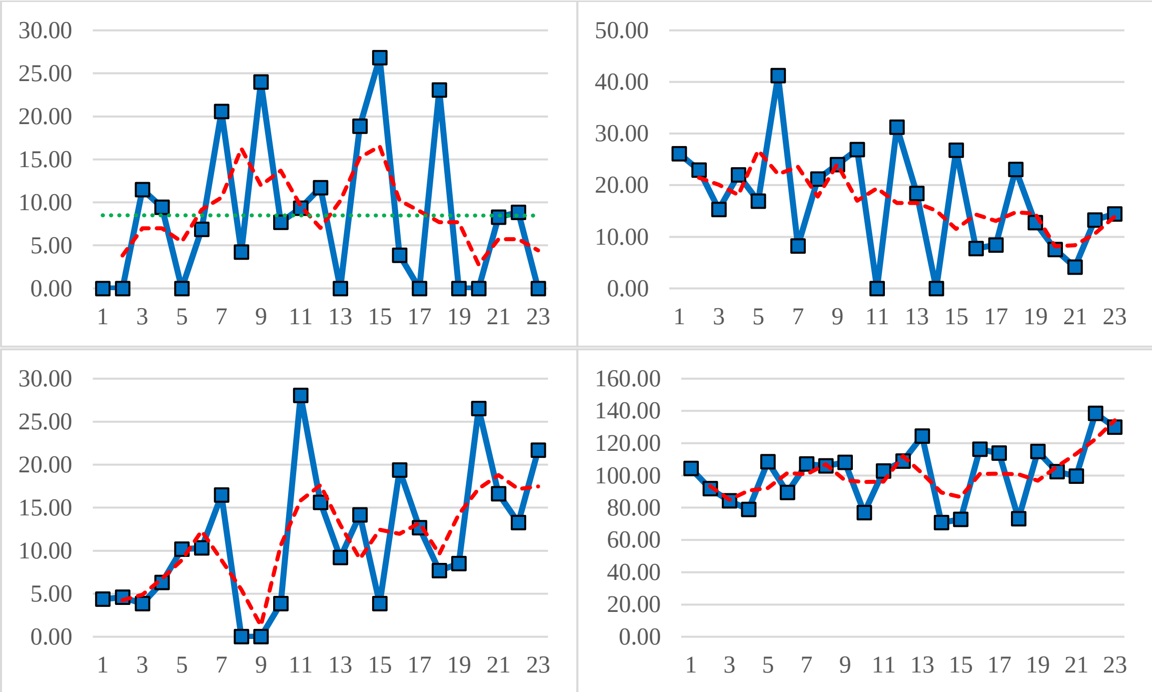


Figure 6.13 Syntactic measures specific to the academic genre (top left: ConC; top right: InfC; bottom left: RelC; bottom right: DRPP).

The smoothing algorithm of the conditional clauses (ConC) index shows a sideways trend over the nine-month investigation. Furthermore, the linear trend line of the ConC index also shows a sideways trend over time. The smoothing algorithm of the infinitive clauses (InfC) displays a downward trend over the nine-month investigation, whereas the smoothers of the relative clauses (RelC) and the prepositional phrases (DRPP) indices display upward trends over the nine months. In other words, there were increases in the use of the relative clauses (RelC) and the prepositional phrases (DRPP) in Emese's written data. However, there were decreases in the use of infinitive clauses (InfC) and there were no changes in the usage of conditional clauses (ConC) over the nine-month investigation.

6.2.3 Accuracy

As shown in Figure 6.14, the EFC/C index fluctuated between 0.68 and 0.79 from data points 1 to 11, but then wilder fluctuations can be observed between data points 11 and 15. The EFC/C index was higher between data points 15 and 23 than between data points 1 and 14. Interestingly, the EFC/C index remained over 0.8 from data point 18. The smoothing algorithm of the accuracy (EFC/C) index displays a clear upward trend over the nine-month investigation.

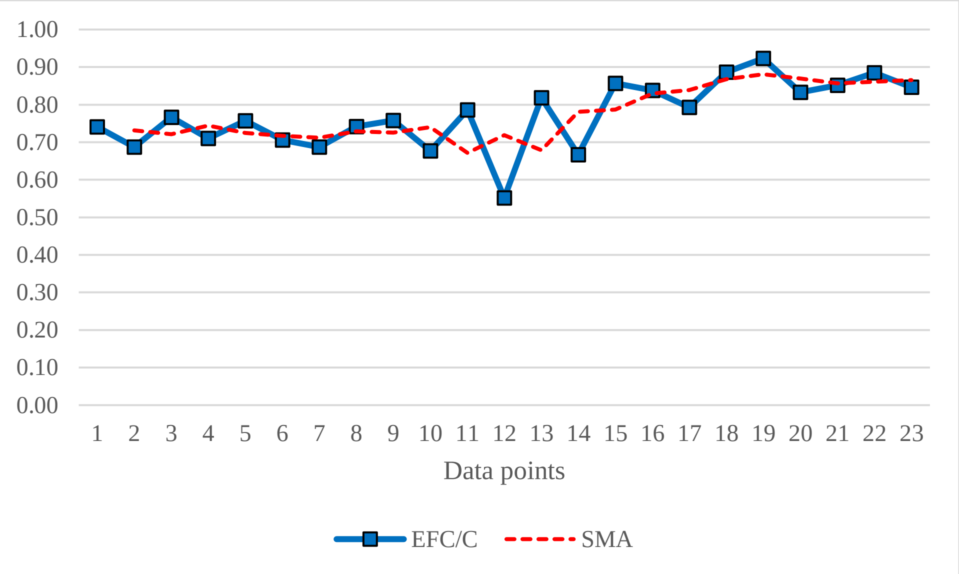


Figure 6.14 Accuracy (EFC/C)

The highest EFC/C value (EFC/C = 0.92) was detected at data point 19, while the lowest EFC/C value (EFC/C = 0.55) was measured at data point 12. The lowest EFC/C value means that there was at least one error in every second clause in Essay 12. The following examples illustrate the low accuracy rate in Essay 12.

Essay excerpt 6.6 (Essay 12)

To [1]summerise I agree with both [2]methodes but only when a student [3]use them together in same time.
(Errors are uncorrected, underlined and numbered in superscript)
(Error-free clause count = 2; clause count = 2; EFC/C = 1)

Essay excerpt 6.7 (Essay 12)

I reckon that the best way to learn a [1]languag is the "blend/concoction" of [2]this two [3]thing.
(Errors are uncorrected, underlined and numbered in superscript)
(Error-free clause count = 1; clause count = 2; EFC/C = 0.5)

In order to gain a deeper insight into the accuracy construct of Emese's written data, this study also looked at the type of errors she made in her essays. The errors were coded and categorised according to the Louvain Error Tag-set, and the results are displayed in Figure 6.15.

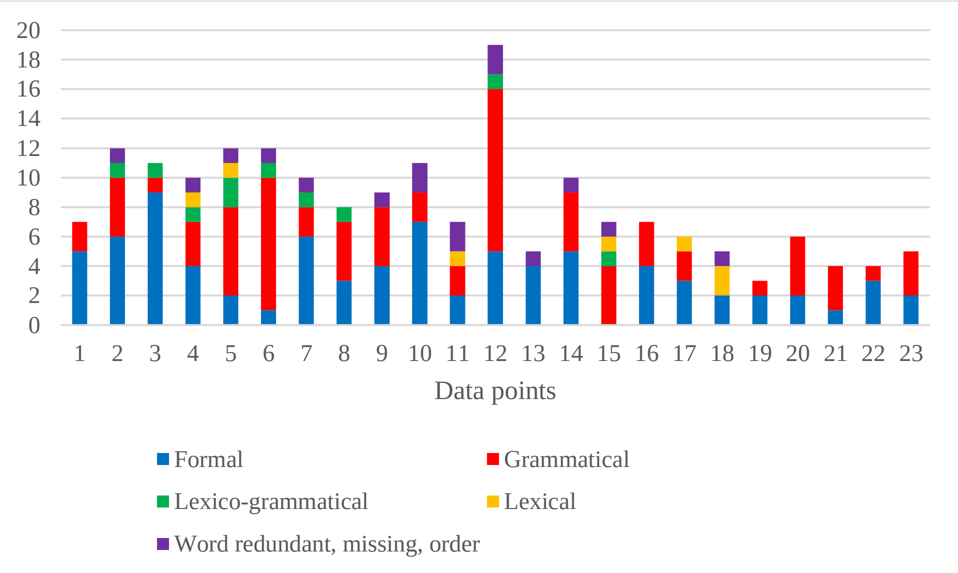


Figure 6.15 Error types

Figure 6.15 shows that the number of errors decreased over the nine months. However, a peak can be observed at data point 12. This finding corresponds to the decrease of the EFC/C index in Figure 6.14. Furthermore, Figure 6.15 shows that Emese made five different types of errors in her essays: (1) formal, (2) grammatical, (3) lexico-grammatical, (4) lexical and (5) word redundant, word missing and word order. Among the five different types of errors, formal errors (morphology and spelling) were the most prevalent (43%) in Emese's essays. In addition, formal errors were the only type of errors which were present in each essay except for Essay 15. The second most common type of error belonged to the grammatical errors category (39%).

Grammatical errors were evident in almost every essay except in Essay 13 and 18. The other three categories remained below 10% and their occurrence was also rather random.

6.3 The Degree of Variability

Section 6.3 explores the degree of variability in lexical and syntactic complexity and accuracy indices in Emese's written data. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity, and (3) accuracy.

6.3.1 Lexical Complexity

The min-max graph of the TTR index does not display clear stages as in Dalma's written data which was expected during the visual inspection of Figure 6.16. Although the bandwidth does not display great differences, the band moved up between data points 6 and 7 and then between data points 16 and 17. The visual inspection of the min-max graph of the TTR index suggests that there might not be statistically significant peaks in Emese's data. These ranges of variability in the lexical variability (TTR) index can be interpreted as follows: although the band moved up three times between data points 1 and 23, the lexical items were used in a regular pattern in Emese's written data over the nine-month investigation.

In the min-max graph of the LSA index, displayed in Figure 6.17, there were greater differences in the bandwidth than in Figure 6.16. In Figure 6.17 we can observe four main stages: (1) a rather wide bandwidth of variability between data points 1 and 11 (from 0.1 to 0.3), (2) a much wider bandwidth of variability between data points 11 and 17 (with a jump from 0.1 to 0.45), (3) a rather wide bandwidth of variability again between data points 17 and 20 (from 0.1 to 0.3) and then (4) a much wider bandwidth of variability between data points 20 and 23 (from 0.15 to 0.4).

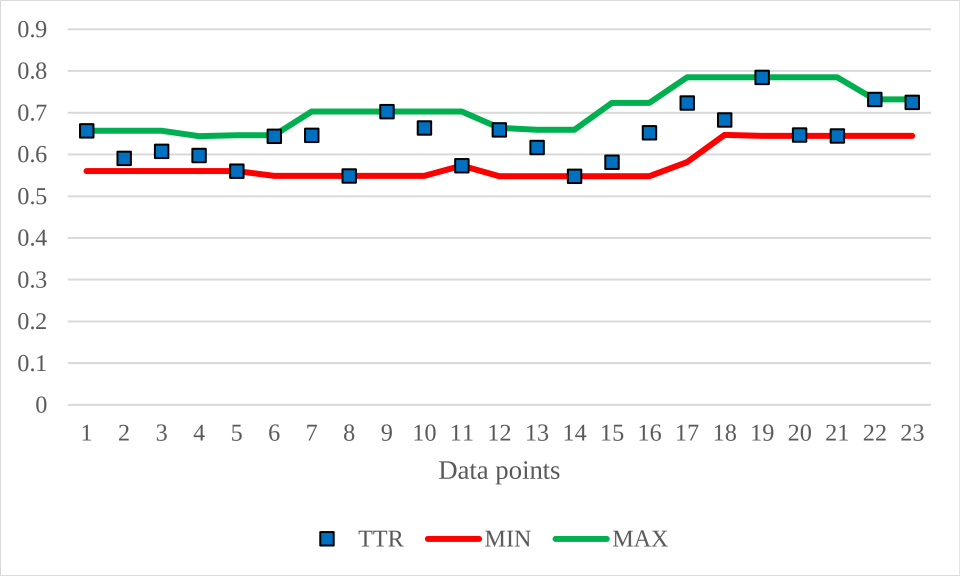


Figure 6.16 Lexical variability (min-max graph)

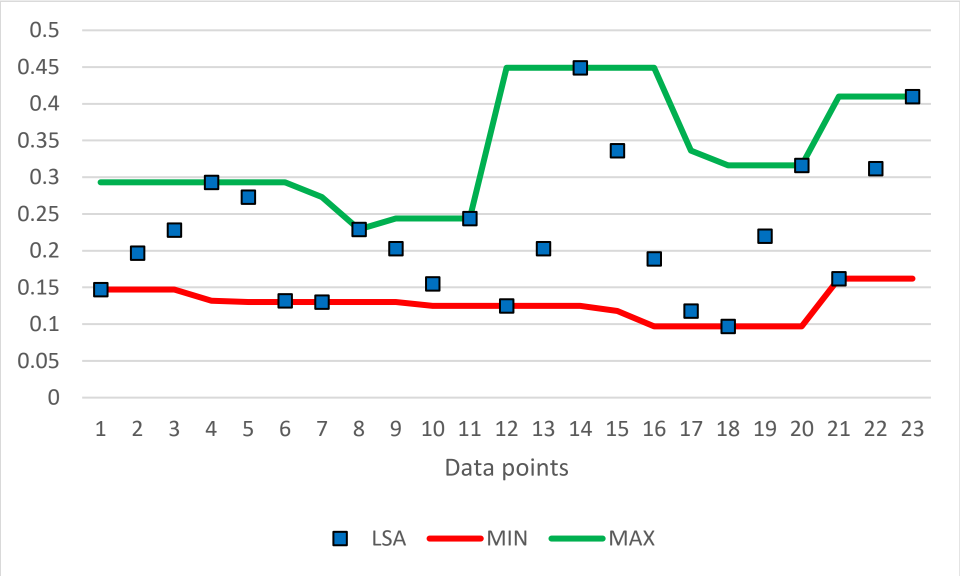


Figure 6.17 Lexical disparity (min-max graph)

The data of the lexical variability (TTR) and the lexical disparity (LSA) indices were resampled, and two separate Monte Carlo analyses were run. Table 6.2 shows that the p-value of the lexical variability (TTR) index was 0.12 and the p-value of the lexical disparity (LSA) index was 0.11. In other words, statistically significant developmental peaks were not detected in the lexical variability (TTR) and the lexical disparity (LSA) indices. The same statistical procedures were repeated for the general lexical complexity (AveWL), the lexical rarity (WRDFRQc) and the Academic Word List (AWL) indices. However, the p-values of the

AveWL, the WRDFRQc and the AWL indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the AveWL, the WRDFRQc, and the LSA indices in Emese's written data.

Table 6.2

P-values for the Lexical and Syntactic Complexity and Accuracy Indices

Index	P-value
AveWL	0.52
TTR	0.12
WRDFRQc	0.40
LSA	0.11
AWL	0.25
FVR	0.21
DC/C	0.20
T/S	0.56
CP/C	0.70
CN/C	0.47
EFC/C	0.58

6.3.2 Syntactic Complexity

In the min-max graph of subordination (Figure 6.18), four main stages can be observed: (1) a narrow bandwidth of variability from data point 1 to 7 (from 0.33 to 0.45), (2) a slightly wider bandwidth of variability from data point 8 to 12 (from 0.26 to 0.40), and then (3) a narrow bandwidth of variability from data point 13 to 18 (from 0.36 to 0.5), and finally (4) a marginally wider bandwidth of variability from data point 18 to 23. These ranges of variability can be interpreted as follows: in the first and the third stages, the dependent clauses were used in a more regular pattern, whereas in the second and the fourth stages dependent clauses were used randomly in Emese's written data.

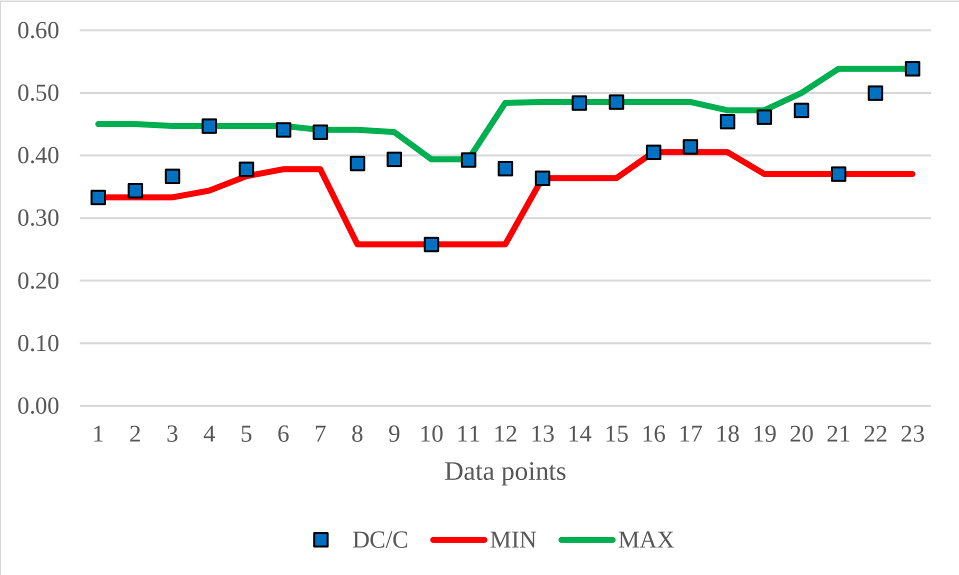


Figure 6.18 Subordination (min-max graph)

Table 6.2 shows that the p-values of the general syntactic complexity (FVR), subordination (DC/C), clausal coordination (T/S), phrasal coordination (CP/C) and phrasal complexity (CN/C) indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the five syntactic complexity indices.

6.3.3 Accuracy

The degree of variability in the construct of accuracy was measured with the error-free clause ratio (EFC/C) index. In the min-max graph (Figure 6.19) of EFC/C index, three main stages can be observed in the bandwidth of variability. The first stage was narrow (from 0.77 to 0.77) from data point 1 to 9. The second stage was rather wide (from 0.55 to 0.8) from data point 10 to 15. The third stage was narrow again (from 0.79 to 0.92) from data point 15 to 23. However, Figure 6.19 also shows that the band moved up from data point 15. These ranges can be interpreted as follows: in the first stage error-free clauses occurred in a regular pattern, while in the second stage, error-free clauses occurred randomly in Emese’s written data. In the last stage, error-free clauses occurred in a more regular pattern again.

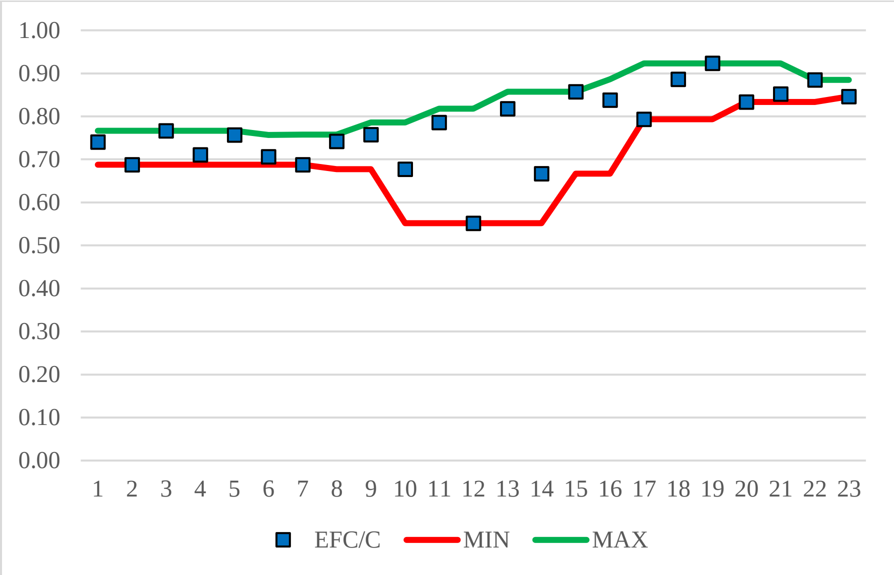


Figure 6.19 Accuracy (min-max graph)

Table 6.2 shows that the p-values of the accuracy (EFC/C) index was 0.58. In other words, a statistically significant developmental peak was not detected in the accuracy (EFC/C) index in Emese's written data.

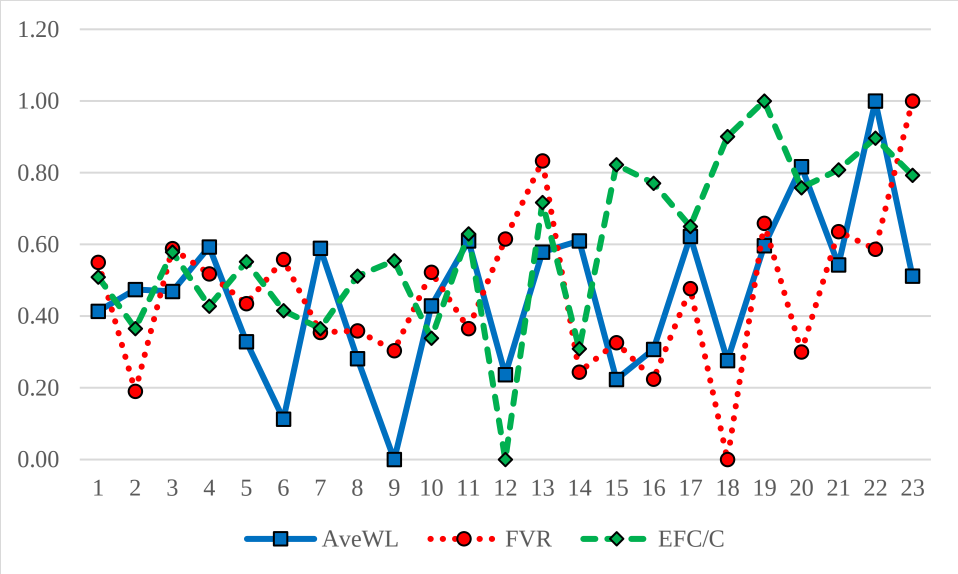
6.4 Interactions

The purpose of this section is to explore the dynamic interactions between lexical and syntactic complexity and accuracy. This section is subdivided into two parts: (1) interactions between general lexical and syntactic complexity and accuracy indices and (2) interactions between lexical variability, subordination, and accuracy indices.

6.4.1 Interactions Between General Lexical and Syntactic Complexity and Accuracy

Figure 6.20 shows (top) the normalised data of the general lexical complexity (AveWL) and general syntactic complexity (FVR) and accuracy (EFC/C) indices. The trajectories of the general lexical complexity (AveWL) and the general syntactic complexity (FVR) exhibit both alternating and parallel patterns. However, between data points 16 and 19, the growth trajectories of the AveWL and FVR indices were clearly parallel. The growth trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) were mainly alternating between data points 1 and 10. However, between data points 10 and 13, the growth trajectories of the AveWL and the EFC/C indices were clearly parallel. The growth trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) indices were both alternating and parallel between data points 13 and 23. The growth trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) indices were parallel between data points 1 to 4. However, from data point 4 the two growth trajectories of the FVR and EFC/C indices became predominantly alternating up to data point 12. Between data points 12 and 16 the two growth trajectories (FVR and EFC/C) were clearly parallel, while between data points 16 and 23 the two trajectories were both alternating and parallel.

Figure 6.20 (bottom) shows the residual plot for the AveWL, FVR and the EFC/C indices. The residual plot also displays both the parallel and the alternating shifts in the trajectories of general lexical complexity (AveWL) and general syntactic complexity (FVR). In addition, the clear parallel shifts were also evident between data points 16 and 19 in the residual plots of the AveWL and FVR indices. Between data points 11 and 12 and data points 20 and 21, a movement above the trend in the general syntactic complexity (FVR) index was accompanied by a movement below it in the general lexical complexity (AveWL) index.



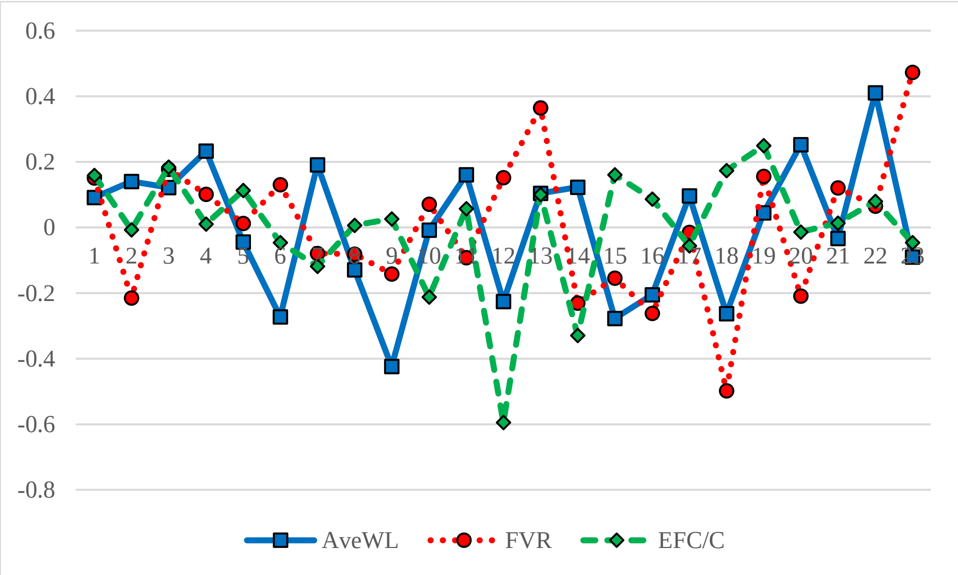


Figure 6.20 General lexical and syntactic complexity and accuracy (top: normalised data, bottom: residual plot)

Figure 6.21 shows the interactions between general lexical (AveWL) and syntactic complexity (FVR) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 6.21 shows that the general lexical complexity (AveWL)-general syntactic complexity (FVR) correlation shifted between moderately negative and moderately positive coefficient values. The general lexical complexity (AveWL)-accuracy (EFC/C) interaction shifted between moderately negative and moderately positive values. The general syntactic complexity (FVR)-accuracy (EFC/C) interaction also shifted between moderately positive and moderately negative values with one peak towards a strongly positive value in the 3rd window.

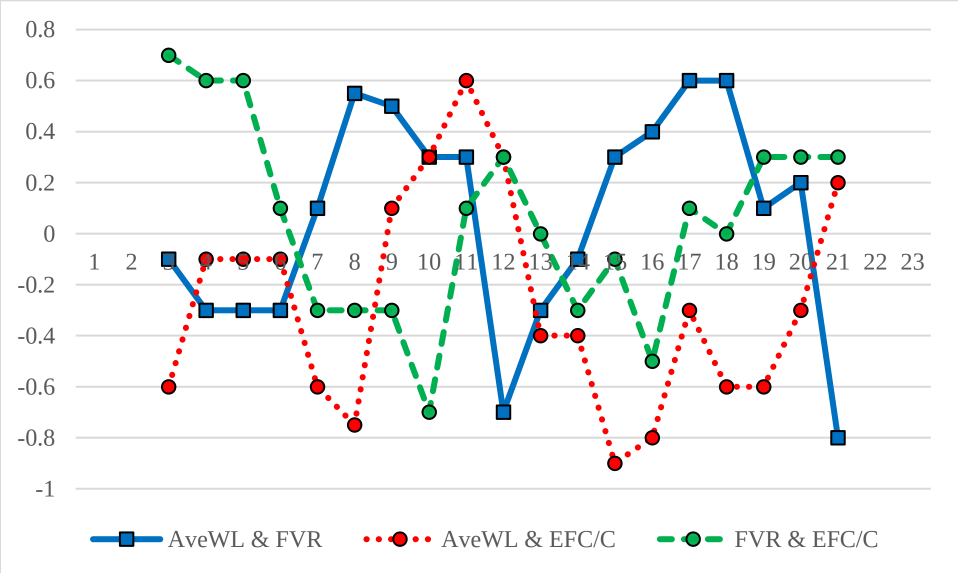


Figure 6.21 Moving correlations (General lexical and syntactic complexity and accuracy)

Table 6.3 shows the Spearman's rank correlation coefficient values between the general lexical and syntactic complexity and accuracy indices. None of the correlations was statistically significant.

Table 6.3

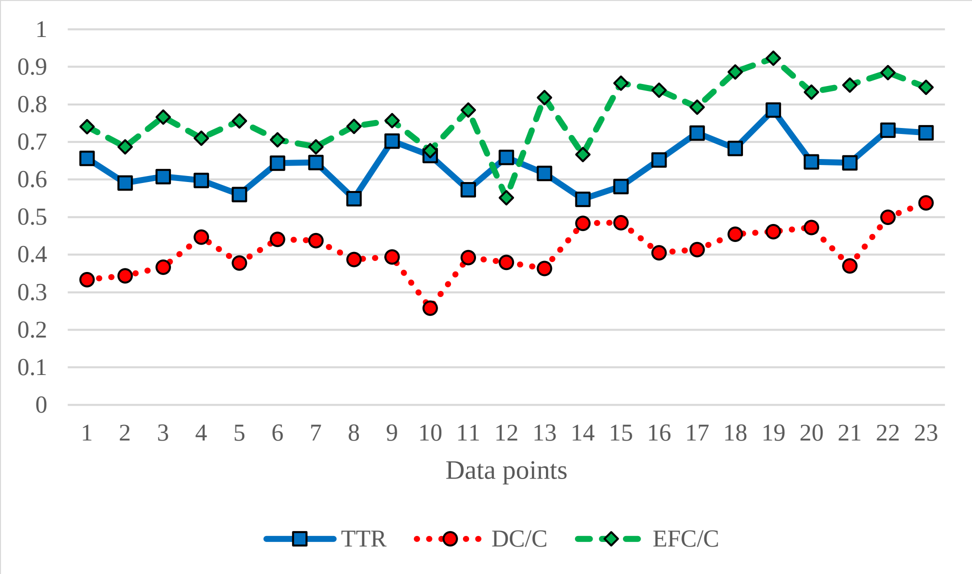
Correlation Matrix (General Lexical and Syntactic Complexity and Accuracy)

		AveWL	FVR	EFC/C
AveWL	ρ	1.000	.152	.171
	Sig. (2-tailed)		.488	.434
	N	23	23	23
FVR	ρ	.152	1.000	.135
	Sig. (2-tailed)	.488		.538
	N	23	23	23
EFC/C	ρ	.171	.135	1.000
	Sig. (2-tailed)	.434	.538	
	N	23	23	23

6.4.2 Interactions Between Lexical Variability, Subordination, and Accuracy

Figure 6.22 shows (top) the raw data values of the lexical variability (TTR), the subordination (DC/C) and the accuracy (EFC/C). Figure 6.22 shows that the growth trajectories of the lexical variability (TTR) and subordination (DC/C) indices were both alternating and parallel over the nine months. The trajectories of the lexical variability (TTR) and accuracy (EFC/C) indices were clearly parallel between data points 1 and 4. However, between data points 4 and 8 a clear alternating pattern is visible in the trajectories of the TTR and EFC/C indices. The trajectories of lexical variability (TTR) and accuracy (EFC/C) display both parallel and alternating shifts between data points 9 and 23. The growth trajectories of subordination (DC/C) and accuracy (EFC/C) were both parallel and alternating over the nine months. However, there were periods where clear patterns are visible. For example, parallel shifts can be observed between data points 8 and 12.

Figure 6.22 (bottom) shows the residual plot for the lexical variability (TTR), subordination (DC/C), and accuracy (EFC/C) indices. The residual plot shows both parallel and alternating shifts for the trajectories of the lexical variability (TTR) and subordination (DC/C). However, clear parallel shifts can be seen between data points 4 and 10. The residual plot shows both alternating and parallel variability patterns in the trajectories of the lexical variability (TTR) and accuracy (EFC/C) indices. However, there were clear parallel shifts between data points 1 and 4, and clear alternating patterns between data points 15 and 18. Likewise, the growth trajectories of the subordination (DC/C) and accuracy (EFC/C) indices display both parallel and alternating shifts over the nine months with a clear parallel variability pattern between data points 8 and 12, and clear alternating patterns between data points 3 and 6.



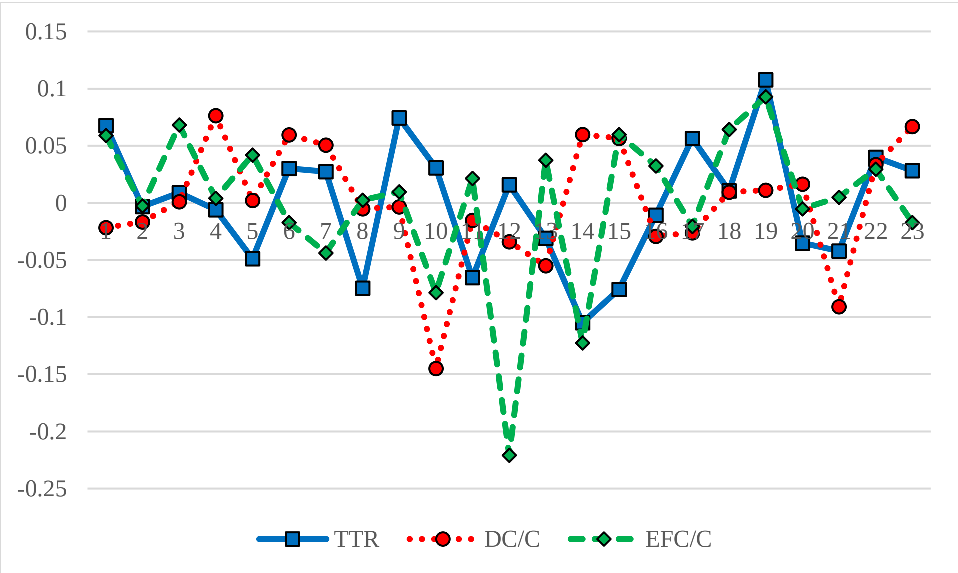


Figure 6.22 Lexical variability, subordination and accuracy (top: data values, bottom: residual plot)

Figure 6.23 shows the interactions between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 6.23 shows that the lexical variability (TTR)-subordination (DC/C) correlation shifted between moderately negative and moderately positive values, with one peak towards strongly negative value in window 12. The lexical variability (TTR)-accuracy (EFC/C) correlation also shifted between moderately negative and positive values with one peak towards a strongly negative value in window 6. The subordination (DC/C)-accuracy (EFC/C) correlation also shifted between moderately negative and moderately positive values, with one peak towards a strongly positive value in window 10. However, between window 8 and 21, the subordination-accuracy correlation was predominantly positive suggesting a positive relationship between the two indices.

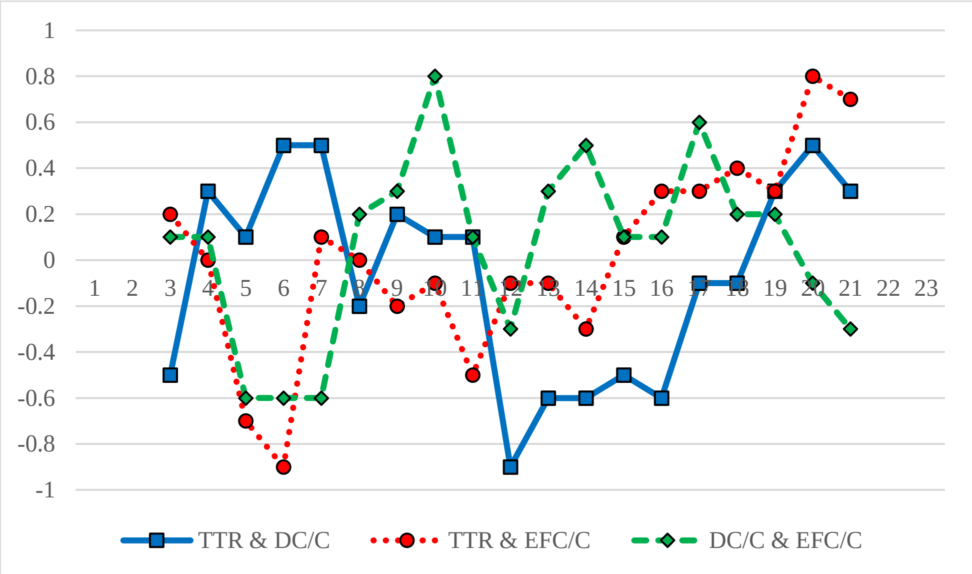


Figure 6.23 Moving correlations (Lexical variability, subordination, and accuracy)

Table 6.4 shows the correlation coefficient values between the lexical variability (TTR), subordination (DC/C) and the accuracy (EFC/C) indices.

Table 6.4

Correlation Matrix (Lexical Variability, Subordination, and Accuracy)

		TTR	DC/C	EFC/C
TTR	ρ	1.000	.230	.402
	Sig. (2-tailed)		.291	.057
	N	23	23	23
DC/C	ρ	.230	1.000	.427*
	Sig. (2-tailed)	.291		.042
	N	23	23	23
EFC/C	ρ	.402	.427*	1.000
	Sig. (2-tailed)	.057	.042	
	N	23	23	23

Note. *Correlation is significant at the 0.05 level (2-tailed)

The correlation coefficient between the lexical variability (TTR) and subordination (DC/C) was slightly positive (0.230) but not statistically significant. Likewise, the correlation coefficient between the lexical variability (TTR) and the accuracy (EFC/C) was also positive (0.402) and not statistically significant. Conversely, the correlation coefficient between the subordination (DC/C) and accuracy (EFC/C) indices was positive (0.427) and statistically significant ($p<0.05$).

In summary, both the raw data values and the residuals exhibited parallel and alternating patterns for lexical variability, subordination and accuracy indices. The moving correlations also confirmed the shifts between moderately negative and positive correlation coefficient values. However, the subordination (DC/C)-accuracy (EFC/C) correlation exhibited predominantly positive values in the moving correlation plot and their overall correlation coefficient value was positive and statistically significant.

6.5 Modelling

The purpose of this section is to test the hypothesised interactions between lexical variability (TTR) and subordination (DC/C), between lexical variability (TTR) and accuracy (EFC/C), and between subordination (DC/C) and accuracy (EFC/C). Based on the variability analyses (section 6.4.2) the following hypotheses were formulated:

1. weak support between lexical variability (TTR) and subordination (DC/C),
2. weak support between lexical variability (TTR) and accuracy (EFC/C),
3. weak support between subordination (DC/C) and accuracy (EFC/C).

To test the abovementioned hypotheses, a model of three connected growers was employed.

6.5.1 Model Setup

Three growers were specified in the model which corresponded with the three constructs under investigation in section 6.4.2 that is: (1) lexical variability (TTR), (2) subordination (DC/C), and (3) accuracy (EFC/C). The modelling procedure is the same as in Chapter 5 (see section 4.8.4 for a discussion). Table 6.5 shows the numerical values that were assigned to the parameters in the model.

Table 6.5

Property and Relational Parameters

		Grower A	Grower B	Grower C
Property	Initial value	0.657	0.3333	0.7407
parameters	Rate	0.01	0.01	0.01
	Carrying capacity	1	1	1
Relational		to A	to B	to C
parameters	from A	-	0.02	0.02
	from B	0.02	-	0.02
	from C	0.02	0.02	-

The second step included the configuration of the relational parameters. Table 6.5 also shows the relational parameters used in Emese's data.

After specifying the property and relational parameters, the data were simulated. The outcome of 40 iterations of the configured equations was obtained by the data simulation. Figure 6.24 shows the result of the simulations after 40 iterations.

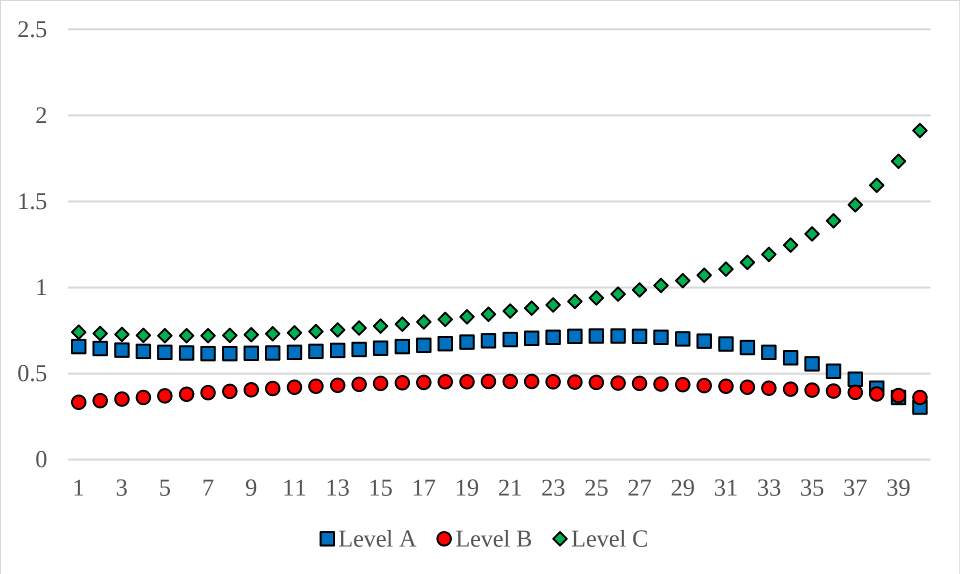


Figure 6.24 Growth model

6.5.2 Optimization

Before evaluating the model by correlating the model with the data, with the linear trends, and the cubic spline interpolations, the relational and property parameters were optimized. Table 6.6 shows the optimized parameter values.

Table 6.6

Optimized Parameter Values

		Grower A	Grower B	Grower C
Property parameters	Rate	-0.091285	0.15715	-0.07860
Relational parameters		to A	to B	to C
	from A	-	-0.0436	-0.0894
	from B	0.32092	-	0.20641
	from C	-0.1254	-0.0654	-

The optimization of the parameters made it possible to assess the overall fit of the model to the data. First, the sum of squared residuals between the model and the data was compared, which was 0.22053 in Emese's data. Second, the model outcome was visually contrasted with the data. Figure 6.25 shows that model resembles the data reasonably well.

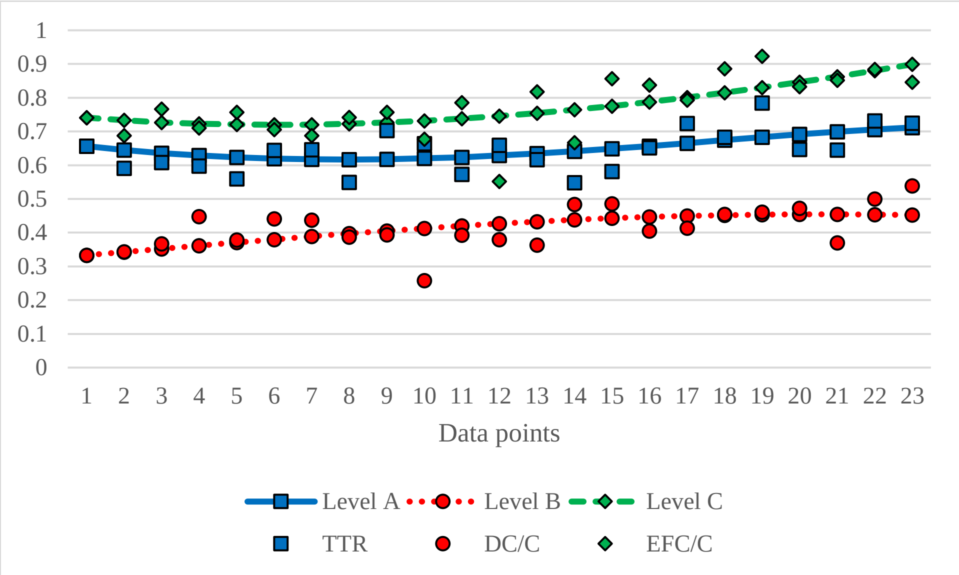


Figure 6.25 Data and model (fit = 0.22053)

The optimized parameter values confirm two of the three hypotheses that informed the model setup. The hypothesised supportive relationship between lexical variability (TTR) and subordination (DC/C) was confirmed by the model since the aggregated parameter value between Level A and B was positive (0.27736). However, the hypothesised supportive relationship between lexical variability (TTR) and accuracy (EFC/C) was not confirmed by the model because the aggregated relational parameter was negative (-0.2148) between Level A and C. The hypothesised supportive relationship between subordination (DC/C) and accuracy (EFC/C) was corroborated by the positive aggregated parameter value (0.14102) between Level B and C.

6.5.3 Evaluation of the Model

Figure 6.26 presents the data, the linear trends, the cubic spline interpolations, and model outcome. Visual inspection of these plots indicates that the model resembles a combination of the smoothing splines and the linear trends. The model was correlated with the data, the linear trends, and the smoothing splines.

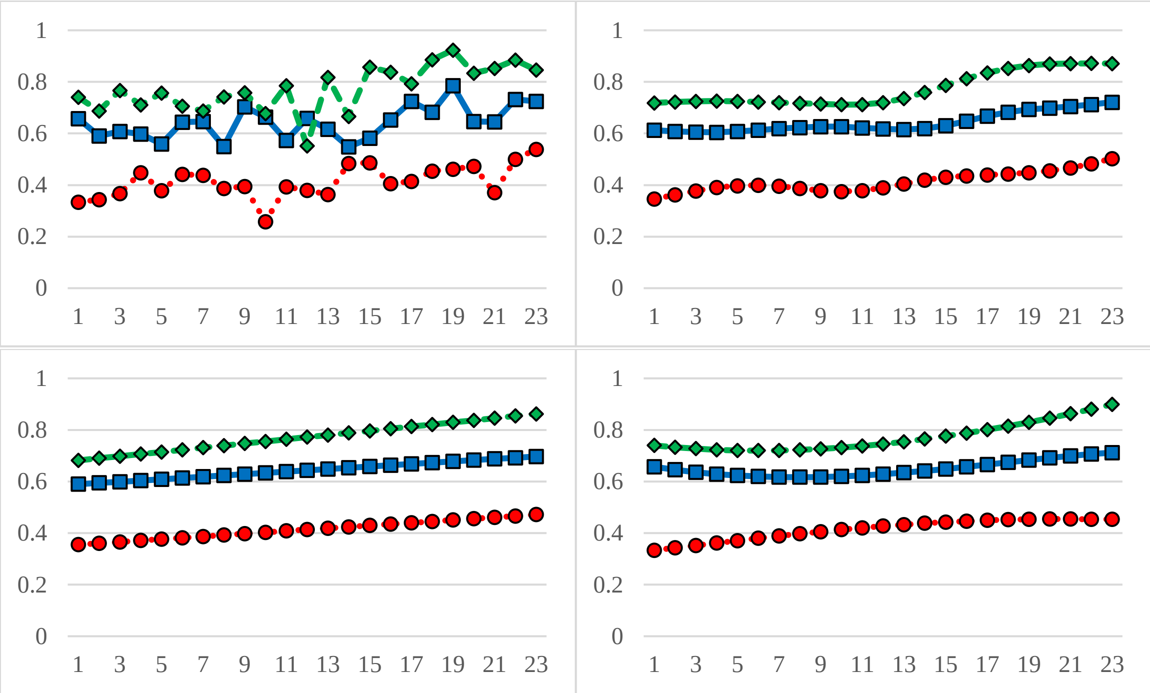


Figure 6.26 Data (top left), linear trends (top right), smoothing splines (bottom left) and model (bottom right)

Table 6.7 shows that the model positively correlated at all three levels with the original data.

Table 6.7

Correlations Between the Model and the Data, Linear Trends, and Splines

			Model-data	Model-linear	Model-spline
Level A	TTR	ρ	.473*	.700**	.648**
		Sig. (2-tailed)	.022	.000	.000
		N	23	23	23
Level B	DC/C	ρ	.559*	.987**	.879**
		Sig. (2-tailed)	.006	.000	.000
		N	23	23	23
Level C	EFC/C	ρ	.701**	.876**	.820**
		Sig. (2-tailed)	.000	.000	.000
		N	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed);** Correlation was significant at the 0.01 level (2-tailed).

In addition to the correlation between the model and the data, the linear trends, and the smoothing splines, another type of correlation analysis was performed to confirm the goodness-of-fit of the model. Correlation matrixes between the various categories in the data, the linear trends, the smoothing splines, and the model were compared to see whether the model replicated the surface level interactions of the data. Table 6.8 shows the correlation matrixes for the data, the linear trends, the smoothing splines, and the model.

Table 6.8

Correlation Matrixes

			Data	Linear	Spline	Model
TTR	DC/C	ρ	.225	1.000**	.758**	.687**
		Sig. (2-tailed)	.300	.000	.000	.000
		N	23	23	23	23
TTR	EFC/C	ρ	.402	1.000**	.625**	.925**
		Sig. (2-tailed)	.057	.000	.001	.000
		N	23	23	23	23
DC/C	EFC/C	ρ	.438*	1.000**	.906**	.863**
		Sig. (2-tailed)	.036	.000	.000	.000
		N	23	23	23	23

Note. *Correlation is significant at the 0.05 level (2-tailed);** Correlation is significant at the 0.01 level (2-tailed).

Table 6.8 shows that correlation matrixes in the data, the linear trends, the smoothing splines and the model outcome show very similar patterns, particularly those of the smoothing splines and the model outcome. The correlation coefficients between all combinations were positive.

6.6 Motivation

The purpose of this section is to explore the development of Emese's motivational system. This section is subdivided into four parts: (1) goals for academic writing, (2) self-regulatory processes, (3) attractor states, and (4) attitude.

6.6.1 Goals for Academic Writing

During the initial interview, Emese reported that she would not write in English in her future studies at university because the subjects would be taught in Hungarian. However, she added that she would like to study abroad for a couple of months in an Erasmus programme. Therefore, Emese wanted to take the IELTS exam. She also emphasised that passing the IELTS exam was her main goal during the EAP programme. Emese also reported that she would write essays, curriculum vitae and reports in her future career and occupation. Her main goal was to speak fluently with correct grammar in her future career. Her usual method of writing in English includes collecting ideas, ordering the ideas, and composing. Emese also added that she sometimes revises her essays, but she does not like reading her finished composition. During revisions, she said that she tries to focus on vocabulary, especially synonyms. She also added that she would like to improve her grammar in writing. Emese also added that she hates writing in English because she thinks that she is not good at it. However, she also added that she would try to change her negative attitude to writing.

6.6.2 Self-Regulatory Processes

Figure 6.27 shows that Emese tended to focus on self-observation and self-evaluation processes simultaneously over the course of investigation. Figure 6.27 also shows that the frequency of self-observation processes decreased, while the frequency of self-evaluation processes increased over time. Figure 6.27 also shows that Emese did not set any goals over the entire period of investigation. Although Emese’s interview data did not contain any goals, a shift can be observed from self-observation to self-evaluation at data point 9.

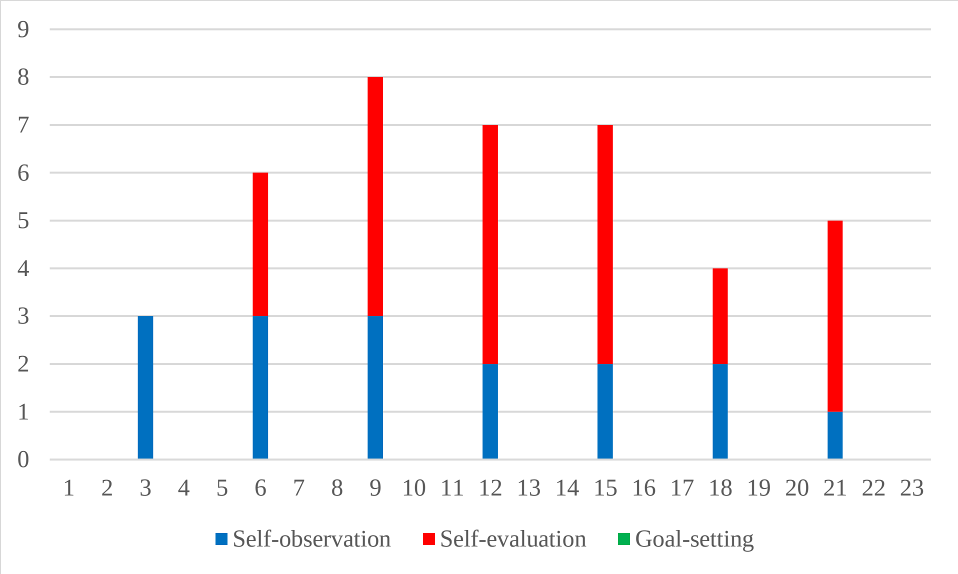


Figure 6.27 Self-regulatory processes

Table 6.9 provides a detailed breakdown of the range of Emese's self-regulatory processes over the nine months.

Table 6.9

Self-Regulatory Processes

		Data point							
		3	6	9	12	15	18	21	Total
Self-observation	Language	1	1	1	0	1	0	0	
	Composing processes	1	1	1	1	0	1	1	
	Content	1	1	1	1	1	1	0	
	Quality	0	0	0	0	0	0	0	
	Total	3	3	3	2	2	2	1	16
Self-evaluation	Language	0	1	2	3	1	1	2	
	Composing processes	0	0	0	0	1	0	1	
	Content	0	1	2	1	2	0	1	
	Quality	0	1	1	1	1	1	0	
	Total	0	3	5	5	5	2	4	24
Goal-setting	Language	0	0	0	0	0	0	0	
	Composing processes	0	0	0	0	0	0	0	
	Content	0	0	0	0	0	0	0	
	Quality	0	0	0	0	0	0	0	
	Total	0	0	0	0	0	0	0	0

The number of self-observation thematic units was evenly distributed (ranging between 1 and 3 thematic units) over time. However, the types of self-observation thematic units was not evenly distributed since Emese mainly focused on her language, composing processes, and the content of the essays. The number of self-evaluation thematic units was not evenly distributed over time. The number of self-evaluation thematic units started to increase at data point 6 and remained five thematic units between data points 9 and 15. However, a decrease can be observed in the number of self-evaluation thematic units from data point 18. Table 6.9 also shows that Emese did not set any goals over time.

The following examples (Interview excerpt 6.1-6.3) show that Emese's focus did not shift from self-observation to self-evaluation regarding the content of her essays over the nine months. Interview excerpt 6.1, 6.2 and 6.3 contain Emese's impressionistic descriptions of her actions containing somewhat superficial descriptions of her own writing. In addition, the specificity of Emese's descriptions of her L2 learning experience did not shift from a lack of tangibility to more elaborate and detailed explanations.

Interview excerpt 6.1 (Self-observation: content, data point 3)

Interviewer: How did you compose this text (Essay3)?
Emese: I just thought about the ideas. I decided to write about the native language teacher's positive aspects and then the negative.

Interview excerpt 6.2 (Self-observation: content, data point 9)

Interviewer: Can you tell me how you approached this piece of writing?
Emese: I tried to thinking about advantages and disadvantages

Interview excerpt 6.3 (Self-observation: content, data point 15)

Emese: At first, I was thinking about what to write and I knew that I won't disagree with the statement and then I tried to continue this essay.

However, the following examples (Interview excerpt 6.4 and 6.5) show that Emese's focus shifted from self-observation to self-evaluation. Emese concentrated on language in Interview excerpt 6.4 and 6.5. Interview excerpt 6.4 contains evaluative comments that involved some cognitive processes (comparing and reasoning).

Interview excerpt 6.4 (Self-evaluation: language, data point 6)

Interviewer:	How did you form the sentences?
Emese:	I don't know. I didn't do directly. I just write so that's why I think I'm not really improved.
Interviewer:	How about vocabulary?
Emese:	I think the first one's vocabulary was better.
Interviewer:	Why was it better?
Emese:	I don't know, more words, or advanced words.

Interview excerpt 6.5 also contains evaluative comments that involved cognitive processes (comparing and reasoning).

Interview excerpt 6.5 (Self-evaluation: language, data point 12)

Interviewer:	Compared to the previous pieces of writing do you think it was an improvement?
Emese:	Yeah, I think it was a little bit because I used some words maybe which are more advanced and I saw the first one (Essay 3) for example, and it was a simple essay. So there wasn't any interesting word.

6.6.3 Attractor States

The aim of this subsection was to identify the main attractor states that made up Emese's L2 learning experience and then to explore the development of these states. The first dominant state mentioned by Emese was boredom. She repeatedly expressed her boredom with the specific writing tasks and writing in general. The following excerpt (Interview excerpt 6.6) exemplifies her boredom or disinterest during the EAP course.

Interview excerpt 6.6 (data point 21)

Interviewer: What about the ideas themselves so was it easy for you to think of ideas for this topic?

Emese: Not really, I think I don't really know the other I think this was the worst… (referring to Essay 21) because it's not that interesting… I don't know it… it wasn't difficult it just not that interesting.

In the final interview, Emese reiterates that some of the topics were not interesting to her. She also pointed out that her interest in the topic affected the level of difficulty to collect ideas for a specific topic.

Interview excerpt 6.7 (data point 21)

Emese: There were one or two topic which was… not that interesting… because when I liked the topic… and I have a lot of ideas… and it always better… than if I'm bored… and I don't like to write that essay.

This subsection clearly shows that the boredom Emese experienced during one part of the EAP course might have had a detrimental effect on the development of her self-regulatory processes and consequently her L2 writing development.

6.6.4 Attitude

According to Csizér et al. (2010), attitude, goals and motivated behaviour are a closely interconnected co-adaptive systems. The evolution of Emese's self-regulatory process was probably hindered by her negative attitude to writing and her low self-efficacy belief. Emese constantly expressed her negative attitude to writing over time (Interview excerpt 6.8-6.10).

Interview excerpt 6.8 (data point 3)

Interviewer:	How do you feel when you write in English?
Emese:	I hate writing in English
Interviewer:	Why do you hate?
Emese:	I don’t know, I think because I’m not good at it… but I try to change this… because I hate that I hate writing in English.

Interview excerpt 6.9 (data point 9)

Interviewer:	Were you happy with the level of vocabulary that you were using?
Emese:	Not really… because I think it’s not my best writing… I don’t know, because I don’t really… want to write… anything… I don’t know… it’s just one day… it’s not a problem for me… but some days it’s a little problem.

Interview excerpt 6.10 (data point 15)

Interviewer:	So thinking about the language in this particular piece of writing, did you try out any new words or phrases that you haven’t used much before?
Emese:	I don’t think so. Not really… I couldn’t really focus on this topic… and you know I don’t really like writing… so I really hate.

However, Emese’s attitude to L2 writing showed fluctuations over time. For example, at data point 18, she expressed a more positive attitude to writing (Interview excerpt 6.11). Furthermore, her attitude to L2 writing might have been affected by her emotional factors experienced in her milieu.

Interview excerpt 6.11 (data point 18)

Interviewer	You mentioned previously that you hate writing in general. So do you think that this essay maybe it was a bit more interesting or?
Emese	Yes a little bit.
Interviewer	a little bit
Emese	or, or I don’t know. I just, it’s a good day, I don’t know.

However, Emese’s attitude to writing changed back from positive to negative at data point 21. Furthermore, she admitted that she also hated writing in her mother tongue.

Interview excerpt 6.12 (data point 21)

Interviewer:	Do you like writing in English?
Emese:	No, not really… because I don’t even like writing in Hungarian
Interviewer:	Why is it?
Emese:	I don’t know because I maybe speak a lot… and it is easier to me to speak… and not writing.

Interview excerpt 6.13 gives further evidence that emotional factors played an important role in Emese’s motivational system.

Interview excerpt 6.13 (data point 15)

Interviewer:	Well anything else that you would like to share with me in connection with this writing?
Emese:	No I don’t think so
Interviewer:	Emotional background?
Emese:	You know emotional background.

It was later discovered but was not recorded with any audio devices by the researcher that Emese broke up with his boyfriend on that day (at data point 15). It might be speculated that her emotional state influenced her motivational system, consequently her L2 writing on that day at data point 15. Indeed, at data point 15 the lexical variability (TTR) index was relatively lower (0.59) than in the other data points (see Figure 6.3). Furthermore, one of the highest subordination (DC/C) value was recorded at data point 15 (see Figure 6.10). This result might suggest that when Emese’s motivational system was negatively affected, she employed previously acquired syntactic structures such as dependent clauses instead of employing more phrasal structures. In dynamic parlance, the motivational system co-adapted with the linguistic system. Moreover, Emese’s data suggest that she could not acquire the characteristics of academic writing as suggested by Biber et al. (2011). She made her writing more complex by employing more dependent clauses instead of using more phrasal embeddings.

In addition to motivation, external factors such as time to learn might have contributed to the stagnation of linguistic and motivational systems, as illustrated in Interview excerpt 6.14:

Interview excerpt 6.14 (data point 21)

Interviewer:	Do you think your writing improved over the last six months?
Emese:	No… I don't really… think so… because I think… I haven't got enough time… to learn at home… check new words… and grammar.

Interview excerpts 6.8-6.14 illustrated that Emese's attitude, emotional factors, and external factors such as time to learn might have affected her L2 writing development.

6.7 Summary

All three constructs of linguistic complexity exhibited nonlinear developmental trends in Emese's written data over the nine-month period. No developmental peaks were detected in the lexical and syntactic complexity and accuracy indices over the nine months. The growth trajectories along with the residual plots, and the moving correlation plots indicated that the interactions between lexical and syntactic complexity and accuracy were dynamic. Both the magnitude and the polarity of these relationships fluctuated over the nine months. Furthermore, correlation coefficients were also calculated to explore the surface interactions between the three constructs. The model confirmed the weak supportive association between lexical variability (TTR) and subordination (DC/C) and the weak supportive relationship between subordination (DC/C) and accuracy (EFC/C). However, the weak supportive relationship between lexical variability (TTR) and accuracy (EFC/C) was not confirmed. The evolution of Emese's self-regulatory processes was also investigated and it was found that her focus shifted from self-observation to self-evaluation over time. Moreover, the interview data lacked the emergence of the third phase of self-regulation: goal-setting. Therefore, attractor states were identified in her interview data, and it was found that boredom might have contributed to the stagnation of self-regulatory processes and consequently of her L2 writing. Furthermore, Emese's negative attitude might have caused to the stagnation of her L2 writing development.

Chapter 7. Case Study Three

Chapter 7 presents the changes in the lexical and syntactic complexity and accuracy indices in Levente's written data. Furthermore, the changes in his self-regulatory processes are presented. There are six main sections: (1) the participant and the context, (2) the developmental profile, (3) the degree of variability, (4) interactions, (5) modelling, and (6) motivation.

7.1 The Participant and the Data

Levente (pseudonym), a 22-year-old third-year Bachelor of Engineering student, was the third participant of this study. His mother tongue is Hungarian, and he had been learning English for 10 years at the beginning of the data collection. Levente's English language education started at the age of 12 at a prestigious secondary school in Budapest. He attended three English classes on a weekly basis over six years. In grade 10 Levente took The European Language Certificates language test which placed him at a B2 CEFR level. Therefore, it was not compulsory for him to attend English classes at his secondary school. Consequently, Levente had not learned English between grade 10 and 12. At the end of the secondary school he successfully took the final exam in English at medium level.

During his Bachelor studies he did not engage in formal language education apart from occasional private English language classes. Levente wanted to take the IELTS language exam because he wanted to obtain an MA degree in Engineering. Some courses are offered in English at his university. Therefore, he enrolled in the EAP course and attended 2-hour long classes on a weekly basis.

The data for this case study were collected using the same procedure outlined in Chapter 4. Table 7.1 presents Levente's written corpus, interviews, and timeline.

Table 7.1

Written Corpus, Interviews, and Timeline

Data point	Code	Writing prompt	Date of submission	Word count	Interview Type	Interview Duration
1	Essay 1	WP1	19/09/2014	233		
2	Essay 2	WP2	03/10/2014	214		
3	Essay 3	CWP4	17/10/2014	320	Initial	5:14
					Retrospective 1	2:21
4	Essay 4	WP3	31/10/2014	239		
5	Essay 5	WP4	07/11/2014	252		
6	Essay 6	CWP5	14/11/2014	264	Retrospective 2	2:50
7	Essay 7	WP5	28/11/2014	244		
8	Essay 8	WP6	05/12/2014	244		
9	Essay 9	CWP3	19/12/2014	279	Retrospective 3	7:52
10	Essay 10	WP7	19/12/2014	280		
11	Essay 11	WP8	09/01/2015	243		
12	Essay 12	CWP6	23/01/2015	302	Retrospective 4	7:34
13	Essay 13	WP9	30/01/2015	246		
14	Essay 14	WP10	06/02/2015	230		
15	Essay 15	CWP2	27/02/2015	274	Retrospective 5	6:21
16	Essay 16	WP11	27/02/2015	229		
17	Essay 17	WP12	06/03/2015	260		
18	Essay 18	CWP1	20/03/2015	275	Retrospective 6	8:21
19	Essay 19	WP13	27/03/2015	279		
20	Essay 20	WP14	03/04/2015	254		
21	Essay 21	CWP4	03/04/2015	292	Retrospective 7	6:50
					Final	15:18
22	Essay 22	WP15	17/04/2015	262		
23	Essay 23	WP16	08/05/2015	314		
		Total word count		6029	Total duration	1:02:41

Note. Controlled data points are highlighted. WP = Writing prompt; CWP = Controlled writing prompt.

7.2 The Developmental Profile

This section is a detailed description of the developmental profiles of Levente's academic writing pertaining to complexity and accuracy. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity, and (3) accuracy.

7.2.1 Lexical Complexity

Figure 7.1 shows that all five indices developed nonlinearly and fluctuated to different degrees over the nine months. As in the previous two case studies, the general lexical complexity (AveWL) and Academic Word List (AWL) indices show similar patterns, while the lexical variability (TTR), lexical rarity (WRDFRQc) and lexical disparity (LSA) indices exhibit different trajectories.

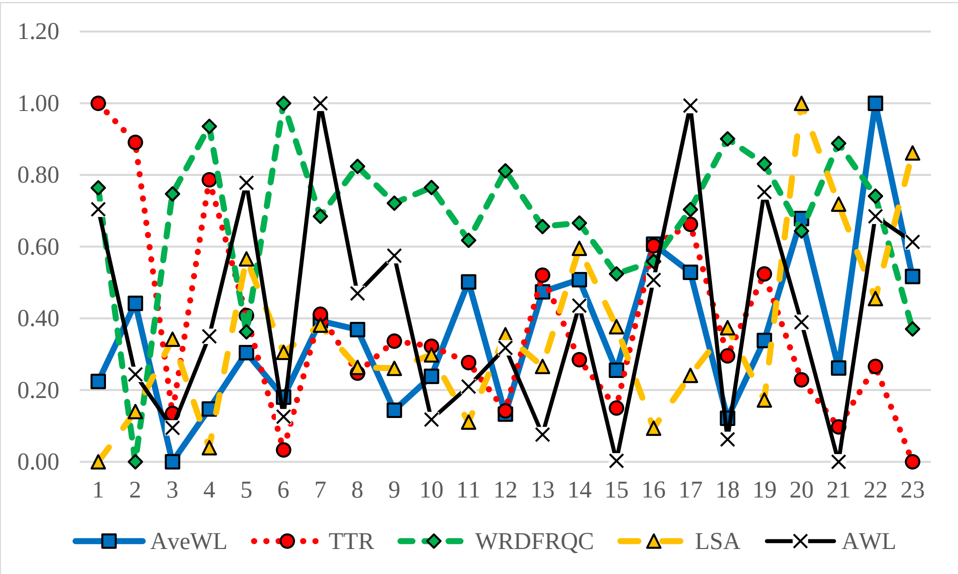


Figure 7.1 Lexical complexity

The average word length (AveWL) index was used to measure general lexical complexity. Figure 7.2 shows that the AveWL index fluctuated between 4.01 and 5.35. The lowest AveWL index was measured at data point 3 (AveWL = 4.01), while the highest AveWL value was gauged at data point 22 (AveWL = 5.35). The smoothing algorithm of the general lexical complexity (AveWL) index displays an upward trend over the nine-month investigation.

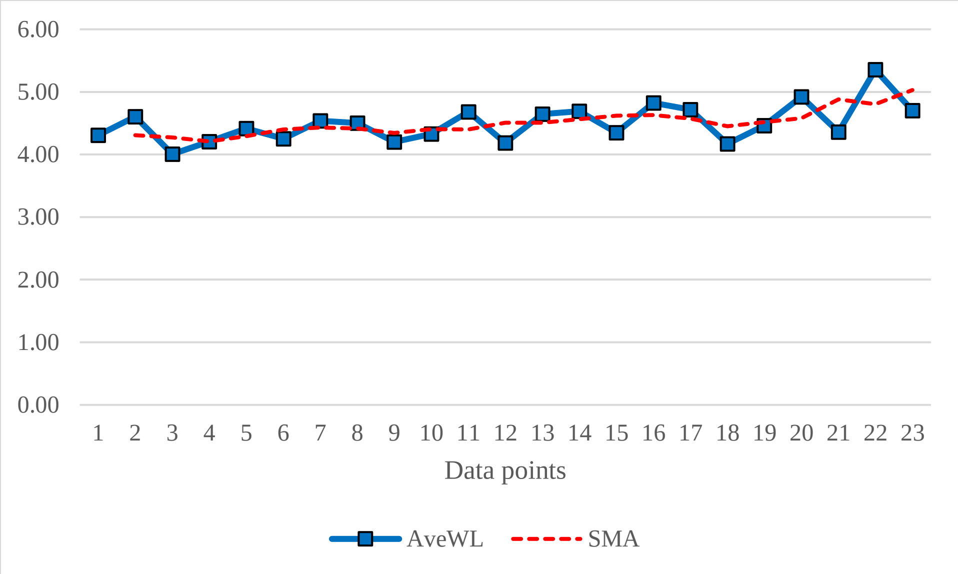


Figure 7.2 General lexical complexity (AveWL)

Figure 7.3 shows that the lexical variability (TTR) index fluctuated between 0.52 and 0.79 over the nine months. From data point 1 to 6 a gradual decrease can be observed in the lexical variability (TTR) index. However, from data point 7 to 23, the TTR index fluctuated slightly without a discernible trend.

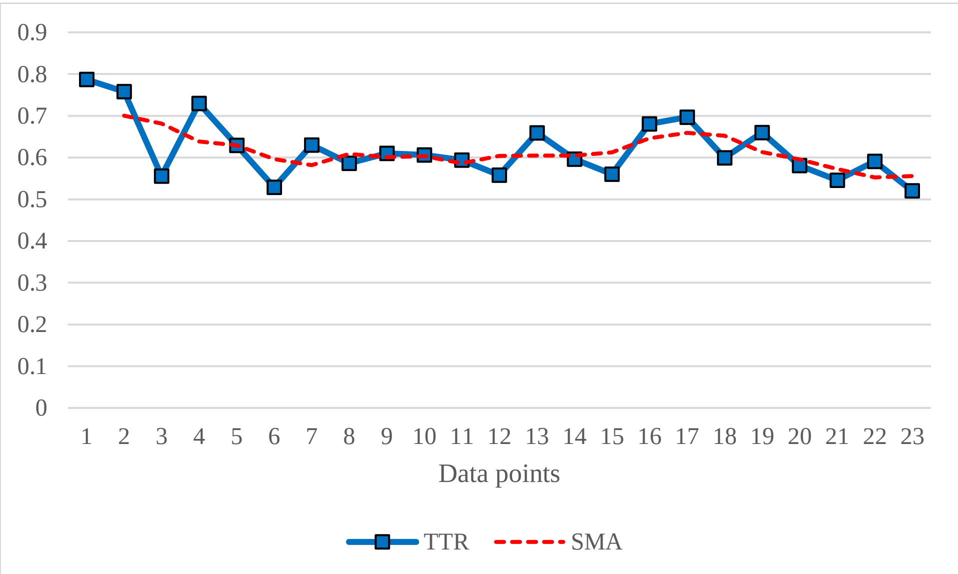


Figure 7.3 Lexical variability (TTR)

The TTR index started from its highest point at data point 1 (TTR = 0.79) and reached its nadir at data point 23 (TTR = 0.52). The smoothing algorithm of the lexical variability (TTR) index displays a downward trend between data points 2 and 7 and between data points 18 and 23. However, the smoother of the TTR index shows a sideways trend between data points 8 and 17.

The second property of lexical diversity, rarity, was measured by the CELEX log minimum frequency of content words index (WRDFRQc) and shown in Figure 7.4. The lexical rarity (WRDFRQc) index shows more fluctuations, especially between data points 1 and 7, than the general lexical complexity (AveWL) and the lexical variability (TTR) indices over the nine months.

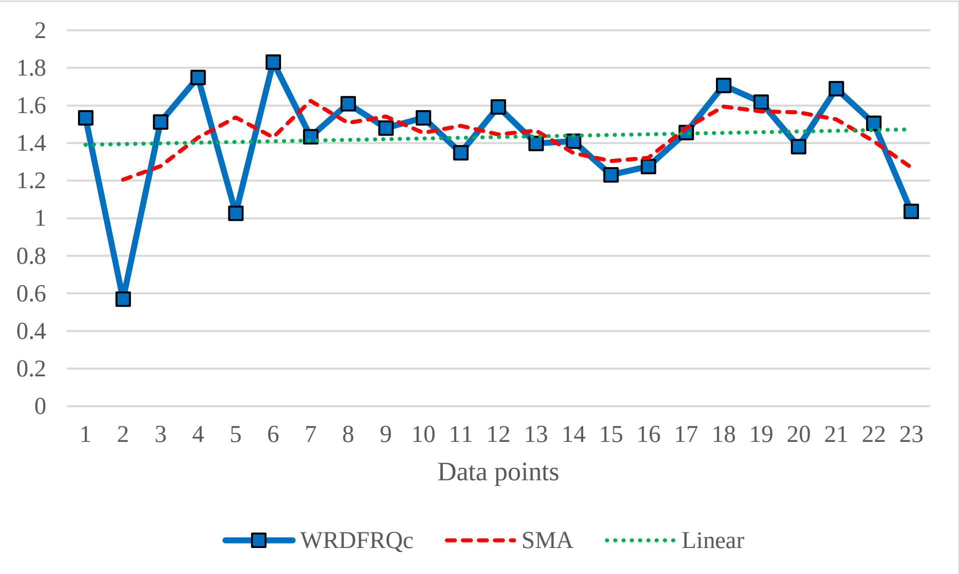


Figure 7.4 Lexical rarity (WRDFRQc)

The highest lexical rarity (WRDFRQc) value (WRDFRQc = 1.83) was measured at data point 6, while the lowest WRDFRQc value (0.57) was detected at data point 2. The lexical rarity (WRDFRQc) was relatively low at data point 2, 5, and 23 which indicated that the Essay 2, Essay 5, and Essay 23 included less frequent words compared to the rest of the essays. The smoothing algorithm of the lexical rarity (WRDFRQc) index displays a sideways trend over the nine-month investigation. However, the linear trend line of the WRDFRQc index shows a slightly upward trend over the nine months.

The third property of lexical diversity, disparity, was measured by the latent semantic analysis index (LSA). Although the LSA index (Figure 7.5) fluctuated wildly, a small increase can be observed over the nine months. The lowest lexical disparity (LSA) value (LSA = 0.07) was measured at data point 1, while the highest LSA value (LSA = 0.47) was detected at data point 20. The smoothing algorithm of the lexical disparity (LSA) index displays an upward trend over the nine-month investigation.

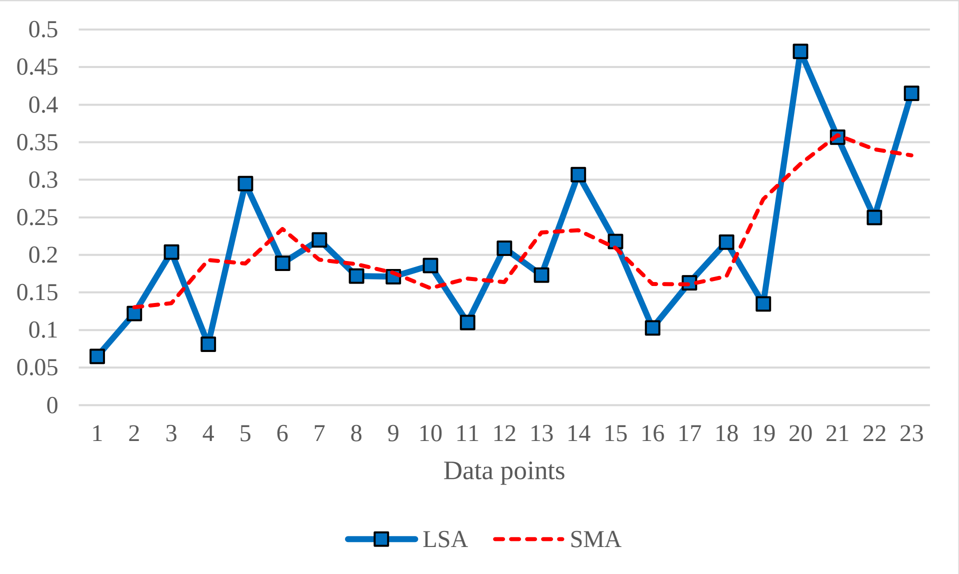
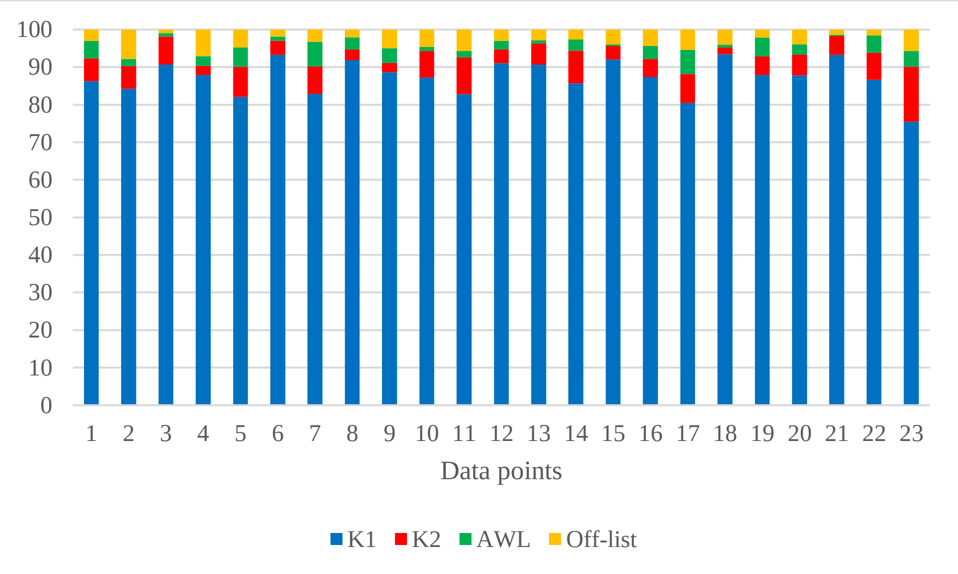


Figure 7.5 Lexical disparity

To investigate the extent of genre-relevant lexical choice in Levente's academic writing, the percentage of academic words in his essays was estimated by the Academic Word List (AWL) measure. Figure 7.6 (top) shows the four frequency bands: K1 is the list of the most frequent 1000-word families, K2 is the list of the second most frequent 1000-word families, the Academic Word List, and off-list words. Figure 7.6 (top) shows that the highest percentage of K1 words was measured at data point 18 (K1 words = 93.45%), while the lowest percentage of K1 words was detected at data point 23 (K1 words = 75.48%). The low percentage of K1 words at data point 23 was attained by the increase of the percentages of K2 words. Indeed, one of the highest percentages of K2 words (K2 words = 14.65%) was detected at data point 23. In order to gain a better understanding of the use of the words from the Academic Word List, a separate graph was plotted in Figure 7.6 (bottom).



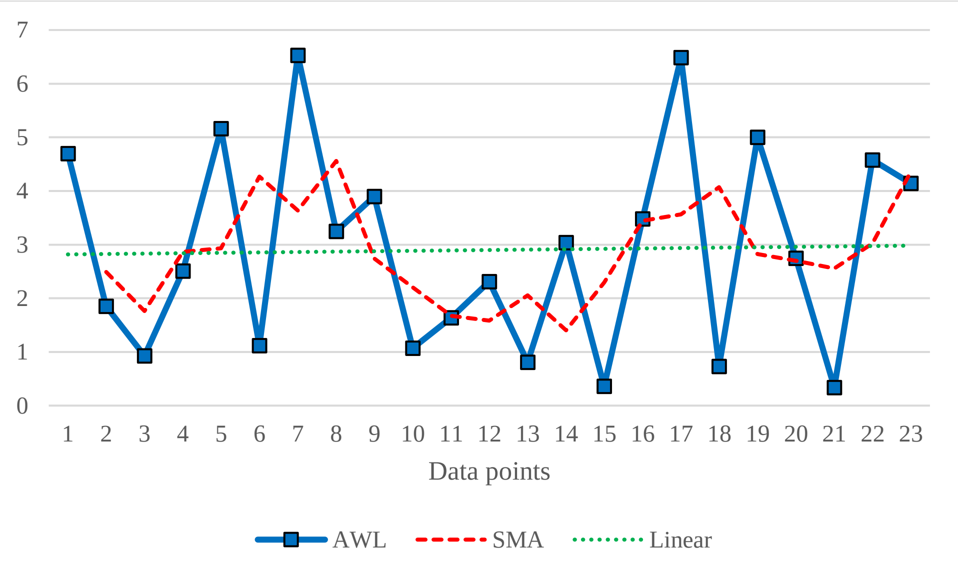


Figure 7.6 Lexical sophistication (top), and Academic Word List (bottom)

The AWL index shows wild fluctuations over the nine months. The AWL index reached its zenith (AWL = 6.53%) at data point 7, while the lowest point (AWL = 0.34%) was measured at data point 21. The smoothing algorithm of the Academic Word List (AWL) index displays a sideways trend over the nine-month investigation. However, the linear trend line of the AWL index displays a slightly upward trend over the nine months.

7.2.2 Syntactic Complexity

Figure 7.7 displays the five syntactic indices used in this study. It can be observed that the FVR and the CN/C indices display similar trajectories since they tap into a similar construct, phrasal complexity. Figure 7.7 also shows that subordination (DC/C) increased over the nine months, whereas the general syntactic complexity (FVR) and phrasal complexity (CN/C) displayed discernible patterns.

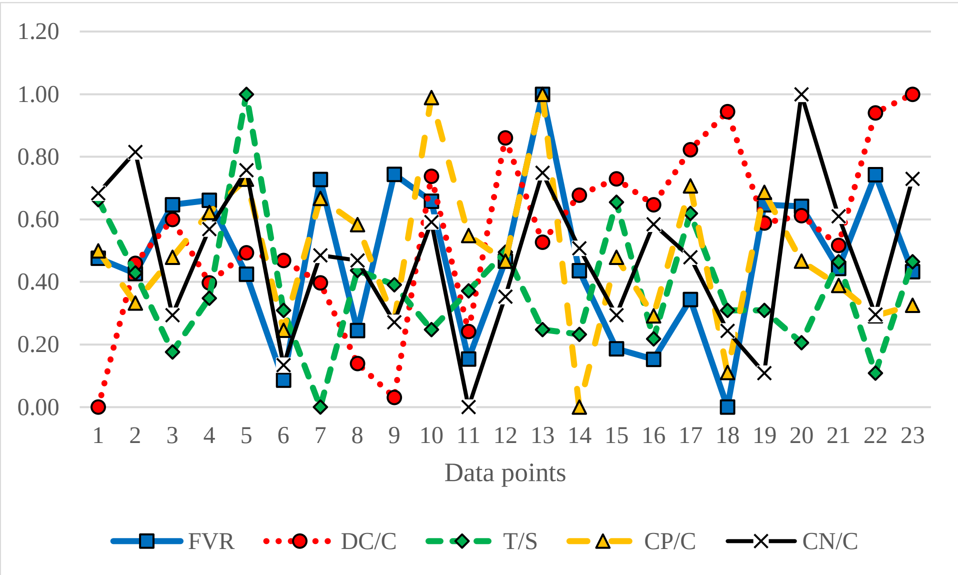


Figure 7.7 Syntactic complexity

General syntactic complexity was measured by the finite verb ratio (FVR) in Levente's written samples. The FVR index, shown in Figure 7.8, was relatively stable between data points 1 and 5, but then it started to fluctuate between data points 6 and 18. However, between data points 19 and 23 the general syntactic complexity (FVR) index was once again relatively stable. The lowest FVR value (FVR = 6.55) was measured at data point 18, while the highest value (FVR = 9.11) was gauged at data point 13. The smoothing algorithm of the general syntactic complexity (FVR) index displays a sideways trend over the nine-month investigation. However, the linear trend line of the FVR index shows a slightly downward trend over the nine months.

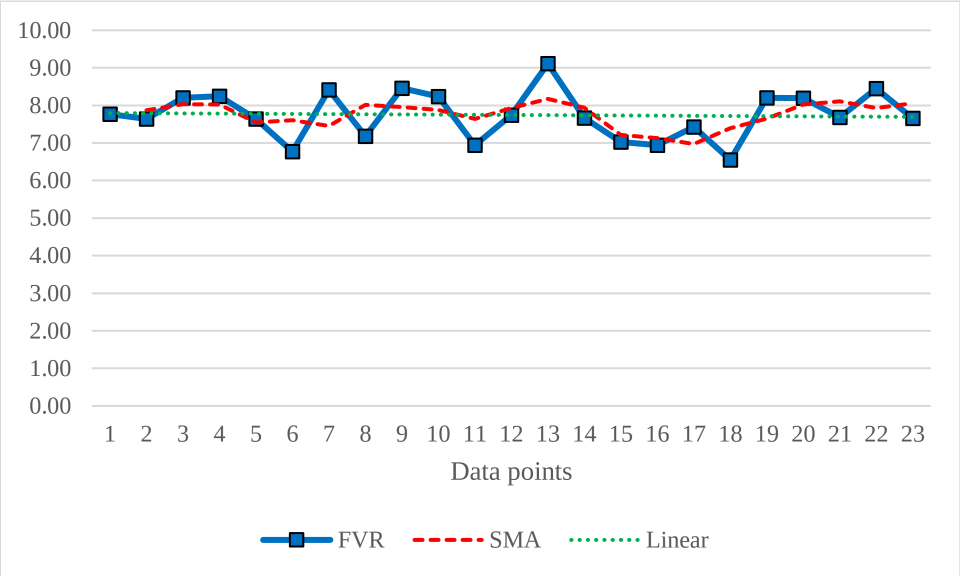


Figure 7.8 General syntactic complexity (FVR)

The following examples, taken from Levente's Essay 18 and Essay 13, illustrate the difference in the general syntactic complexity (FVR) index. Essay excerpt 7.1 was extracted from Essay 18 where the FVR was the lowest (6.55). In contrast, Essay excerpt 7.2 was taken from Essay 13 where the FVR index was the highest (9.11). The finite verbs are underlined, and the example sentences are left uncorrected.

Essay excerpt 7.1 (Essay 18)

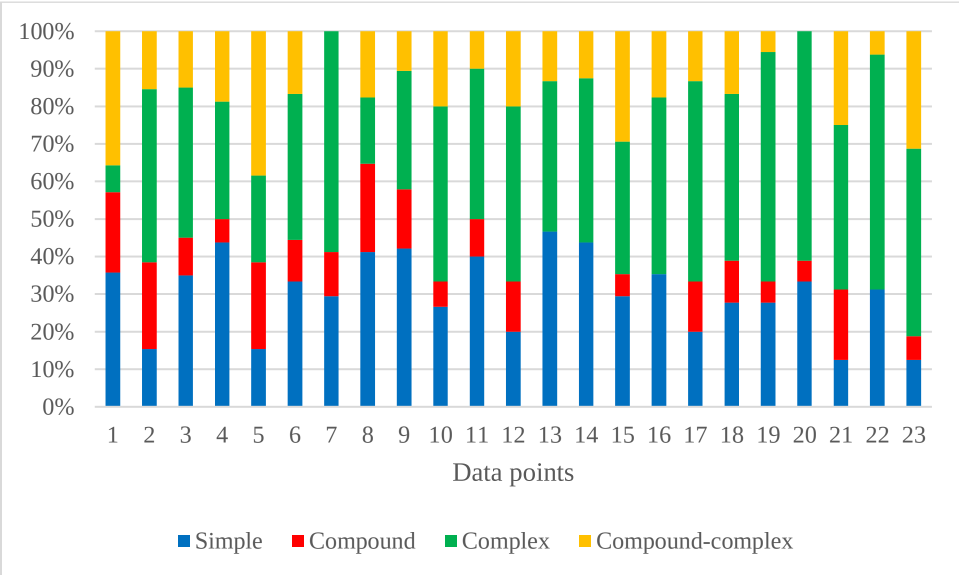
Learning a language [1]<u>is</u> not just depending on that where you [2]<u>live</u>.
(Errors are uncorrected; finite verb is underlined and numbered in superscript)
(Word count = 12; finite verb count = 2; FVR = 6)

Essay excerpt 7.2 (Essay 13)

However, it [1]<u>is</u> a completely great idea because students [2]<u>can</u> earn a lot of experiences during a year of travel or work.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 22; finite verb count = 2; FVR = 11)

The two sentences exhibit different levels of syntactic complexity, although they were the same type of sentence i.e. complex. The ratio of the words per finite verbs clearly indicates that Essay excerpt 7.2 (FVR = 11) is syntactically more complex than Essay excerpt 7.1 (FVR = 6).

Levente's written samples were also coded for sentence types in order to gain a better insight into the syntactic structures he used in his written data. The distribution of the different sentence types is presented in a bar chart in Figure 7.9.



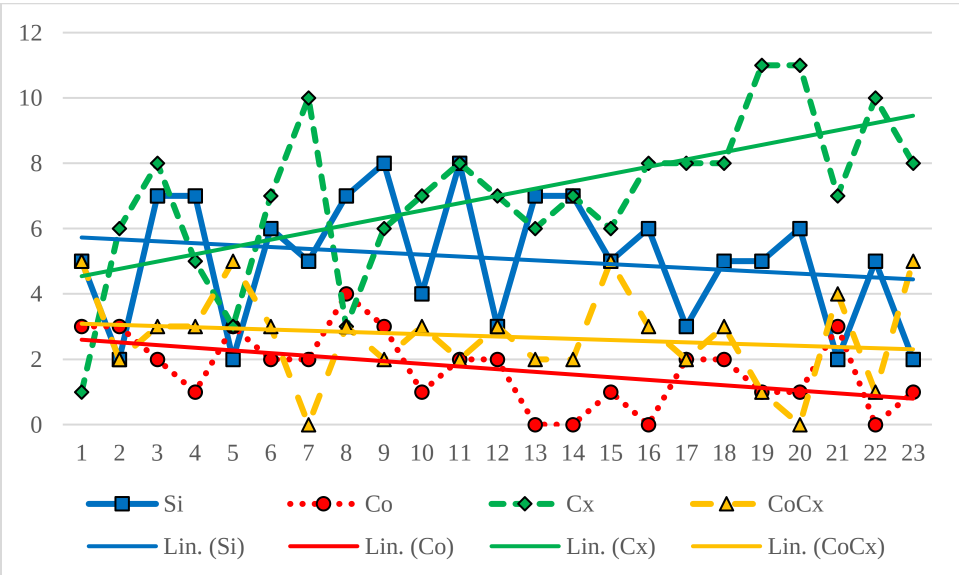


Figure 7.9 Sentence types (top: distribution; bottom: trend)

Figure 7.9 (top) shows the distribution of sentence types. Visual inspection of Figure 7.9 indicates that complex sentences dominated in Levente's written data. Indeed, Levente's written corpus contained 161 complex, 117 simple, 62 complex-compound and 39 compound sentences. Figure 7.9 (bottom) also shows that the instances of the complex sentences increased over the nine months, whereas the instances of the simple sentences decreased, along with compound and complex-compound sentences. This phenomenon does not confirm the current findings of academic writing which suggest that advanced users use more complex nominalisation rather than clausal embeddings in academic writing (Biber & Gray, 2010). Furthermore, the increase of complex sentences in Levente's data is different from Dalma's data, in which an increase in the use of simple sentences and a decrease in the use of complex sentences can be observed. However, the increase of complex sentences in Levente's data is similar to Emese's data in which the use of complex sentences increased over the nine months. Simple sentences were present in every essay, while compound sentences were absent in Essay 13, Essay 14, Essay 16 and Essay 22. Not surprisingly, complex sentences were detected in each essay, while complex-compound sentences were non-existent in Essay 7 and 20.

Due to the preponderance of the complex sentences in Levente's written data, one would expect a correspondingly high value of subordination. Indeed, Figure 7.10 shows that the subordination (DC/C) index fluctuated slightly around 0.4 which suggests that virtually every second clause was dependent.

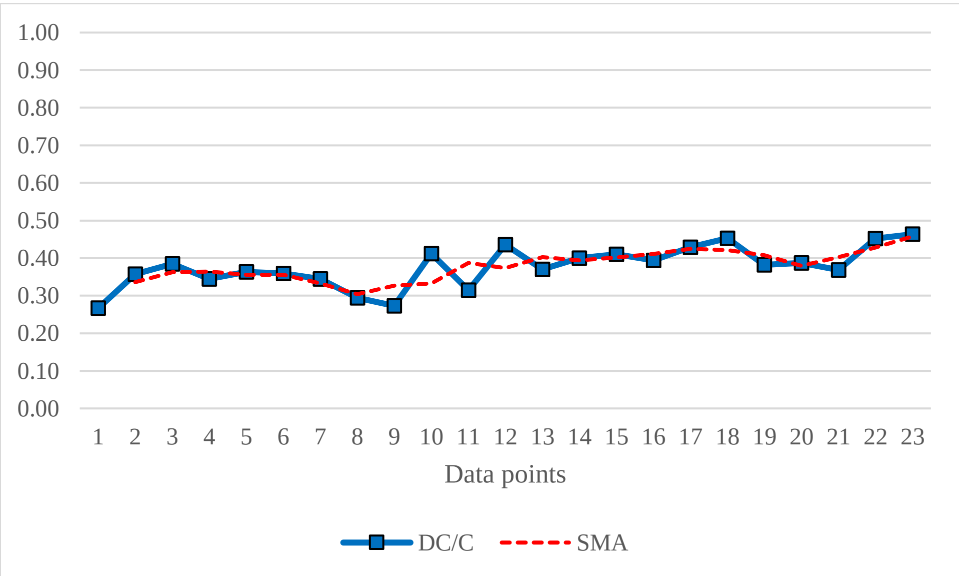


Figure 7.10 Subordination

The subordination (DC/C) index started from its lowest point (0.27) at data point 1 and reached its highest point (0.46) at data point 23. The smoothing algorithm of the subordination (DC/C) index displays an upward trend over the nine-month investigation. The following example, Essay excerpt 7.3, illustrates the relatively high DC/C value at data point 23 where the highest DC/C value (0.46) was measured.

Essay excerpt 7.3 (Essay 23)

All in all I think [1]that popular and fascinating events like the football world cup are good [2]because they can release the tension between countries.
(Errors are uncorrected; dependent clauses are underlined and numbered in superscript)
(Dependent clause count = 2; clause count = 3; DC/C = 0.67)

In addition to subordination, Levente also employed both clausal and phrasal coordination in his essays. The clausal coordination (T/S) index, shown in Figure 7.11, displays slight fluctuations over the nine months.

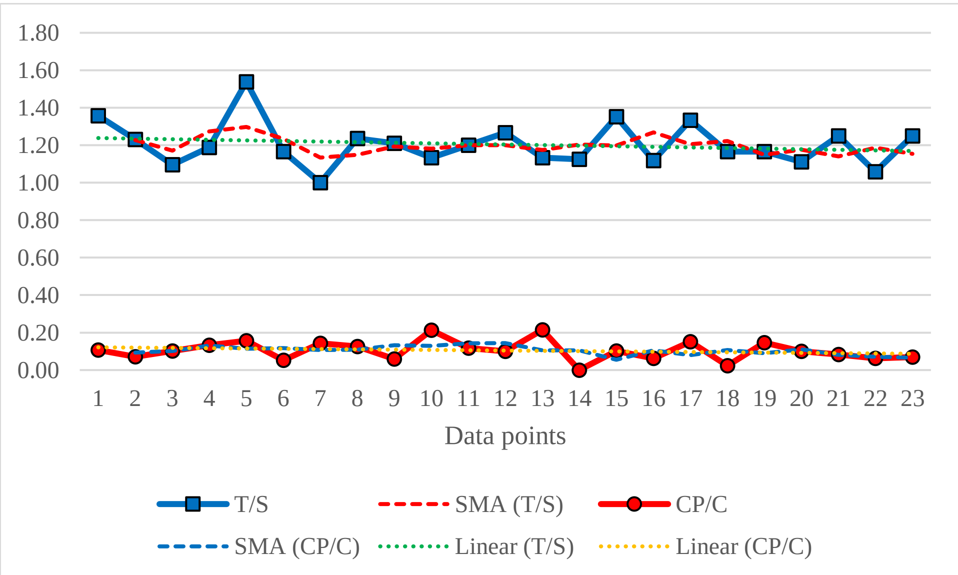


Figure 7.11 Coordination

The lowest clausal coordination (T/S) value (T/S = 1) was measured at data point 7, while the highest T/S value (T/S=1.54) was detected at data point 5. The smoothing algorithm of the clausal coordination (T/S) index (red dashed line) displays sideways trend over the nine months. However, the linear trend line of the T/S index (green dotted line) shows a slightly downward trend over the nine months. In Essay 5, the ratio of the number of T-units to the number of sentences was highest (1.54), which means that there were approximately one and half T-units in each sentence. This finding was not surprising since the proportion of the compound-complex sentences was the highest in Essay 5 (see Figure 7.9).

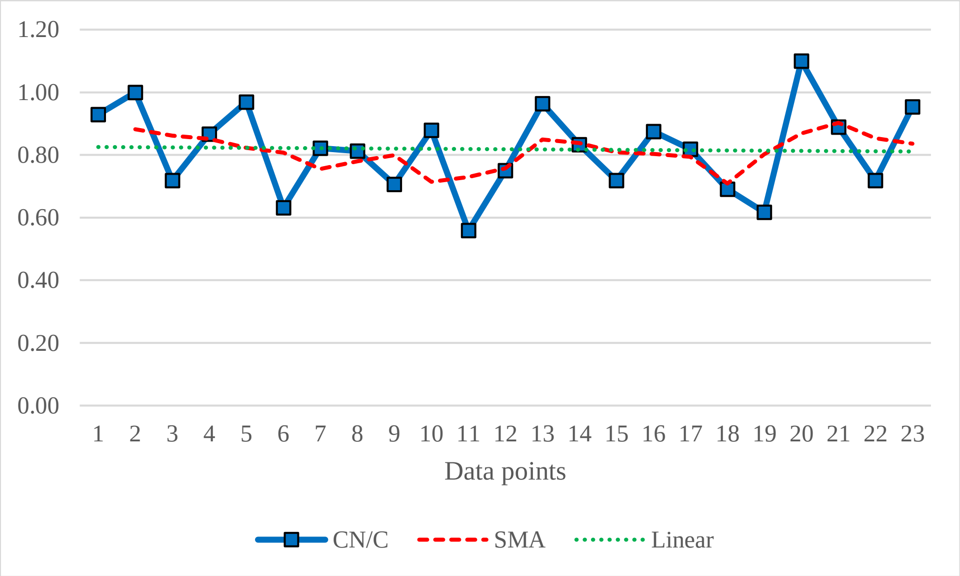
In addition to clausal coordination, Levente also employed phrasal coordination to make his sentences longer, consequently making his writing more complex. Figure 7.11 shows that the CP/C index reached its zenith at data point 13 (CP/C = 0.21). This finding suggests that on average one coordinate phrase was used for every four clauses in Essay 13, as illustrated in Essay excerpt 7.4.

> Essay excerpt 7.4 (Essay 13)
>
> In some [1]American or English countries young people are encouraged [2]to work or travel for year between [3]finishing high school and starting university studies.
> (Errors are uncorrected; coordinate phrases are underlined and numbered in superscript)
> (Coordinate phrase count = 3; clause count = 1; CP/C = 3)

The smoothing algorithm of the phrasal coordination (CP/C) index (blue dashed line) shows a sideways trend over the nine-month investigation. Likewise, the linear trend of the CP/C index (yellow dotted line) shows a sideways trend over the nine months.

Phrasal complexity was measured by the complex nominal per clause (CN/C) and the mean number of modifiers per noun phrase (SYNNP) indices. Figure 7.12 (top) shows that the CN/C index fluctuated wildly as opposed to the other syntactic indices which remained constant. The lowest CN/C index value (CN/C = 0.56) was measured at data point 11, while the highest value (CN/C = 1.10) was gauged at data point 20. The smoothing algorithm of the phrasal complexity (CN/C) index displays a sideways trend over the nine-month investigation.

However, the linear trend line of the CN/C index shows a slightly downward trend over the nine months.



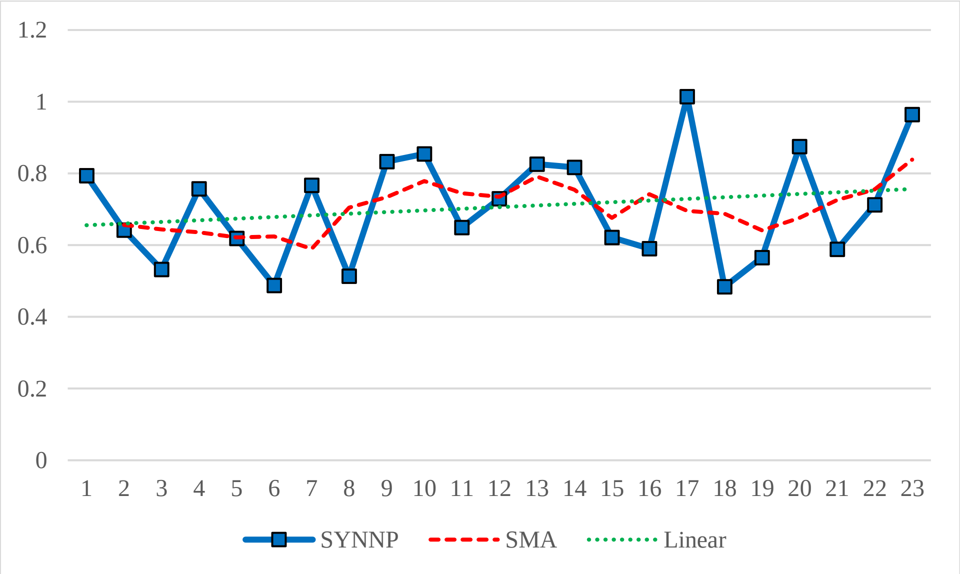


Figure 7.12 Phrasal complexity (CN/C and SYNNP)

Figure 7.12 (bottom) also shows the mean number of modifiers per noun phrase index. The lowest SYNNP index (SYNNP = 0.484) was measured at data point 18, while the highest SYNNP index (SYNNP = 1.014) was gauged at data point 17. The smoothing algorithm of the SYNNP index displays a sideways trend over the nine-month investigation. However, the linear shows a clear upward trend over the nine months.

Essay excerpt 7.5 exemplifies the high phrasal complexity (CN/C) index (CN/C = 1.10) at data point 20.

Essay excerpt 7.5 (Essay 20)

Music played an [1]incredibly important part of the [2]people's everyday life for many [3]hundreds of years.
(Errors are uncorrected; complex nominals are underlined and numbered in superscript)
(Complex nominal count = 3; clause count = 1; CN/C = 3)

Figure 7.13 shows the normed rate of occurrence of conditional clauses (ConC), the normed rate of occurrence of infinitive clauses (InfC), the normed rate of occurrence of relative clauses (RelC), and the incidence score of prepositional phrases (DRPP) indices.

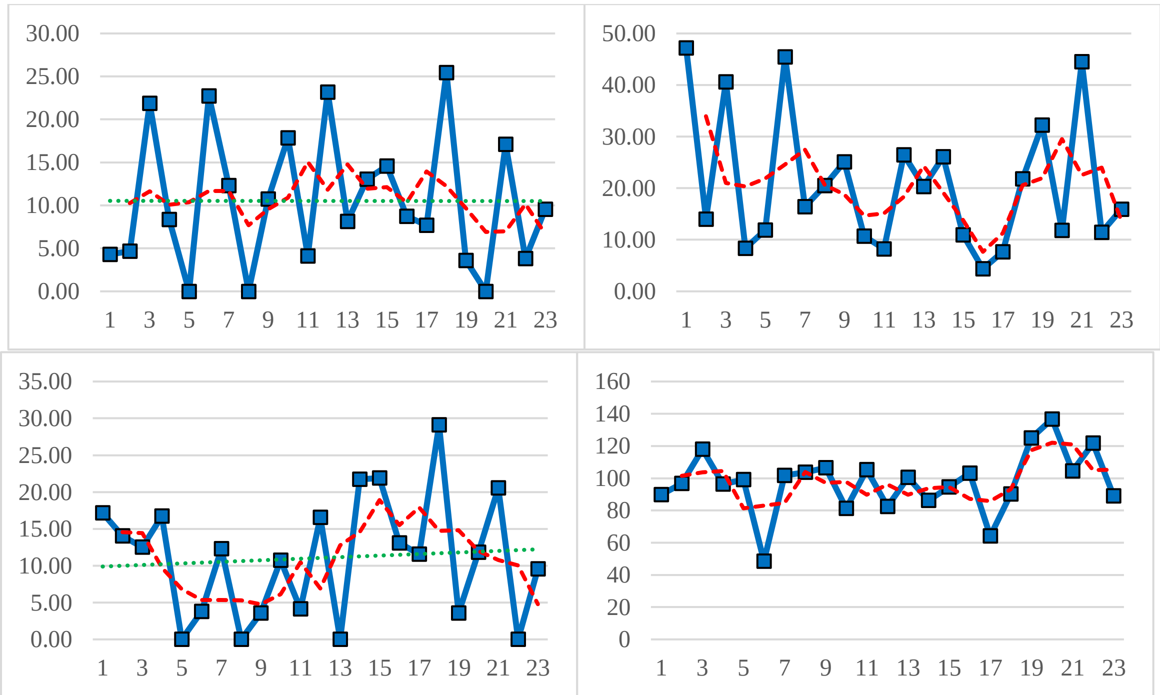


Figure 7.13 Syntactic measures specific to the academic genre (top left: ConC; top right: InfC; bottom left: RelC; bottom right: DRPP).

The smoothing algorithms of the conditional clauses (ConC) and the relative clauses (RelC) indices display sideways trends over the nine-month investigation. However, the linear trend line of the relative clauses (RelC) index shows a slightly upward trend whereas the linear trend line of the conditional clauses (ConC) index shows a sideways trend. The smoothing algorithm of the infinitive clauses (InfC) index shows a downward trend over time. Conversely, the smoothing algorithm of the prepositional phrases (DRPP) index displays an upward trend over the nine-month investigation. In other words, Levente tended to use more relative clauses and prepositional phrases in his written data over the nine-month investigation. Conversely, Levente started to use fewer infinitive clauses in his essays, while there were no changes in the use of conditional clauses.

7.2.3 Accuracy

In addition to the constructs of syntactic and lexical complexity, accuracy was also measured in Levente's written samples by the ratio of error-free clauses to the total number of clauses (EFC/C) index. As shown in Figure 7.14, the EFC/C index fluctuated slightly over the nine months.

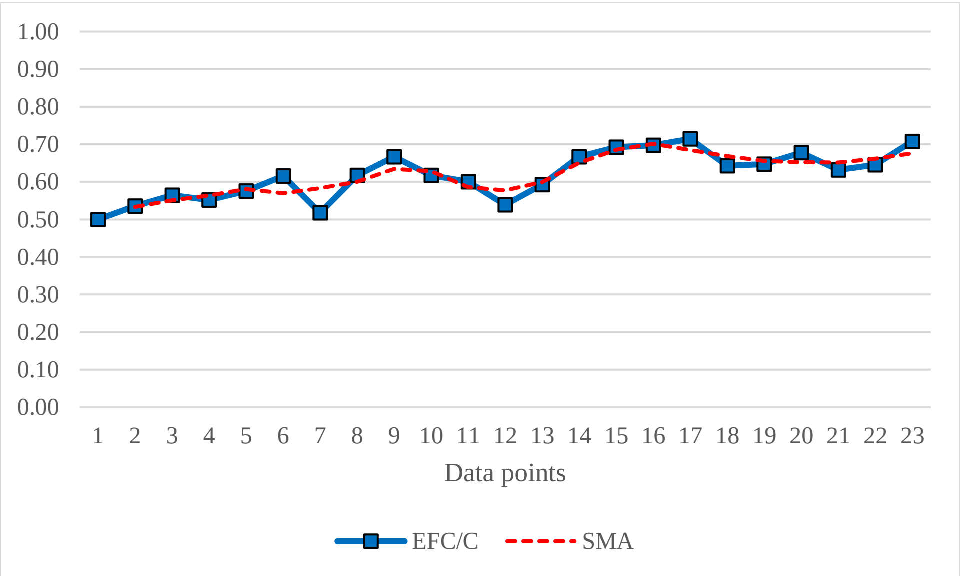


Figure 7.14 Accuracy

However, between data points 12 and 17, a gradual increase can be observed in the trajectory of the accuracy (EFC/C) index. The smoothing algorithm of the accuracy (EFC/C) index displays a clear upward trend over the nine-month investigation. The lowest EFC/C value (EFC/C = 0.50) was measured at data point 1, while the highest (EFC/C = 0.71) was gauged at data point 14. The lowest EFC/C value means that there was at least one error in every second clause in Essay 1. The following example illustrates the low accuracy rate in Essay 1.

> Essay excerpt 6.4 (Essay 1)
>
> [1]Over all, it is not very bad that we have to have [2]varius posts in our employment but if we plan our [3]occuppation better we can save a lot of time and energy.
> (Errors are uncorrected, underlined and numbered in superscript)
> (Error-free clause = 2; clause = 4; EFC/C = 0.5)

In order to gain a deeper insight into the accuracy construct of Levente's written data, this study also looked at the type of errors he made in his essays. The errors were coded and categorised according to the Louvain Error Tagset, and the results are displayed in Figure 7.15. Figure 7.15 shows that the number of errors decreased over time along with the error-free clause ratio (EFC/C), displayed in Figure 7.14. Furthermore, Figure 7.15 shows that Levente made five different types of errors in his essays: (1) formal, (2) grammatical, (3) lexico-grammatical, (4) lexical, and (5) word redundant, word missing and word order. Among the five different types of errors, formal (morphology and spelling) (40%) and grammatical (40%) errors were the most prevalent. Moreover, formal and grammatical errors were the only types of errors which were present in each essay. The third most common type of error belonged to the word redundant, word missing or word order errors category (10%). However, grammatical errors were not evident in every essay. The other two categories remained below 10% and their occurrence was also random.

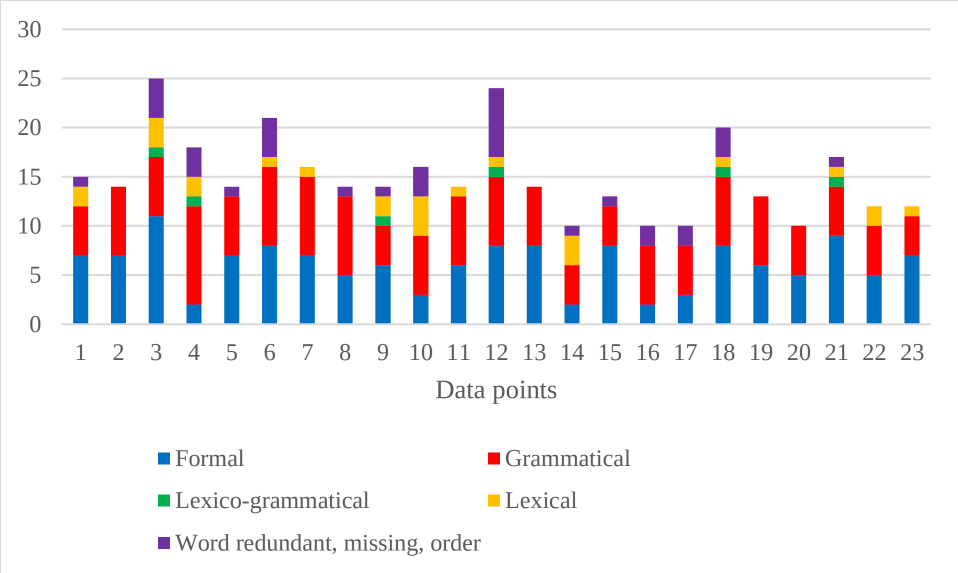


Figure 7.15 Error type

7.3 The Degree of Variability

Section 7.3 explores the degree of variability in lexical and syntactic complexity and accuracy in Levente's data. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity and (3) accuracy.

7.3.1 Lexical Complexity

In the min-max graph of lexical disparity (LSA) four main stages can be observed (see Figure 7.16). In the first stage the bandwidth was wide (from 0.66 to 0.29) from data point 1 to 7. In the second short stage, from data point 8 to 11, the bandwidth became narrow (from 0.11 to 0.22) but then in the third stage, from data point 12 to 16, the bandwidth became wider (from 0.10 to 0.31) again as in the first stage. In the last stage the bandwidth became even wider (from 0.13 to 0.47) from data point 18 to 23. The visual inspection of the min-max graph of the lexical disparity (LSA) index suggests that the developmental peak at data point 20 might be statistically significant.

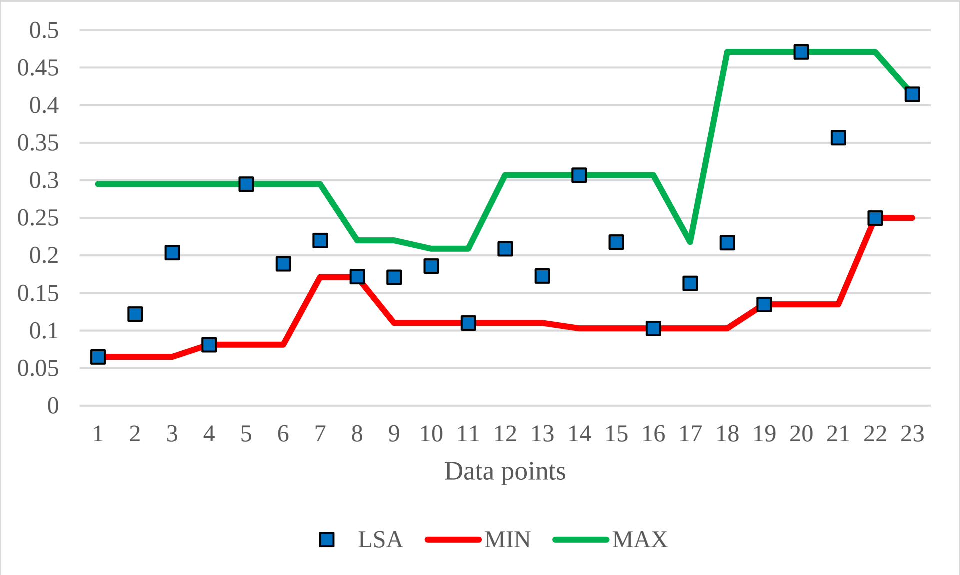


Figure 7.16 Lexical disparity (min-max graph)

The data of the lexical disparity (LSA) index were resampled, and a Monte Carlo analysis was run. Table 7.2 shows that the p-value in the lexical disparity (LSA) index was 0.19. In other words, statistically significant developmental peaks were not detected in the lexical disparity (LSA) index.

Table 7.2

P-values for the Complexity and Accuracy Indices

Index	P-value
AveWL	0.31
TTR	0.85
WRDFRQc	0.64
LSA	0.19
AWL	0.70
FVR	0.65
DC/C	0.55
T/S	0.60
CP/C	0.95
CN/C	0.19
EFC/C	0.72

The same statistical procedures were repeated for the general lexical complexity (AveWL), the lexical variability (TTR), the lexical rarity (WRDFRQc) and the Academic Word List (AWL) indices. However, the p-values in the AveWL, the TTR, the WRDFRQc, and the AWL indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the AveWL, the TTR, the WRDFRQc, and the AWL indices in Levente's written data.

7.3.2 Syntactic Complexity

In the min-max graph of phrasal complexity (CN/C) index (Figure 7.17), two main stages can be observed. The first bandwidth, from data point 1 to 17, was relatively wide and it moved down. Although there was a certain degree of change in the width of the bandwidth (from 0.72 to 1.00, from 0.63 to 0.97, from 0.56 to 0.96), these changes were not robust. The second main stage started at data point 18 and finished at data point 23. In the second stage the bandwidth was slightly wider (from 0.62 to 1.10) than in the first stage.

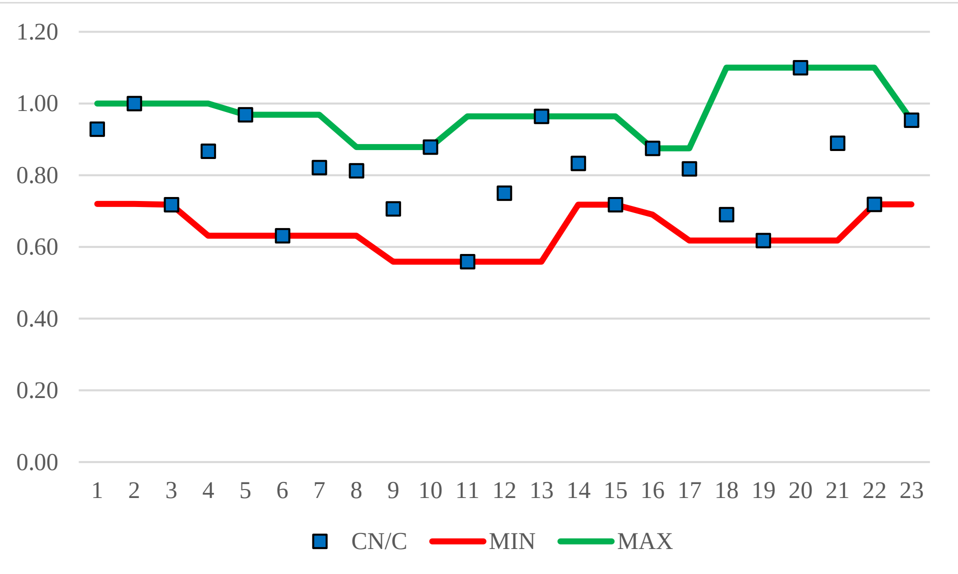


Figure 7.17 Phrasal complexity (min-max graph)

Table 7.2 shows that the p-values in the general syntactic complexity (FVR), the subordination (DC/C), the clausal coordination (T/S), the phrasal coordination (CP/C) and the phrasal complexity (CN/C) indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the five syntactic complexity indices in Levente's written data.

7.3.3 Accuracy

The degree of variability in the construct of accuracy was measured with the error-free clause ratio (EFC/C) index. In the min-max graph (Figure 7.18) of accuracy (EFC/C) some variability can be observed, but as the min-max graph shows, there were no substantial differences in the bandwidth over the nine months, even though the band itself moved up between data points 14 and 17.

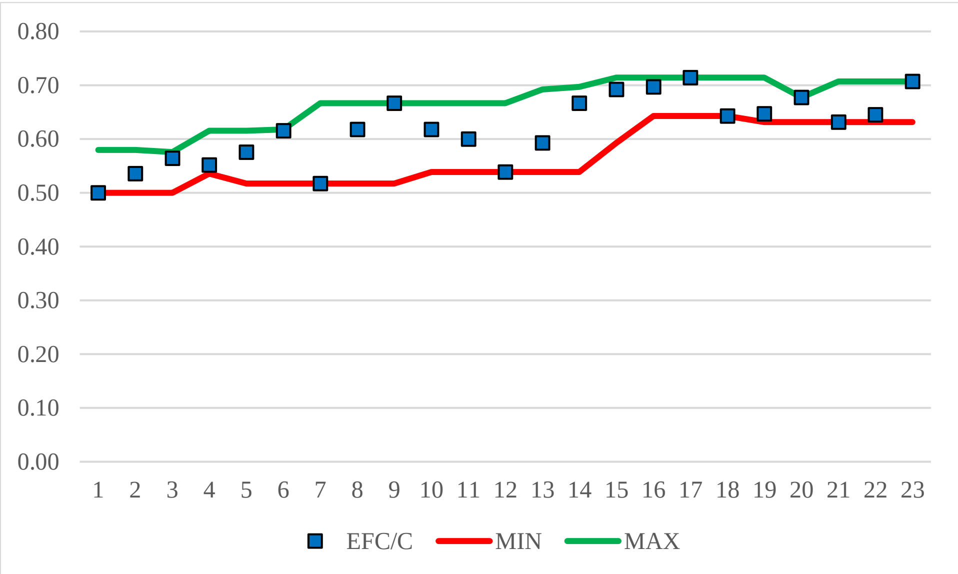


Figure 7.18 Accuracy (min-max graph)

Table 7.2 shows that the p-value in the accuracy (EFC/C) index was above 0.05. In other words, a statistically significant developmental peak was not detected in the accuracy (EFC/C) index in Levente's written data.

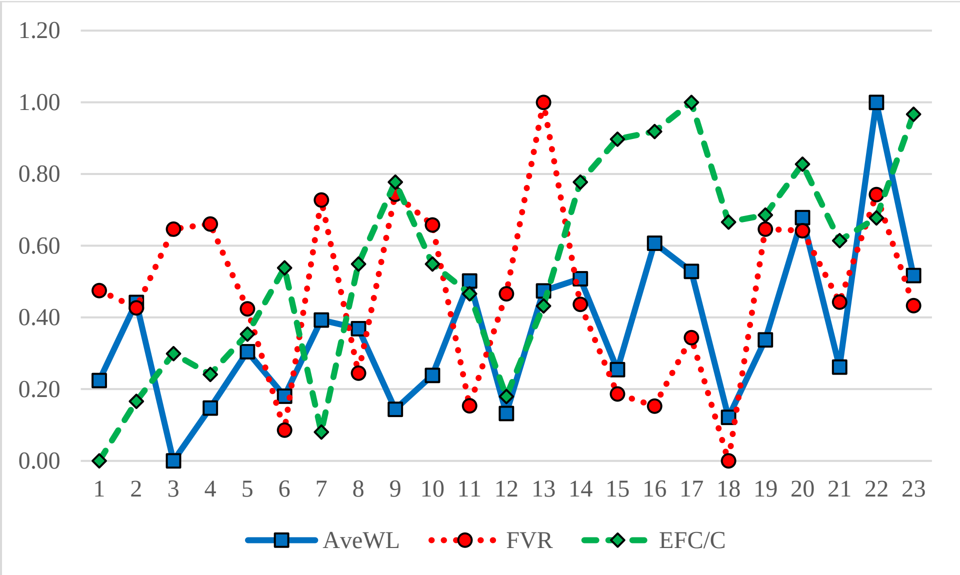
7.4 Interactions

The purpose of this section is to explore the dynamic interactions between lexical and syntactic complexity and accuracy. This section is subdivided into two parts: (1) interactions between general lexical and syntactic complexity and accuracy indices and (2) interactions between lexical variability, subordination, and accuracy indices.

7.4.1 Interactions Between General Lexical and Syntactic Complexity and Accuracy

Figure 7.19 shows (top) the normalised data of the general lexical complexity (AveWL) and general syntactic complexity (FVR) and the accuracy (EFC/C) indices. The trajectories of the general lexical complexity (AveWL) and the general syntactic complexity (FVR) exhibit both alternating and parallel patterns. However, between data points 17 and 19 and between data points 20 to 23, the growth trajectories were clearly parallel. The growth trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) were mainly parallel but there were alternating shifts between data points 5 to 11. Conversely, there were clear parallel shifts between the general lexical complexity (AveWL) and accuracy (EFC/C) trajectories between data points 17 and 22. The growth trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) were both alternating and parallel over the nine months. However, between data points 8 and 11 there were clear parallel shifts, whereas between data points 5 and 7 and between data points 13 and 16 there were clear alternating patterns.

Figure 7.19 (bottom) shows the residual plot for the AveWL, FVR and the EFC/C indices. The trajectories of the general lexical complexity (AveWL) and general syntactic complexity (FVR) were both parallel and alternating. There were clear parallel shifts between data points 20 and 23, while there were clear alternating patterns between data points 8 and 12. In these shifts, a movement above the trend in the general lexical complexity (AveWL) index was accompanied by a movement below it in the general syntactic complexity (FVR) index and vice versa, for example between data points 2 and 3. The trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) also show both parallel and alternating shifts.



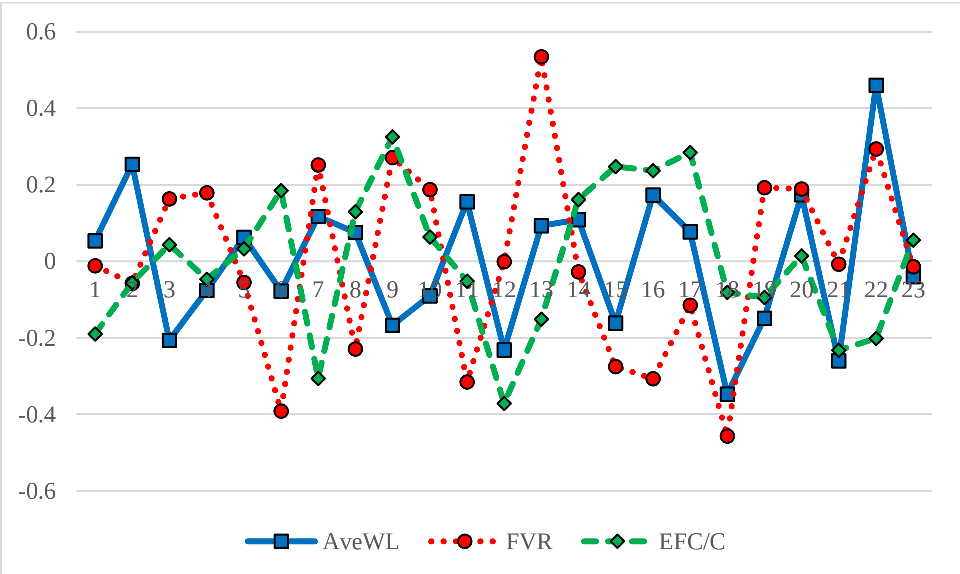


Figure 7.19 General lexical and syntactic complexity and accuracy (top: normalised data, bottom: residual plot)

Between data points 5 and 11 there were clear alternating shifts, while between data points 11 and 14 there were clear parallel patterns. The growth trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) also show both parallel and alternating shifts. In these shifts, a movement above the trend in the general syntactic complexity (FVR) index is accompanied by a movement below it in the accuracy (EFC/C) index and vice versa, for example at data points 7 and 8.

Figure 7.20 shows the interactions between general lexical (AveWL) and syntactic complexity (FVR) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 7.20 shows that the general lexical complexity (AveWL)-general syntactic complexity (FVR) correlation shifted between strong negative and moderately positive values. The general lexical complexity (AveWL)-accuracy (EFC/C) correlation shifted between strong negative values and strong positive values. However, Figure 7.20 shows that the trajectory of the AveWL-EFC/C correlation was mainly positive. The general syntactic complexity (FVR)-accuracy (EFC/C) correlation shifted between moderately positive and negative values over the nine months.

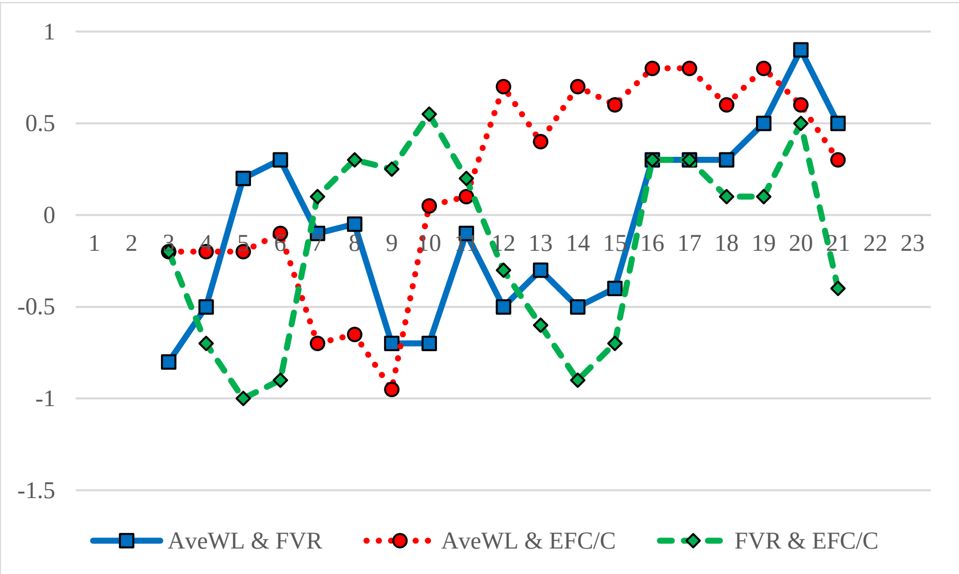


Figure 7.20 Moving correlations (General lexical and syntactic complexity and accuracy)

Table 7.3 shows the correlation coefficient values between the general lexical and syntactic complexity and accuracy indices. The correlation between the general lexical complexity (AveWL) and accuracy (EFC/C) was positive and statistically significant. The general lexical complexity (AveWL)-general syntactic complexity (FVR) and the general syntactic complexity (FVR)-accuracy (EFC/C) correlations were not statistically significant.

Table 7.3

Correlation Matrix (General Lexical and Syntactic Complexity and Accuracy)

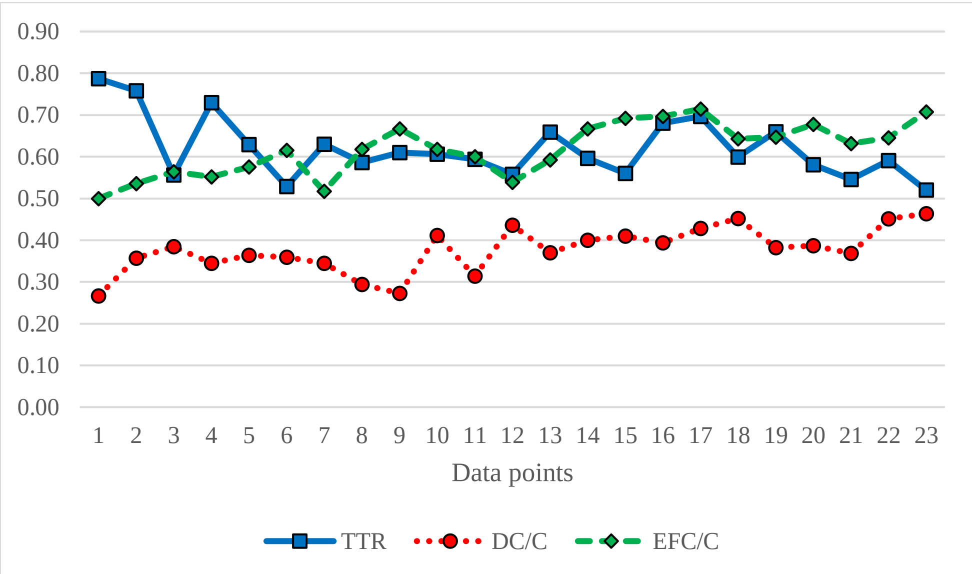
		AveWL	FVR	EFC/C
AveWL	ρ	1.000	-.022	.462*
	Sig. (2-tailed)		.920	.026
	N	23	23	23
FVR	ρ	-.022	1.000	-.245
	Sig. (2-tailed)	.920		.260
	N	23	23	23
EFC/C	ρ	.462*	-.245	1.000
	Sig. (2-tailed)	.026	.260	
	N	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed)

7.4.2 Interactions Between Lexical Variability, Subordination, and Accuracy

Figure 7.21 shows (top) the raw data values of the lexical variability (TTR), the subordination (DC/C) and the accuracy (EFC/C) indices. Figure 7.21 shows that the growth trajectories of the lexical variability (TTR) and subordination (DC/C) were both alternating and parallel over the nine months. Likewise, the trajectories of lexical variability (TTR) and accuracy (EFC/C) were both parallel and alternating. Conversely, the trajectories of the subordination (DC/C) and accuracy (EFC/C) were predominantly parallel.

Figure 7.21 (bottom) shows the residual plot for the lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices. The trajectories of lexical variability (TTR) and subordination (DC/C) were mainly alternating, especially between data points 1 and 5. In these shifts, a movement above the trend in the lexical variability (TTR) index was accompanied by a movement below it in the subordination (DC/C) index and vice versa, for example between data points 12 and 13 and between data points 15 and 16. Figure 7.21 (bottom) also shows that growth trajectories of lexical variability (TTR) and accuracy (EFC/C) were mainly alternating. In these shifts, a movement above the trend in the lexical variability (TTR) index was accompanied by a movement below it in the accuracy (EFC/C) index and vice versa, for example between data points 2 and 5 and between data points 13 and 14. The growth trajectories of subordination (DC/C) and accuracy (EFC/C) show both parallel and alternating shifts. There was a clear correspondence between the DC/C and the EFC/C trajectories between data points 1 and 5. However, there were clear alternating shifts between data points 7 and 10.



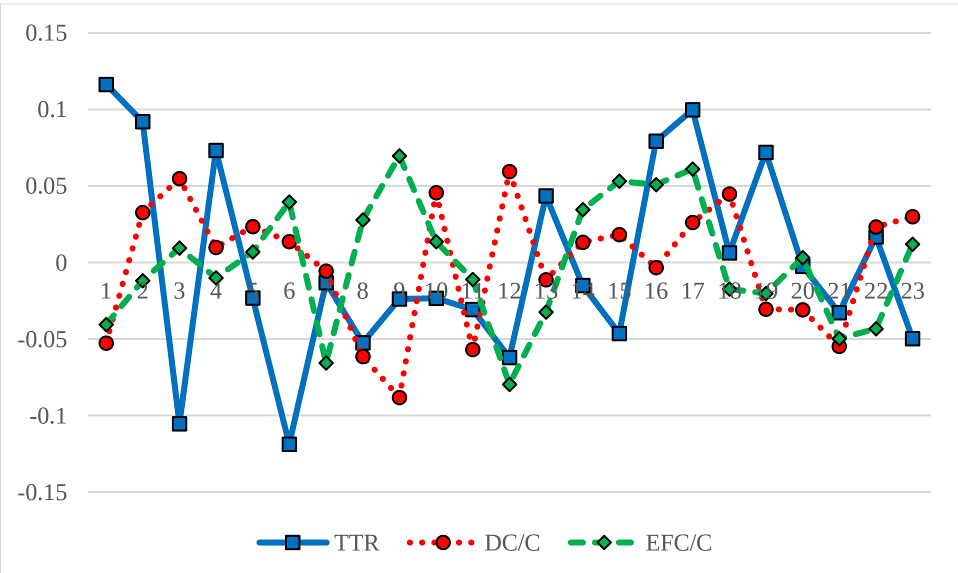


Figure 7.21 Lexical variability, subordination and accuracy (top: data values, bottom: residual plot)

Figure 7.22 shows the interactions between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 7.22 shows that the lexical variability (TTR)-subordination (DC/C) correlation shifted between strong negative and moderately positive values. The lexical variability (TTR)-accuracy (EFC/C) correlation also shifted between strong negative and moderately positive values. Likewise, the subordination (DC/C)-accuracy (EFC/C) correlation also shifted between strong positive and negative values.

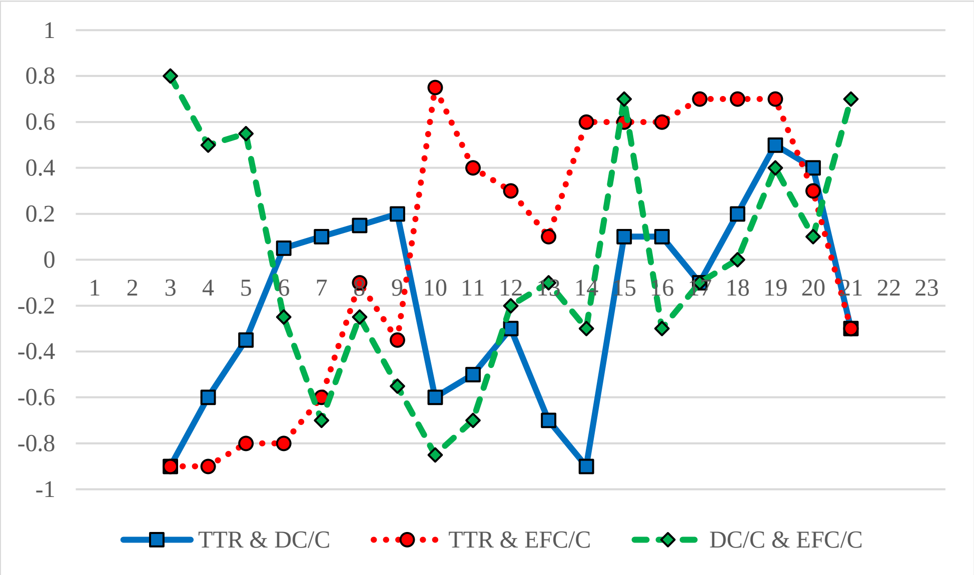


Figure 7.22 Moving correlations (Lexical variability, subordination, and accuracy)

Table 7.4 shows the correlation coefficient values between the lexical variability (TTR), subordination (DC/C) and the accuracy (EFC/C) indices. The correlation coefficient between the lexical variability (TTR) and subordination (DC/C) was negative (-.348) but not statistically significant. Likewise, the correlation between lexical variability (TTR) and accuracy (EFC/C) was negative (-.245) but not statistically significant. However, the correlation between subordination (DC/C) and accuracy (EFC/C) was positive (.533) and highly significant (p<0.01).

Table 7.4

Correlation Matrix (Lexical Variability, Subordination, and Accuracy)

		TTR	DC/C	EFC/C
TTR	ρ	1.000	-.348	-.245
	Sig. (2-tailed)		.103	.260
	N	23	23	23
DC/C	ρ	-.348	1.000	.533*
	Sig. (2-tailed)	.103		.009
	N	23	23	23
EFC/C	ρ	-.245	.533*	1.000
	Sig. (2-tailed)	.260	.009	
	N	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed)

In summary, both the raw data values and the residuals exhibited both parallel and alternating patterns for the lexical variability (TTR), the subordination (DC/C) and the accuracy (EFC/C) indices. Likewise, the moving correlations for all three correlations showed shifts between moderately negative and moderately positive values. However, the correlation coefficient for the subordination-accuracy interaction was positive and statistically significant indicating a supportive relationship.

7.5 Modelling

The purpose of this section is to test the hypothesised interactions between lexical variability (TTR) and subordination (DC/C), between lexical variability (TTR) and accuracy (EFC/C), and between subordination (DC/C) and accuracy (EFC/C). Based on the variability analyses (section 7.4.2) the following hypotheses were formulated:

1. weak competition between lexical variability (TTR) and subordination (DC/C),
2. weak competition between lexical variability (TTR) and accuracy (EFC/C),
3. moderate support between subordination (DC/C) and accuracy (EFC/C).

To test the abovementioned hypotheses, a model of three connected growers was used.

7.5.1 Model Setup

Three growers were specified in the model which corresponded with the three constructs under investigation in the previous section (Section 7.4.2) that is: (1) lexical variability (TTR), (2) subordination (DC/C), and (3) accuracy (EFC/C). Table 7.5 shows the numerical values that were assigned to the parameters in the model.

Table 7.5

Property and Relational Parameters

		Grower A	Grower B	Grower C
Property parameters	Initial value	0.787	0.2667	0.5
	Rate	-0.05	0.05	0.05
	Carrying capacity	1	1	1
Relational parameters		to A	to B	to C
	from A	-	-0.0175	-0.0125
	from B	-0.0175	-	0.0265
	from C	-0.0125	0.0265	-

Table 7.5 also shows the relational parameters used in Levente's data. After specifying the property and relational parameters, the data were simulated. The outcome of 100 iterations of the configured equations was obtained by the data simulation. Figure 7.23 shows the result of the simulations after 100 iterations.

7.5.2 Optimization

Before evaluating the model by correlating the model with the data, with the linear trends, and the cubic spline interpolations, the relational and property parameters were optimized. Table 7.6 shows the optimized parameter values.

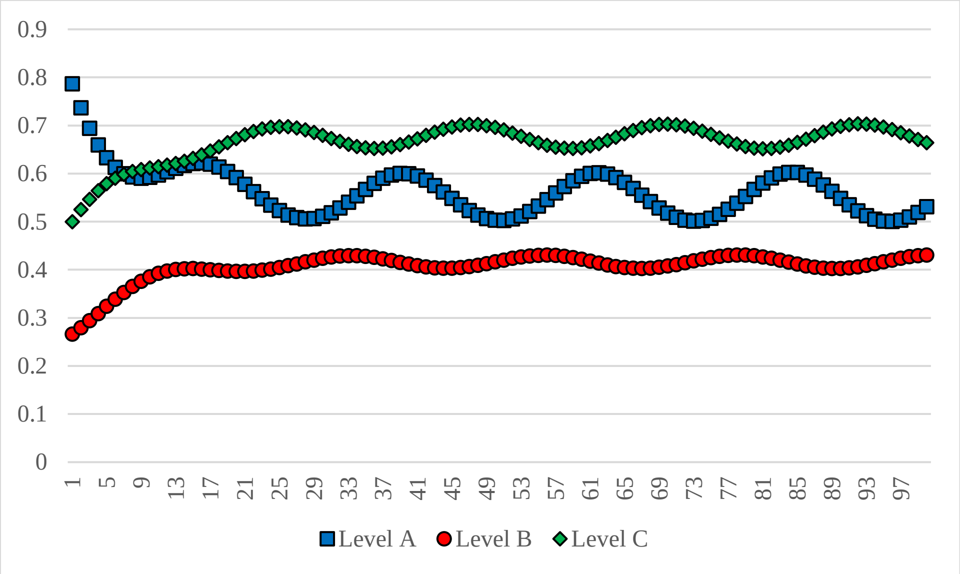


Figure 7.23 Growth model

Table 7.6

Optimized Parameter Values

		Grower A	Grower B	Grower C
Property parameters	Rate	0.24866	0.49955	-0.15153
Relational		to A	to B	to C
parameters	from A	-	-0.26679	0.21885
	from B	0.84275	-	-0.17176
	from C	-0.68318	-0.21300	-

The optimization of the parameters made it possible to assess the overall fit of the model to the data. First, the sum of squared residuals between the model and the data was compared, which was 0.14842 in Levente's data. Second, the model outcome was visually contrasted with the data. Figure 7.24 shows that model resembles the data quite well.

The optimized parameter values confirm one of the three hypotheses that informed the setup of the model. The hypothesised weak competitive relationship between lexical variability (TTR) and subordination (DC/C) was not confirmed by the model since the aggregated parameter values between Level A and B was positive (0.57596). However, the hypothesised weak competitive relationship between lexical variability (TTR) and accuracy (EFC/C) was confirmed by the model because the aggregated relational parameter was negative (-0.46433) between Level A and C. The hypothesised moderately supportive relationship between subordination (DC/C) and accuracy (EFC/C) was not corroborated by the negative aggregated parameter value (-0.38476) between Level B and C.

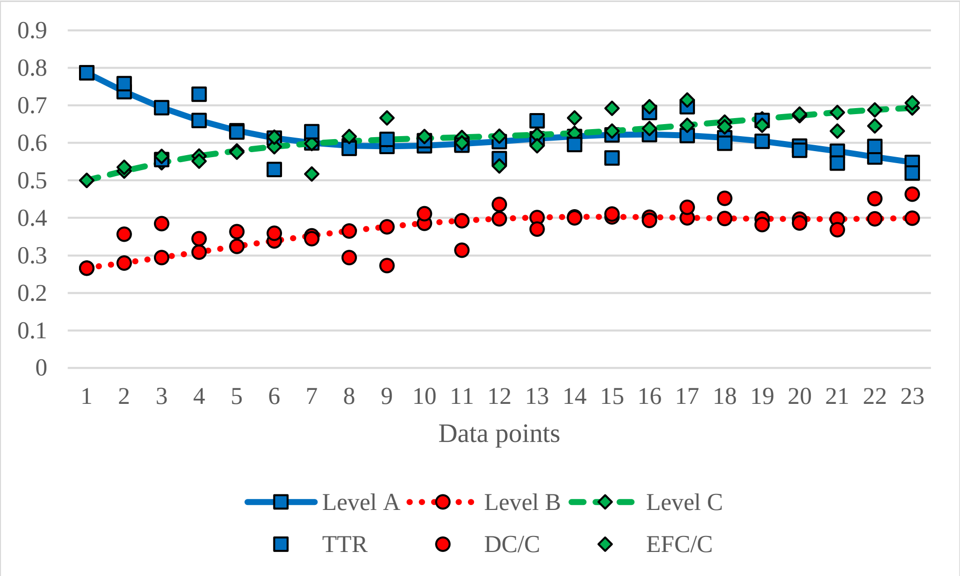


Figure 7.24 Data and model (fit = 0.14842)

7.5.3 Evaluation of the Model

Figure 7.25 presents the data, the linear trends, the cubic spline interpolations, and model outcome. Visual inspection of these plots indicates that the model resembles a combination between the smoothing splines and the linear trends. Level A clearly resembles the corresponding smoothing spline, while Level B and C were similar to a combination of the corresponding linear trends and smoothing splines.

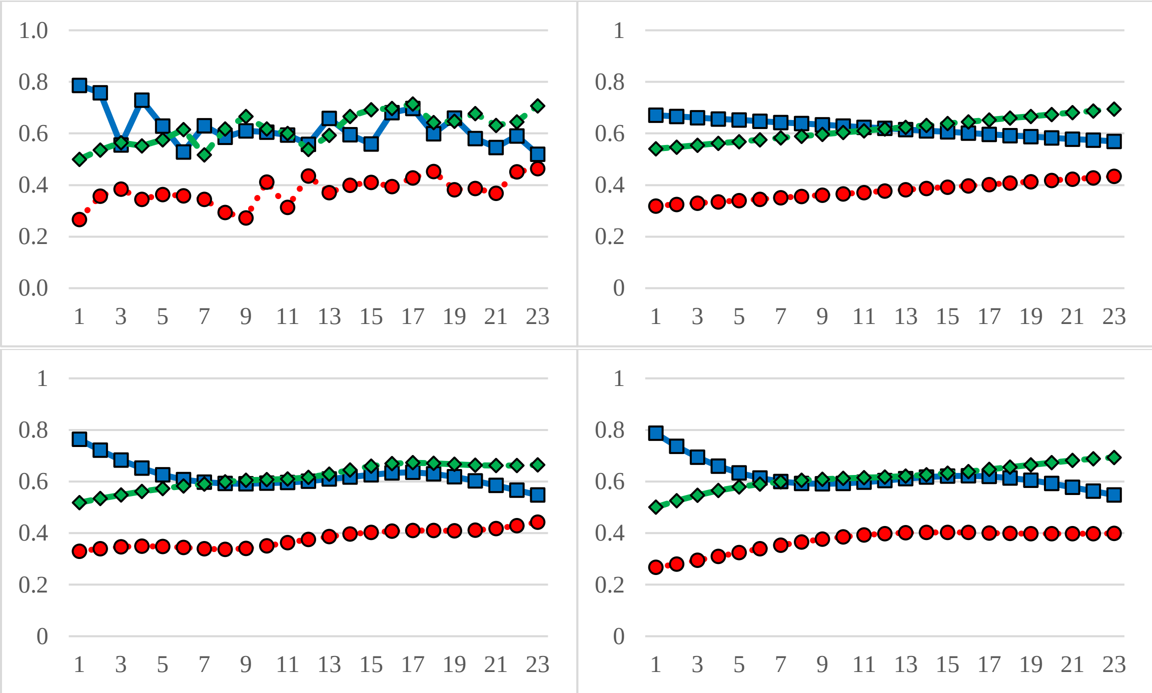


Figure 7.25 Data (top left), linear trends (top right), smoothing splines (bottom left) and model (bottom right)

The model was correlated with the data, the linear trends, and the smoothing splines. Table 7.7 shows that the model correlated positively at all three levels with the original data.

Table 7.7

Correlations Between the Model and the Data, Linear Trends, and Splines

			Model-data	Model-linear	Model-spline
Level A	TTR	ρ	.536**	.626**	.960**
		Sig. (2-tailed)	.009	.002	.000
		N	23	23	23
Level B	DC/C	ρ	.671**	.778**	.723**
		Sig. (2-tailed)	.000	.000	.000
		N	23	23	23
Level C	EFC/C	ρ	.766**	1.000**	.927**
		Sig. (2-tailed)	.000	.000	.000
		N	23	23	23

Note. ** Correlation was significant at the 0.01 level (2-tailed).

In addition to the correlation between the model and the data, the linear trends, and the smoothing splines, another type of correlation analysis was performed to confirm the goodness-of-fit of the model. Correlation matrixes between the various categories in the data, the linear trends, the smoothing splines, and the model were compared to ascertain whether the model replicates the surface level interactions of the data. Table 7.8 shows the correlation matrixes for the data, the linear trends, the smoothing splines, and the model.

Table 7.8

Correlation Matrixes

			Data	Linear	Spline	Model
TTR	DC/C	ρ	-.348	-1.000**	-.293	-.270
		Sig. (2-tailed)	.103	.000	.174	.212
		N	7	25	25	25
TTR	EFC/C	ρ	-.245	-1.000**	-.263	-.626**
		Sig. (2-tailed)	.260	.000	.225	.002
		N	7	25	25	25
DC/C	EFC/C	ρ	.533**	1.000**	.872**	.778**
		Sig. (2-tailed)	.009	.000	.000	.000
		N	7	25	25	25

Note. ** Correlation was significant at level 0.01 (2-tailed).

Table 7.8 shows the correlation matrixes were highly similar in the data, the smoothing spline and the model in the lexical variability (TTR) and subordination (DC/C), and in the lexical variability (TTR) and accuracy (EFC/C) correlations. In addition, there were similarities in the subordination (DC/C) and accuracy (EFC/C) correlations. However, these similarities were not as great as in the previous two correlations.

7.6 Motivation

The purpose of this section is to explore the development of Levente’s motivational system. This section is subdivided into four parts: (1) goals for academic writing, (2) self-regulatory processes, and (3) attractor states.

7.6.1 Goals for Academic Writing

During the initial interview, Levente reported that he would not write in English in his future Bachelor studies at university. However, he also said that he would like to take the IELTS language test because he would do his Master studies in English at a university in Budapest or abroad. He also added that he would write reports and his curriculum vitae in English in his future career. His usual method of writing in English includes collecting ideas and revising the text. He also added that he usually checks his spelling during the revision of his compositions. Levente also reported that he would write on topics related to engineering. Therefore, he said that he would like to focus on IELTS Writing Task 1. He also added that he would improve his vocabulary because he admitted that he has a “small vocabulary” and he would review his grammar during the EAP course. Levente also admitted that he rarely reads in English.

7.6.2 Self-Regulatory Processes

Figure 7.26 shows the evolution of Levente’s self-regulatory processes. Figure 7.26 shows that Levente focused mainly on self-observation processes between data point 3 and 9. However, from data point 12 a shift can be observed from self-observation to self-evaluation. For example, Levente’s interview data contained four self-evaluation thematic units, while only one self-observation thematic unit at data point 12. Between data points 15 and 21, the preponderance of self-evaluation thematic units can be observed in Levente’s self-regulatory processes. Figure 7.26 also demonstrates that Levente did not set any goals over the nine-month investigation.

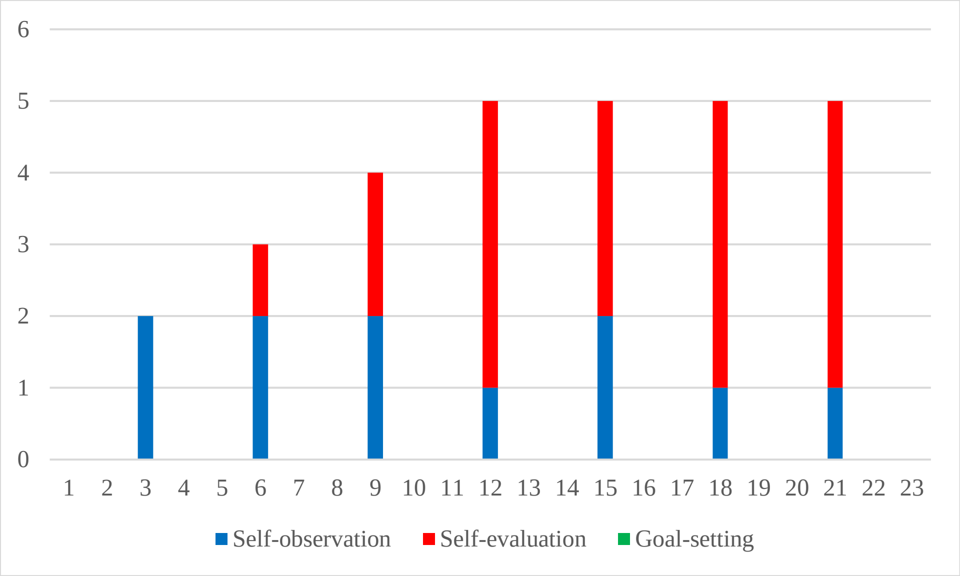


Figure 7.26 Self-regulatory processes

Table 7.9 provides a detailed breakdown of the range of Levente's self-regulatory processes over the nine months.

Table 7.9

Self-Regulatory Processes

		Data point							
		3	6	9	12	15	18	21	Total
Self-observation	Language	1	0	0	1	1	0	0	
	Composing processes	1	1	1	0	1	1	1	
	Content	0	1	1	0	0	0	0	
	Quality	0	0	0	0	0	0	0	
	Total	2	2	2	1	2	1	1	11
Self-evaluation	Language	0	1	2	2	1	2	3	
	Composing processes	0	0	0	0	0	0	0	
	Content	0	0	0	1	1	1	1	
	Quality	0	0	0	1	1	1	0	
	Total	0	1	2	4	3	4	4	18
Goal-setting	Language	0	0	0	0	0	0	0	
	Composing processes	0	0	0	0	0	0	0	
	Content	0	0	0	0	0	0	0	
	Quality	0	0	0	0	0	0	0	
	Total	0	0	0	0	0	0	0	0

The number of self-observation thematic units was evenly distributed (ranging between 1 and 2 thematic units) over time. However, the types of self-observation thematic units was not evenly distributed since Levente constantly focused on his composing processes, whereas he did not concentrate on the quality of his essays. Furthermore, the number of self-evaluation thematic units was not evenly distributed over time. Between data points 3 and 9 the number of self-evaluation thematic units ranged between 0 and 2. However, an increase can be observed in the number of self-evaluation thematic units at data point 12. This result suggests that there was a slight improvement in his self-regulatory processes. However, no goals were identified in Levente's data.

The following examples show that Levente's focus shifted from self-observation to self-evaluation at around data point 18. Interview excerpt 7.1 is an example of his impressionistic description of his action containing somewhat superficial description of his own writing.

Interview excerpt 7.1 (Self-observation: language, data point 12)

Interviewer:	Did you try out any new words or phrases that you haven't used much before?
Levente:	Yes, maybe I tried some new

However, Interview except 7.2 contains evaluative comments that involved cognitive processes (reasoning).

Interview excerpt 7.2 (Self-evaluation: language, data point 18)

Interviewer:	Are you happy with the level of vocabulary that you were using in this writing?
Levente:	Yes, I am happy with it because there were only one words, word what I couldn't write.

Interview excerpt 7.3 also contains evaluative comments that involved cognitive processes (reasoning).

Interview excerpt 7.3 (Self-evaluation: language, data point 21)

Interviewer:	Why didn't you include those words in the essay? It's not a question that you should include. I'm just asking. Why not? Is there a reason?
Levente:	Maybe because these words is not so automatically useable for me now. I don't know. Maybe this is the reason.

Interview excerpts 7.1-7.3 also show that Levente's description of his L2 learning experience did not contain more specific explanations over time which indicates that his self-regulatory processes did not evolve substantially, although there was a shift from self-observation to self-evaluation at data point 18.

7.6.3 Attractor States

The aim of this subsection was to identify the main attractor states that made up Levente's L2 learning experiences and then to explore the development of these states. The first dominant state mentioned by Levente was boredom. Levente quite often mentioned that he was not interested in the task (see Interview excerpt 7.4).

Interview excerpt 7.4 (data point 9)

Interviewer:	Do you think that this piece of writing is a good representation of how well you can write?
Levente:	Well… it can be but the topic wasn't my… best… I'm not very interested in this topic.

The lexical variability (TTR) value reflects Levente's disinterest in the topic of 'language learning'. The TTR index decreased from the onset of the data collection (Figure 7.3). From data point 1 to 2, the TTR index was relatively high (over 0.7), but when the first sample was written on 'language learning' the TTR index dropped from 0.76 to 0.56. When the second sample was composed on 'language learning' the TTR index was quite low (0.53). Even on the third occasion (data point 9), the TTR index was lower than the first two data points. At data point 12, during the retrospective interview, Levente expressed his disinterest in the topic of 'language learning' once again (Interview excerpt 7.5).

Interview excerpt 7.5 (data point 12)

Interviewer:	Do you think that this piece of writing is a good representation of how well you can write?
Levente:	Well, it isn't because I am better when I have to write about my own ideas so there is no topic
Interviewer:	So you think that this topic is really precise or there is no freedom in it?
Levente:	Yes… kötött (restricted)

The second dominant state mentioned by Levente was anxiety. In his Initial interview at data point 3, Levente described how frustrated he feels when he does not know a lexical item during writing (see Interview excerpt 7.6).

Interview excerpt 7.6 (data point 3)

Interviewer:	How do you feel when you write in English?
Levente:	When I don't know a word I will be angry... and then I always open a next dictionary
Interviewer:	So you are frustrated… or stressed a little bit
Levente:	Yeah.

Low self-efficacy beliefs might have contributed to the stagnation of Levente's self-regulation. At data point 18 and 21 (Interview excerpt 7.7-7.8), Levente revealed that he perceives himself as a "bad writer" even in his mother tongue.

Interview excerpt 7.7 (data point 18)

Interviewer:	How do you see yourself as a writer? So, what kind of problems do you have?
Levente:	I don't know maybe I'm not the best writer in Hungarian, too
Interviewer:	Did you have problems, for example in Hungarian literature or grammar?
Levente:	Not really but it is hard to express myself in a topic what I don't really understand.

Interview excerpt 7.8 (data point 21)

Interviewer:	Do you like writing argumentative essays?
Levente:	No, I don’t like these topics, these types of essays.
Interviewer:	Why?
Levente:	I don’t know because when I learned in Hungarian… I was a bit bad in it… I don’t know, I got bad marks… for every essay.

Interview excerpts 7.4-7.8 show that boredom, anxiety, and low self-efficacy belief might have contributed to the stagnation of his motivational system, consequently his L2 writing development.

7.7 Summary

All three constructs of linguistic complexity exhibited nonlinear developmental trends in Levente’s written data over the nine-month period. No developmental peaks were detected in the lexical and syntactic complexity and accuracy indices over the nine months. The growth trajectories along with the residual plots, and the moving correlation plots indicated that the interactions between lexical and syntactic complexity and accuracy were dynamic. Both the magnitude and the polarity of these relationships fluctuated over the nine months. Furthermore, the correlation coefficients were also calculated to explore the surface interactions between the three constructs. The model confirmed the weak competitive relationship between lexical variability (TTR) and accuracy (EFC/C) indices. Nevertheless, the model did not confirm the weak competitive association between lexical variability (TTR) and subordination (DC/C) and the moderately supportive relationship between subordination (DC/C) and accuracy (EFC/C). The evolution of Levente’s self-regulatory processes were investigated and it was found that Levente’s focus shifted from self-observation to self-evaluation over time. Moreover, the interview data lacked the emergence of the third phase of self-regulation: goal-setting. Therefore, attractor states were identified in his interview data, and it was found that boredom and anxiety might have contributed to the stagnation of self-regulatory processes and consequently of his L2 writing. Furthermore, low self-efficacy belief might have caused the stagnation of his L2 writing development.

Chapter 8. Case Study Four

Chapter 8 presents the changes in the lexical and syntactic complexity and accuracy indices in Avarka's written data. Furthermore, the changes in her self-regulatory processes are presented. There are six main sections: (1) the participant and the context, (2) the developmental profile, (3) the degree of variability, (4) interactions, (5) modelling, and (6) motivation.

8.1 The Participant and Data

Avarka (pseudonym), a 22-year-old third-year Bachelor of Psychology student, was the fourth participant of this study. Her mother tongue is Hungarian, and she had been learning English for 12 years at the beginning of the data collection. Avarka's English language education started at the age of 12 at a prestigious secondary school in Budapest. She attended five classes on a weekly basis. She also had native English teachers at her secondary school. Furthermore, she attended private language schools in order to improve her English and prepare for the matura exam[4]. At the age of 16 Avarka took an advanced level matura exam in English at the beginning of grade 11 and her result was 90%. Apart from English, Avarka learned German, French and Russian.

During her Bachelor studies she did not engage in formal language education apart from occasional private English classes. In order to continue her studies in a Master programme, Avarka wanted to take the IELTS language exam. Therefore, she enrolled in the EAP course and attended 2-hour long classes on a weekly basis.

The data for this case were collected using the same procedure outlined in Chapter 4. Table 8.1 shows Avarka's written corpus and timeline.

[4] The matura, or érettségi vizsga (in Hungarian), is the secondary school exit exam in Hungary.

Table 8.1

Written Corpus and Timeline

	Written sample				Interview	
Data point	Code	Writing prompt	Date of submission	Word count	Type	Duration
1	Essay 1	WP1	19/09/2014	248		
2	Essay 2	WP2	03/10/2014	257		
3	Essay 3	CWP3	17/10/2014	227	Initial	8:30
					Retrospective 1	2:20
4	Essay 4	WP3	31/10/2014	236		
5	Essay 5	WP4	07/11/2014	270		
6	Essay 6	CWP4	28/11/2014	212	Retrospective 2	3:57
7	Essay 7	WP5	05/12/2014	258		
8	Essay 8	WP6	12/12/2014	257		
9	Essay 9	CWP2	19/12/2014	253	Retrospective 3	8:59
10	Essay 10	WP7	19/12/2014	251		
11	Essay 11	WP8	09/01/2015	227		
12	Essay 12	CWP5	23/01/2015	233	Retrospective 4	10:41
13	Essay 13	WP9	30/01/2015	233		
14	Essay 14	WP10	06/02/2015	252		
15	Essay 15	CWP1	27/02/2015	226	Retrospective 5	11:42
16	Essay 16	WP11	27/02/2015	225		
17	Essay 17	WP12	06/03/2015	218		
18	Essay 18	CWP6	20/03/2015	219	Retrospective 6	10:25
19	Essay 19	WP13	20/03/2015	269		
20	Essay 20	WP14	24/03/2015	220		
21	Essay 21	CWP3	27/03/2015	230	Retrospective 7	15:40
					Final	27:47
22	Essay 22	WP15	03/04/2015	244		
23	Essay 23	WP16	08/05/2045	239		
		Total word count		5504	Total duration	2:07:08

Note. Controlled data points are highlighted. WP = Writing prompt; CWP = Controlled writing prompt.

8.2 The Developmental Profile

This section is a detailed description of the developmental profiles of Avarka's academic writing pertaining to complexity and accuracy. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity, and (3) accuracy.

8.2.1 Lexical Complexity

Figure 8.1 shows that all five indices developed nonlinearly and fluctuated to different degrees over the nine months. As in the previous cases, the general lexical complexity (AveWL) and Academic Word List (AWL) indices show similar patterns, while the lexical variability (TTR), lexical rarity (WRDFRQc) and lexical disparity (LSA) indices exhibit different trajectories.

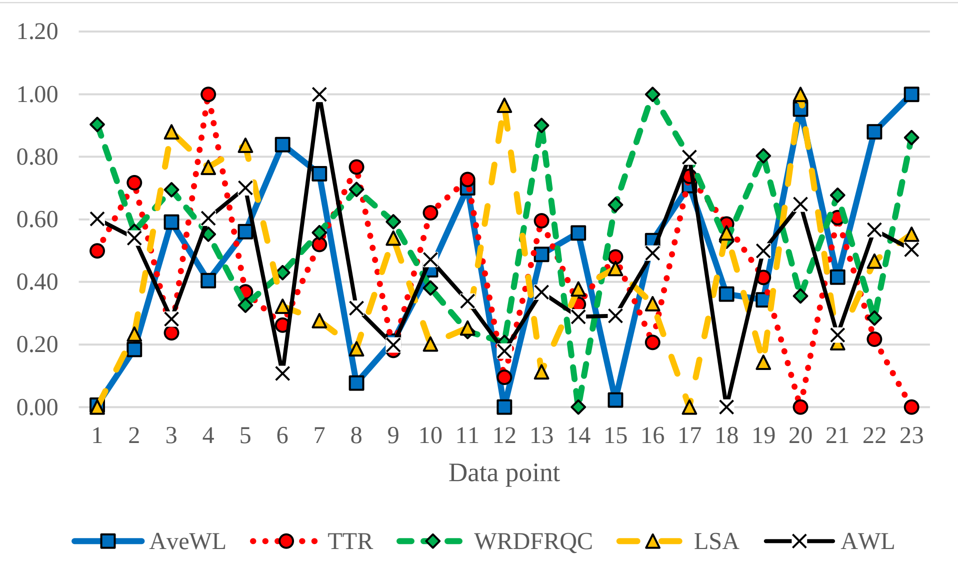


Figure 8.1 Lexical complexity

The average word length (AveWL) index was employed to measure general lexical complexity. Figure 8.2 shows that the AveWL index fluctuated between 4.52 and 5.17. The lowest AveWL index was measured at data points 1 and 12 (AveWL = 4.52), while the highest AveWL value was gauged at data point 20 (AveWL = 5.17). The smoothing algorithm of the general lexical complexity (AveWL) index displays a slightly upward trend over the nine-month investigation.

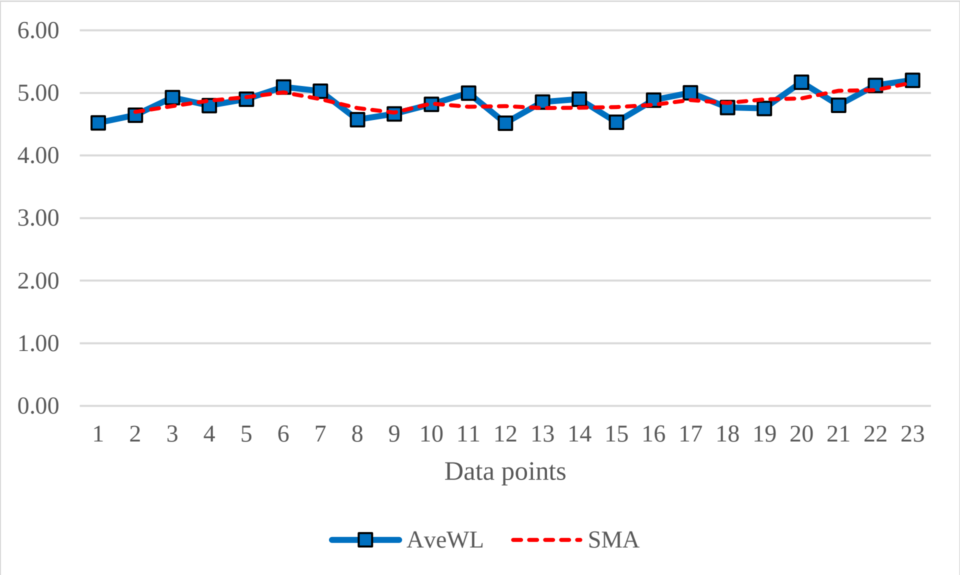


Figure 8.2 General lexical complexity (AveWL)

Figure 8.3 shows that the TTR index fluctuated between 0.64 and 0.84 over the nine months. The TTR index reached its zenith (0.84) at data point 4, while the lowest TTR values (0.64) were measured at data points 20 and 23. The smoothing algorithm of the lexical variability (TTR) index displays a sideways trend between data points 2 and 19. However, the smoother shows a slightly downward trend between data points 19 and 23.

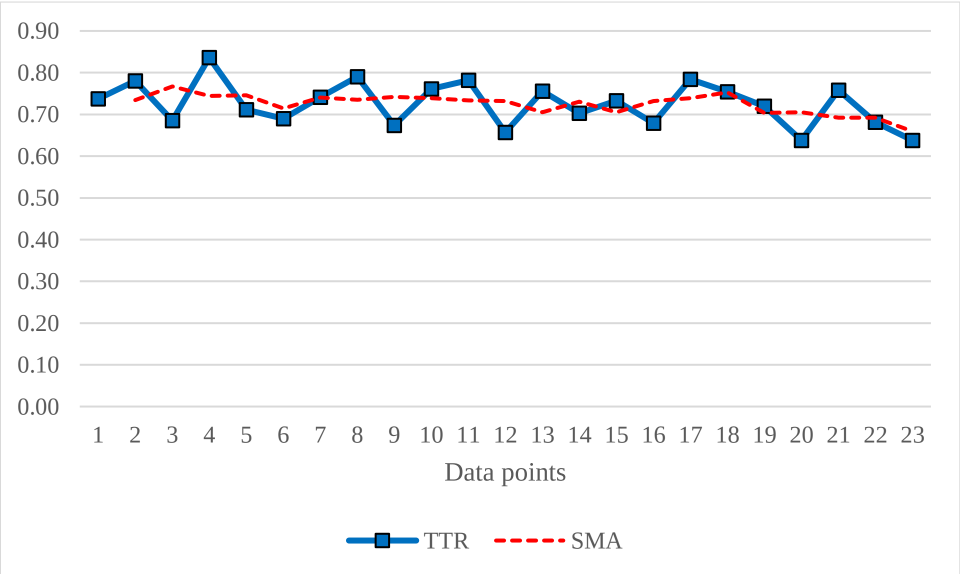


Figure 8.3 Lexical variability (TTR)

The second property of lexical diversity, rarity, was measured by the CELEX log minimum frequency of content words index (WRDFRQc) and shown in Figure 8.4. The WRDFRQc index shows more fluctuations than the AveWL and the TTR indices over the nine months. The highest lexical rarity (WRDFRQc) value (WRDFRQc = 1.62) was gauged at data point 16, while the lowest WRDFRQc value (0.75) was measured at data point 14. The lexical rarity (WRDFRQc) index was below 1.00 at data points 11, 12 and 14. The essays written during these data points included less frequent words than the essays written in the rest of the data points. The smoothing algorithm of the lexical rarity (WRDFRQc) index displays a sideways trend. However, the linear trend line shows a slightly upward trend over the nine-month investigation.

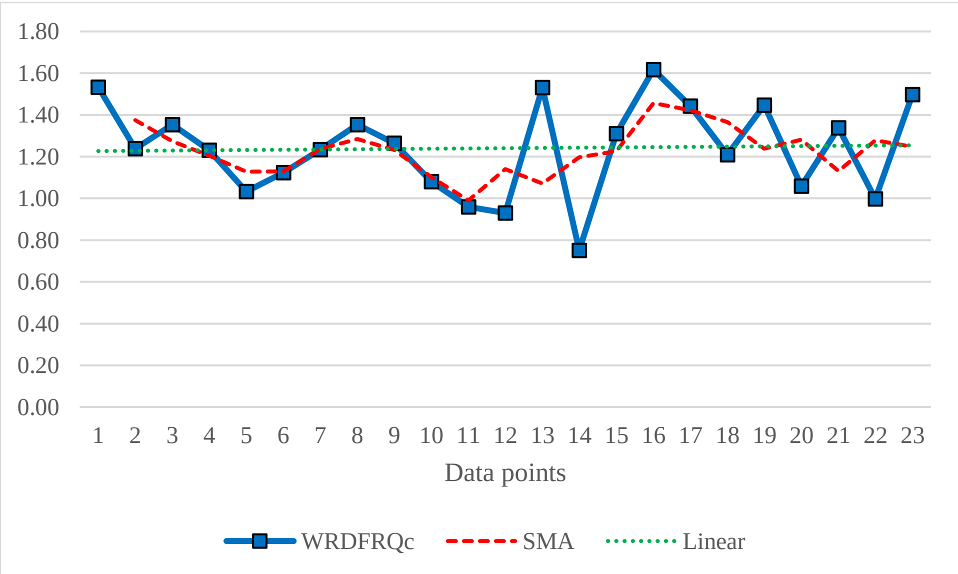


Figure 8.4 Lexical rarity (WRDRFQc)

The third property of lexical diversity, disparity, was measured by the latent semantic analysis index (LSA). The LSA index (Figure 8.5) fluctuated wildly over the nine months. The lowest LSA values (LSA = 0.09) were measured at data points 1 and 17, while the highest LSA value (LSA = 0.35) was detected at data point 20. The smoothing algorithm of the lexical disparity (LSA) index displays a sideways trend over the nine-month investigation.

Furthermore, the linear trend line displays a sideways trend (horizontal) over time which indicates stagnation in the lexical disparity (LSA) index. In other words, there were no changes in cohesion in Avarka's essays.

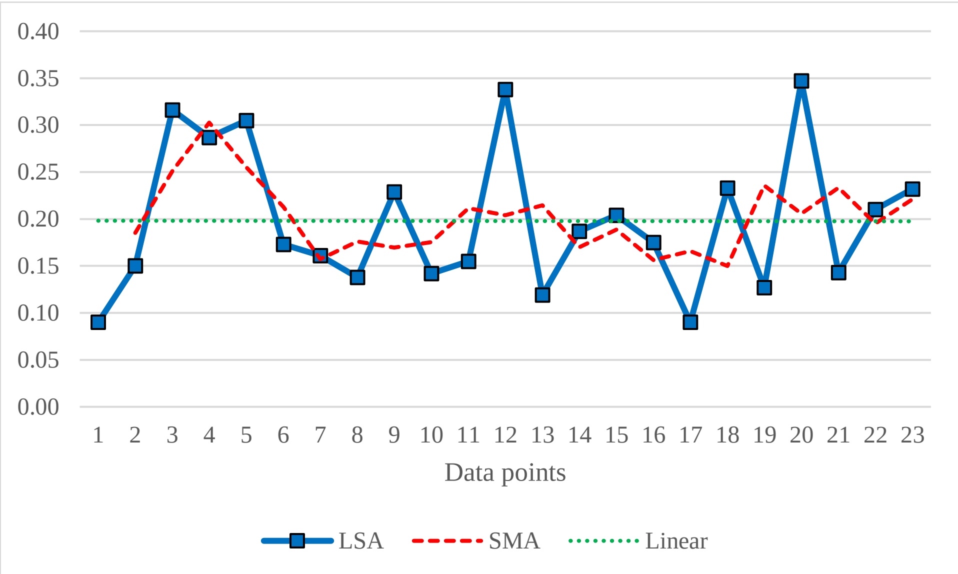


Figure 8.5 Lexical disparity (LSA)

To investigate the extent of genre-relevant lexical choice in Avarka's written data, the percentage of academic words in her essays was estimated by the Academic Word List (AWL) measure. Figure 8.6 (top) shows the four frequency bands: K1 is the list of the most frequent 1000-word families, K2 is the list of the second most frequent 1000-word families, the Academic Word List, and off-list words. Figure 8.6 (top) shows that the highest percentage of K1 words was measured at data point 18 (K1 words = 93.52%), while the lowest percentage of K1 words was detected at data point 5 (K1 words = 75.56%). The low percentage of K1 words at data point 5 was attained by the joint increase of the percentages of K2 words, AWL and off-list words. Indeed, one of the highest percentages of K2 words (9.26%) was detected at data point 5. However, the highest percentage was measured at data point 23 (11.97%). In order to gain a better understanding of the use of the words from the Academic Word List, a separate graph was plotted in Figure 8.6 (bottom).

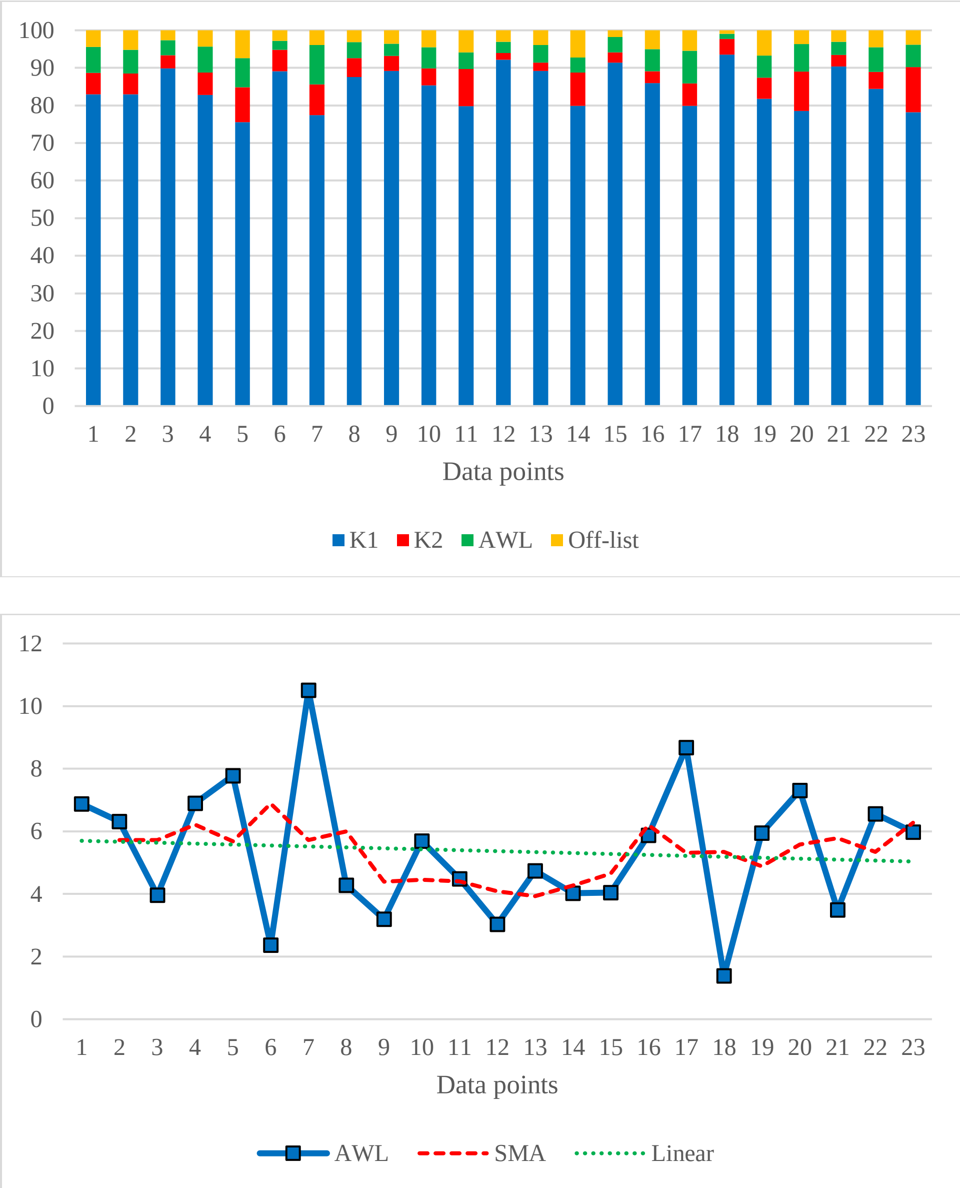


Figure 8.6 Lexical sophistication (top), and Academic Word List (bottom)

The AWL index fluctuated wildly over the nine months, especially between data points 2 and 8, and between data points 16 and 22. A relatively more stable period is visible between data points 8 and 16. The AWL index reached its nadir (1.39%) at data point 18, while the highest AWL index (10.51%) was measured at data point 7. This result indicates that Essay 7 included the highest amount of words from the Academic Word List, whereas Essay 18 contained the fewest words from the AWL list. The smoothing algorithm of the Academic Word List (AWL) index displays a sideways trend over the nine-month investigation. However, the linear trend line displays a slightly downward trend over time.

8.2.2 Syntactic Complexity

Figure 8.7 displays the five syntactic indices used in this study. It can be observed that the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices display similar trajectories since they tap into a similar construct: phrasal complexity. Between data points 12 and 19, even the peaks correspond in the trajectories of FVR and CN/C. In the trajectory of the subordination (DC/C) index, an increase can be observed. In contrast, the FVR index stagnated over the nine months.

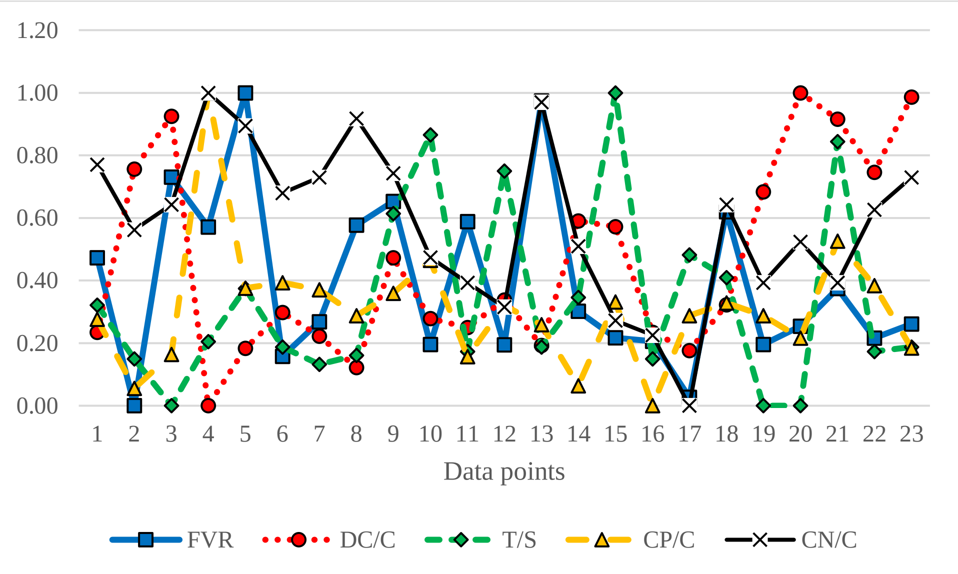


Figure 8.7 Syntactic complexity

General syntactic complexity was measured by the finite verb ratio (FVR) in Avarka's written samples. The FVR index, shown in Figure 8.8, fluctuated wildly between data points 1 and 14. However, between data points 14 and 23, the general syntactic complexity (FVR) index was stable except for a small peak at data point 18. The lowest FVR index was measured (FVR = 8.29) at data point 2. Interestingly, the FVR index reached its zenith (FVR = 11.74) within 3 data points at data point 4. However, another zenith (FVR = 11.65) was also measured at data point 13. The smoothing algorithm of the general syntactic complexity (FVR) index displays a slightly downward trend over the nine-month investigation.

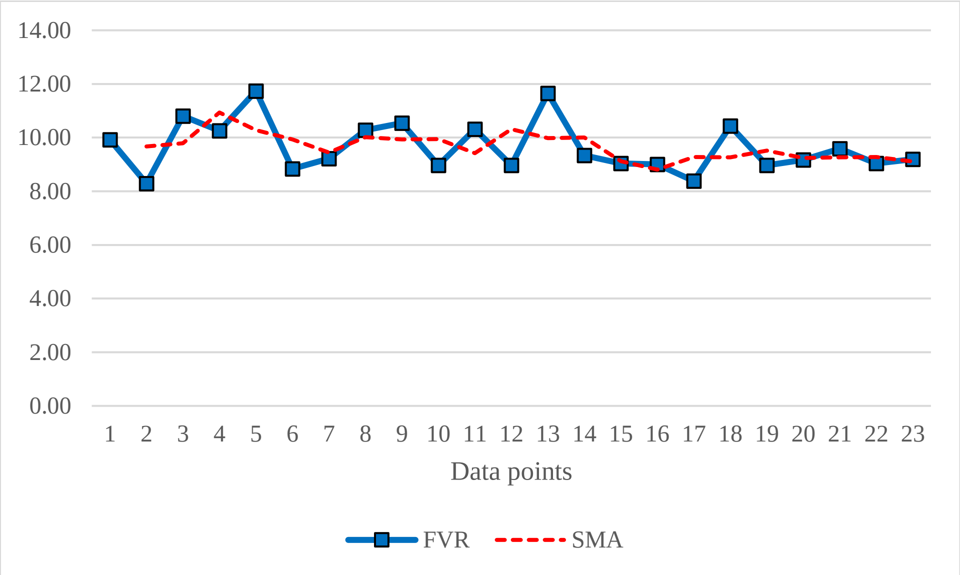


Figure 8.8 General syntactic complexity (FVR)

The following examples, taken from Avarka's Essay 2 and Essay 5, illustrate the difference in the general syntactic complexity (FVR) index. Essay excerpt 8.1 was extracted from Essay 2 where the FVR was the lowest (8.29). In contrast, Essay excerpt 8.2 was taken from Essay 5 where the FVR index was the highest (11.74). The finite verbs are underlined, and the example sentences are left uncorrected.

Essay excerpt 8.1 (Essay 2)

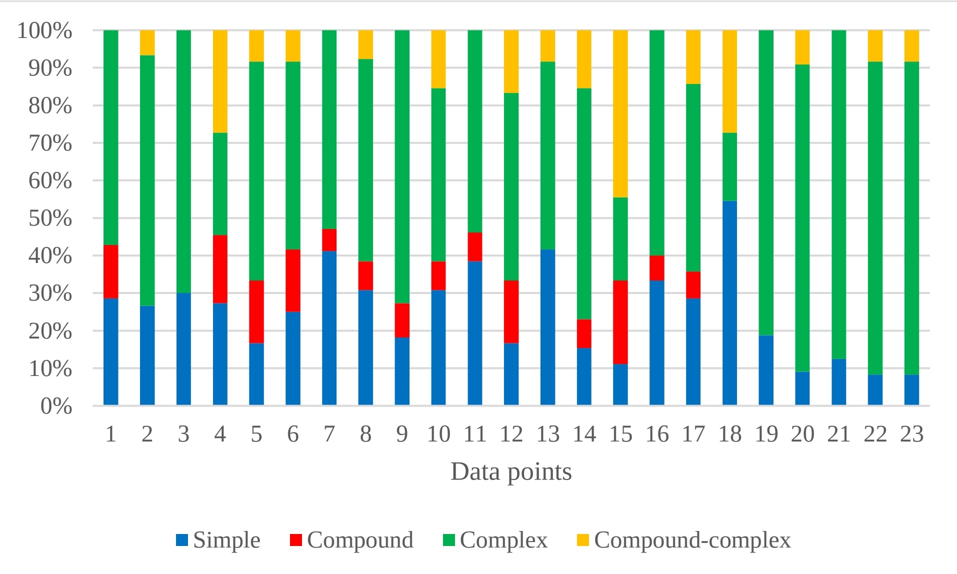
If a child ^{1}<u>is</u> on his or her own, a lot of bad things ^{2}<u>can</u> occur.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 16; finite verb count = 2; FVR = 8)

Essay excerpt 8.2 (Essay 5)

Politicians ^{1}<u>talked</u> to the public with the aid of the radio since during the Second World War the radio ^{2}<u>was</u> the main gadget to contact the people.
(Errors are uncorrected; finite verbs are underlined and numbered in superscript)
(Word count = 27; finite verb count = 2; FVR = 13.5)

The two sentences exhibit different levels of syntactic complexity, although they are the same type of sentence i.e. complex. The ratio of the words per finite verbs clearly indicates that Essay excerpt 8.2 (FVR = 13.5) is more complex than Essay excerpt 8.1 (FVR = 8).

Avarka's written samples were also coded for sentence types in order to gain a better insight into the syntactic structures she used in her academic writing. The distribution of the different sentence types is presented in a bar chart in Figure 8.9.



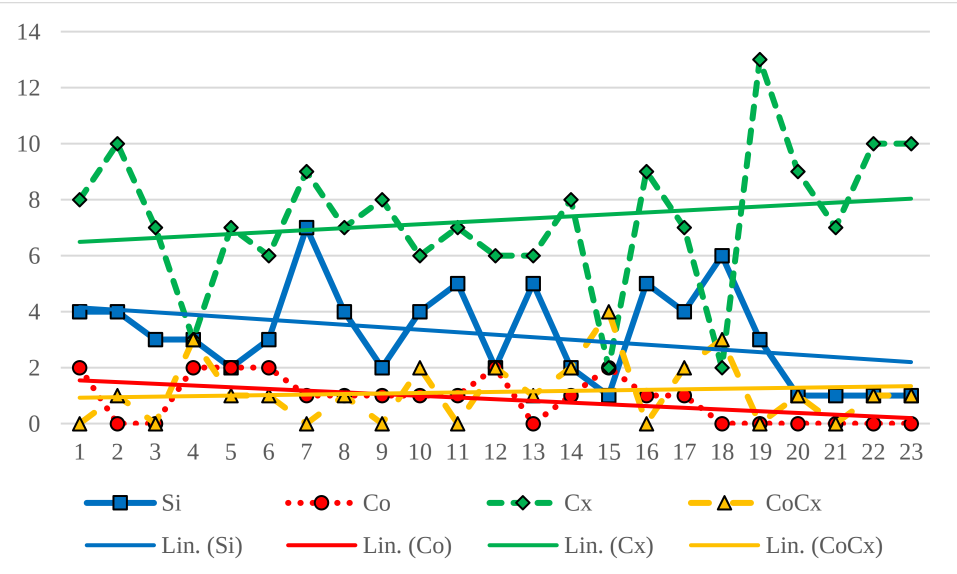


Figure 8.9 Sentence types

Figure 8.9 (top) shows the distribution of sentence types in Avarka's academic writing. Visual inspection of Figure 8.9 indicates that complex sentences dominated in Avarka's written data. Indeed, 58% of the total sentences were complex sentences. The second most common sentence type were simple sentences (26%). The compound and the compound-complex sentences were employed less frequently in Avarka's written data. The third most frequent sentence type were compound-complex sentences (9%), while only 7% of the total sentences were compound sentences. The general syntactic complexity (FVR) index suggested a decrease of the phrasal complexity in Figure 8.8. Indeed, a decreasing trend can be observed in the use of simple sentences (Figure 8.9, bottom). Conversely, an increasing trend can be observed in the use of complex sentences. This phenomenon does not corroborate the current findings of academic writing which suggest that advanced users use more complex nominalisation rather than clausal embeddings in academic writing (Biber et al., 2011). Furthermore, the increase of the complex sentences in Avarka's data is also contrary to Dalma's data (Chapter 5), where an increase in the use of simple sentences and a decrease in the use of complex sentences can be observed. Conversely, the trend found in Avarka's data is similar to Emese's (Chapter 6) and Levente's (Chapter 7) data in which the complex sentences increased over the nine months. Figure 8.9 also shows that simple sentences were present in each essay, however compound sentences were only present in 14 of the 23 essays. For example, in the last six essays Avarka did not use any compound sentences in her essays. Complex sentences were found in every essay, whereas compound-complex sentences were omitted in eight essays over the nine months.

Avarka made her writing more elaborate by employing different syntactic structures like subordination, coordination and phrasal complexity. Subordination was gauged in this study by the ratio of dependent clauses to clauses (DC/C) index. Figure 8.10 shows that the DC/C index was stable between data points 4 and 23, except for a small hump between data points 1 and 4. The smoothing algorithm of the subordination (DC/C) index displays a slightly upward trend over the nine-month investigation. The lowest DC/C index value (DC/C = 0.30) was measured at data point 4, while the highest (DC/C = 0.54) at data point 20. This finding means that every second clause was a dependent clause in Essay 20 as illustrated in Essay excerpt 8.3.

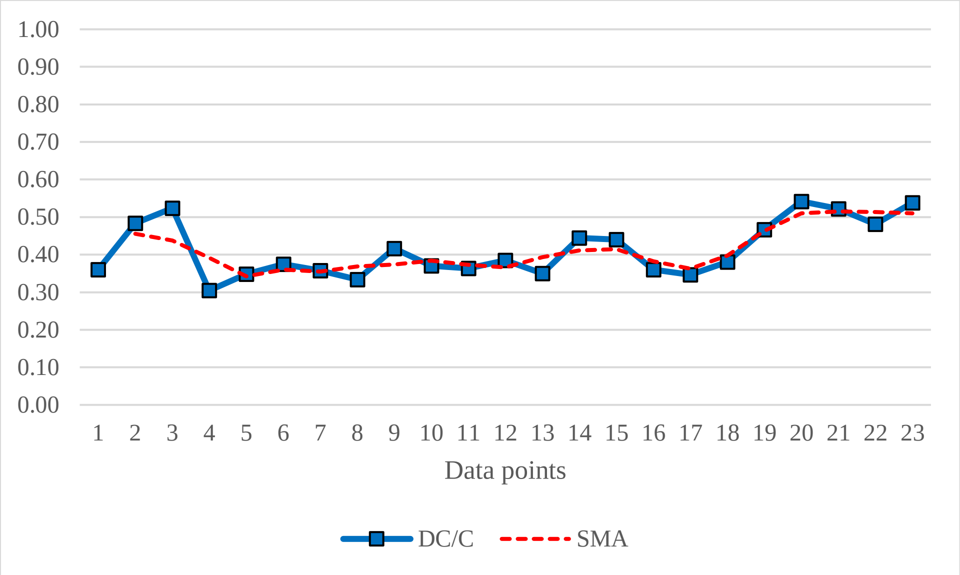


Figure 8.10 Subordination (DC/C)

Essay excerpt 8.3 (Essay 20)

There is no difference between traditional and international music [1]because the two types have different purposes.
(Errors are uncorrected; dependent clauses are underlined and numbered in superscript)
(Dependent clause count = 1; clause count = 2; DC/C = 0.5)

In addition to subordination, Avarka also employed both clausal and phrasal coordination syntactic structures in her essays. The clausal coordination (T/S) index, shown in Figure 8.11, displays two distinct periods. The first period is stable between data points 1 and 8, while the T/S index started to fluctuate more intensely in the second period between data points 8 and 23. The smoothing algorithm of the clausal coordination (T/S) index (red dashed line) shows a sideways trend over the nine-month investigation. However, the linear trend line (green dotted line) displays a slightly upward trend over time.

In addition to clausal coordination, Avarka also employed phrasal coordination structures which consequently made her writing more complex. Figure 8.11 shows that the phrasal coordination (CP/C) index reached its zenith (CP/C = 0.58) at data point 4, while the lowest CP/C index (CP/C = 0) was measured at data point 16.

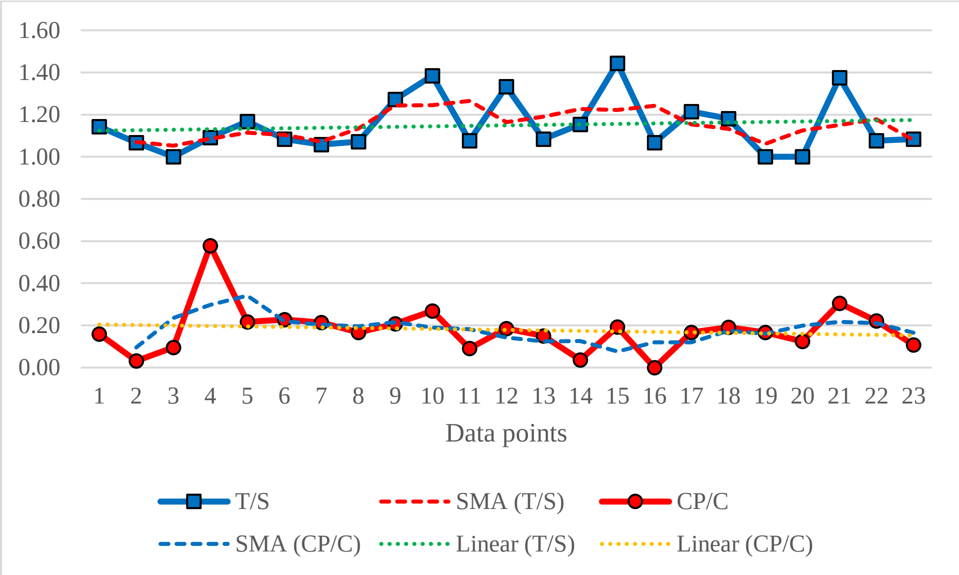


Figure 8.11 Coordination (T/S and CP/C)

This result might indicate that on average one coordinate phrase was used in every second clause in Essay 4. See Essay excerpt 8.4 for an example.

Essay excerpt 8.4 (Essay 4)

However, [1]<u>the most popular channels and newspapers</u> usually cover a lot of different topics.
(Errors are uncorrected; coordinate phrases are underlined and numbered in superscript)
(Coordinate phrase count = 1; clause count = 1; CP/C = 1)

The smoothing algorithm of the phrasal coordination (CP/C) index (blue dashed line) displays a sideways trend over the nine-month investigation. However, the linear trend line (yellow dotted line) displays a slightly downward trend over the nine months.

Phrasal complexity was measured by the complex nominal per clause (CN/C) and the mean number of modifiers per noun phrase (SYNNP) indices. Figure 8.12 shows that the CN/C index fluctuated wildly over the nine months. The highest CN/C value (CN/C = 1.58) was measured at data point 4, while the lowest value (CN/C = 0.63) was detected towards the end of the investigation at data point 17. The smoothing algorithm of the phrasal complexity (CN/C) index displays a downward trend between data points 2 and 16. However, the smoother of the CN/C index shows an upward trend between data points 16 and 23.

Figure 8.12 (bottom) also shows the mean number of modifiers per noun phrase (SYNNP) index. The lowest SYNNP index (SYNNP = 0.673) was measured at data point 6, while the highest SYNNP index (SYNNP = 1.015) was gauged at data point 14. The smoothing algorithm of the SYNNP index displays a sideways trend over the nine-month investigation. However, the linear trend shows a slightly downward trend over the nine months.

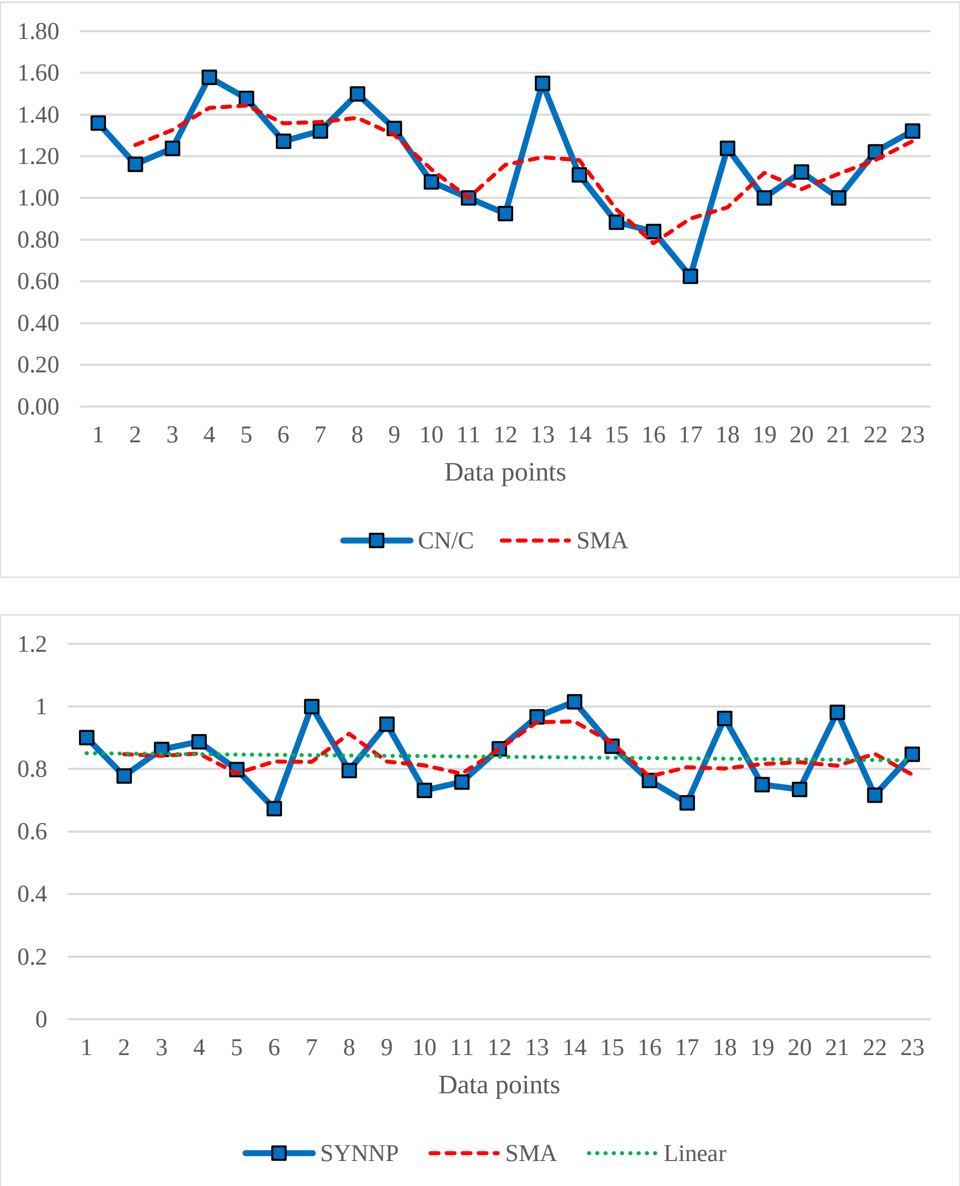


Figure 8.12 Phrasal complexity (CN/C and SYNNP)

Essay excerpt 8.5 exemplifies the high phrasal complexity (CN/C) index value (CN/C = 1.58) measured at data point 4.

Essay excerpt 8.5 (Essay 4)

Nowadays, there is a [1]wide range of newspapers, radio stations and television channels, and the [2]variety of news (that) they present, depends on their [3]main focus of topic. (Errors are uncorrected; complex nouns are underlined and numbered in superscript) (Complex nominal count = 3; clause count = 3; CN/C = 3)

Figure 8.13 shows the normed rate of occurrence of conditional clauses (ConC), the normed rate of occurrence of infinitive clauses (InfC), the normed rate of occurrence of relative clauses (RelC), and the incidence score of prepositional phrases (DRPP) indices.

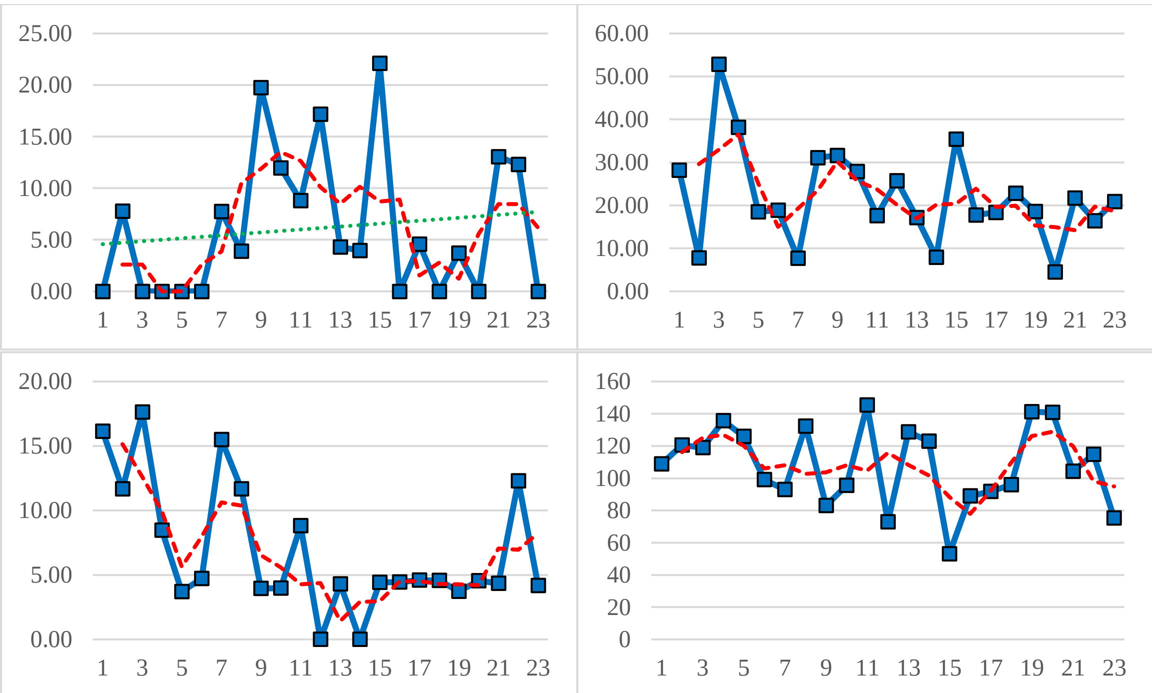


Figure 8.13 Syntactic measures specific to the academic genre (top left: ConC; top right: InfC; bottom left: RelC; bottom right: DRPP).

The smoothing algorithms of the infinitive clauses (InfC), the relative clauses (RelC) and the prepositional phrases (DRPP) display downward trends over the nine-month investigation. The smoothing algorithm of the conditional clauses (ConC) index displays a sideways trend.

However, the linear trend line of the conditional clauses (ConC) index shows an upward trend over the nine months. In other words, there was a decrease in the usage of the infinitive and relative clauses, prepositional phrases. However, there was an increase in the usage of conditional clauses (ConC) in Avarka's written data.

8.2.3 Accuracy

In addition to the constructs of syntactic and lexical complexity, accuracy was also measured in Avarka's written samples by the ratio of error-free clauses to the total number of clauses (EFC/C) index. As shown in Figure 8.14, in the trajectory of the EFC/C index two distinct periods can be distinguished.

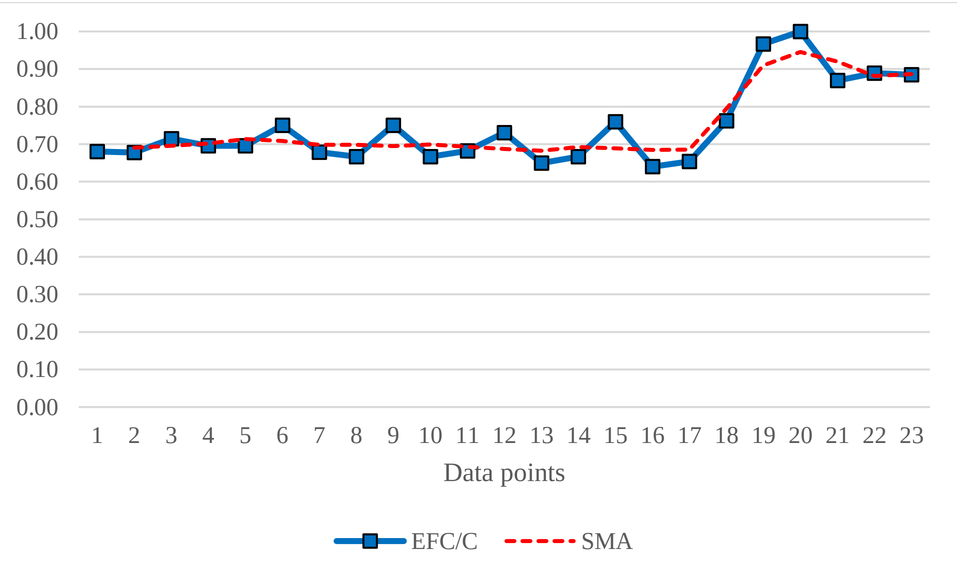


Figure 8.14 Accuracy

The EFC/C index fluctuated between 0.64 and 0.76 between data points 1 and 18 in the first period. However, the EFC/C index rocketed at data point 19 and reached its zenith (EFC/C = 1.00) at data point 20. In the second period, between data points 19 and 23 the EFC/C index was relatively higher than in the first period since it fluctuated between 0.87 and 1.00. The smoothing algorithm of the accuracy (EFC/C) index displays a sideways trend between data points 2 and 16. However, the smoothing algorithm of the EFC/C index shows an upward trend from data point 16 which indicates an improvement in the construct of accuracy in Avarka's written data. The lowest EFC/C value (EFC/C = 0.64) was measured at data point 16, while the highest EFC/C value (EFC/C = 1.00) was detected at data point 20. The 1.00 EFC/C value means that there were no errors in Essay 20. The lowest EFC/C value (0.64) means that there was at least one error in every second clause in Essay 16. Essay excerpt 8.6 illustrates the low accuracy rate in Essay 16.

Essay excerpt 8.6 (Essay 16)

It can be true that the change of the women's position in our current [1]<u>societe</u> might have affected the life of young people.
(Errors are uncorrected, underlined and numbered in superscript)
(Error-free clause count = 1; clause count = 2; EFC/C = 0.5)

In order to gain a deeper insight into the accuracy construct of Avarka's writing, this study also looked at the type of errors she made in her essays. The errors were coded and categorised according to the Louvain Error Tagset, and the results are displayed in Figure 8.15. Figure 8.15 shows that the number of errors decreased over the nine months, especially from data point 19. This result corresponds to the accuracy (EFC/C) index which also indicated a change in trend around data point 19. Figure 8.15 shows that Avarka made five different types of errors in her essays: (1) formal, (2) grammatical, (3) lexico-grammatical, (4) lexical, and (5) word redundant, word missing and word order. Among the five different types of errors formal errors were the most prevalent (57%) in her written data. Formal errors were evident in every essay except in Essay 20, in which no errors were made. The second most common type of errors belonged to the grammatical errors category (29%).

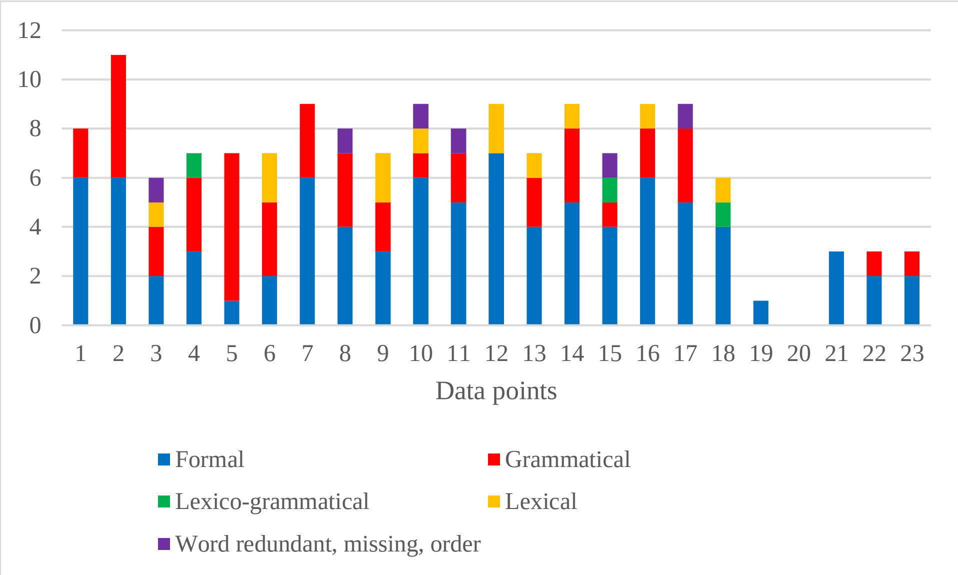


Figure 8.15 Error types

Grammatical errors were present in every essay except for Essay 12, Essay 18, Essay 19, Essay 20, and Essay 21. The third most frequent type of errors belonged to the lexical errors group (8%). However, lexical errors were present in only nine essays out of the 23 essays over the nine months.

8.3 The Degree of Variability

Section 8.3 explores the degree of variability in lexical and syntactic complexity and accuracy in Avarka's data. This section is subdivided into three parts: (1) lexical complexity, (2) syntactic complexity, and (3) accuracy.

8.3.1 Lexical Complexity

The min-max graph of the lexical rarity (WRDFRQc) index (Figure 8.16) shows three clear stages. In the first stage the bandwidth of variability was narrow (from 1.03 to 1.35) between data points 1 and 10. In the second stage the bandwidth of variability became wider (from 0.75 to 1.62) between data points 11 and 16. In the third stage the bandwidth became narrow again (from 1.00 to 1.50) between data points 17 and 23. These ranges can be interpreted as follows. In the first and the third stages the less frequent lexical items were used in a more regular pattern. Conversely, in the second stage less frequent lexical items were employed randomly in Avarka's written data.

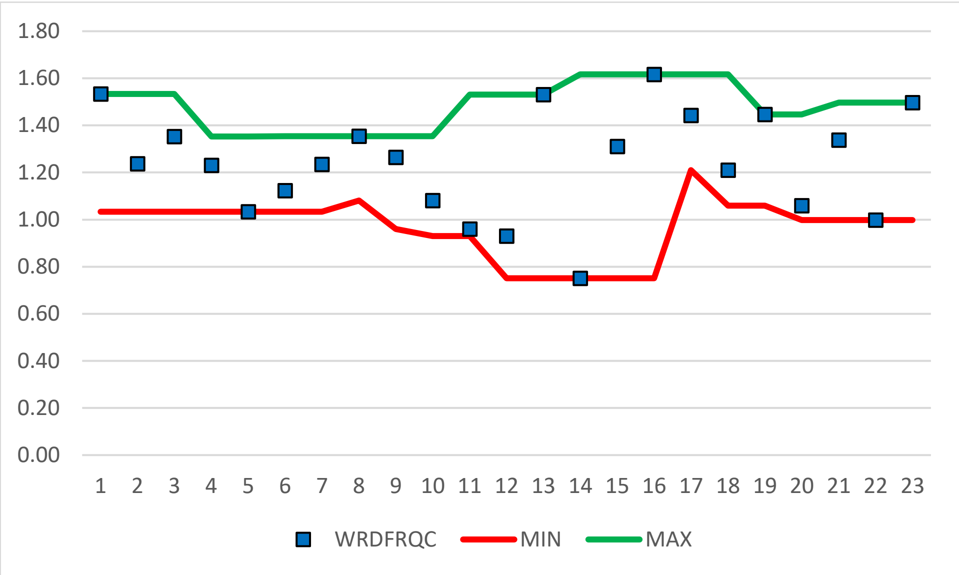


Figure 8.16 Lexical rarity (min-max graph)

The data of the lexical rarity (WRDFRQc) index were resampled, and a Monte Carlo analysis was run. Table 8.2 shows that the p-value of the lexical rarity (WRDFRQc) index was 0.13. In other words, a statistically significant developmental peak was not detected in the lexical rarity (WRDFRQc) index. The same statistical procedures were repeated for the general lexical complexity (AveWL), the lexical variability (TTR), the lexical disparity (LSA) and the Academic Word List (AWL) indices. However, the p-values of the AveWL, the TTR, the LSA and the AWL indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the data of the AveWL, the TTR, the LSA, and the AWL indices.

Table 8.2

P-values for the Complexity and Accuracy Indices

Index	P-value
AveWL	0.47
TTR	0.96
WRDFRQc	0.13
LSA	0.38
AWL	0.82
FVR	0.60
DC/C	0.17
T/S	0.52
CP/C	0.13
CN/C	0.85
EFC/C	0.01*

Note. *Statistically significant at 0.05 level

8.3.2 Syntactic Complexity

In the min-max graph of the phrasal coordination (CP/C) index (Figure 8.17) there were two main stages.

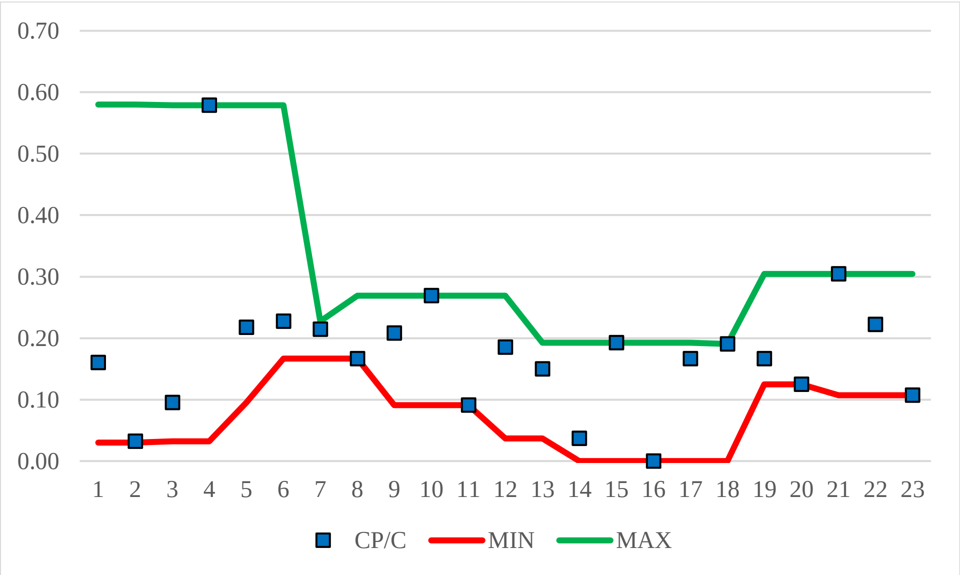


Figure 8.17 Phrasal coordination (min-max graph)

In the first stage the bandwidth of variability was wide (from 0.03 to 0.58) from data point 1 to 6. Conversely, in the second stage the bandwidth was narrow between data points 7 and 23. In addition, in the second stage, the bandwidth moved down between data points 7 to 18 and then it moved up between data points 18 and 23. However, during these moves the width of the band did not change substantially. These ranges can be interpreted as follows. In the first stage of the bandwidth phrasal coordination was employed randomly, while in the second stage phrasal coordination structures were used in a more regular pattern.

Table 8.2 shows that the p-values of the five syntactic complexity indices were above 0.05. In other words, statistically significant developmental peaks were not detected in the five syntactic complexity indices in Avarka's written data.

8.3.3 Accuracy

In the min-max graph of the EFC/C index (Figure 8.18), there were three main stages.

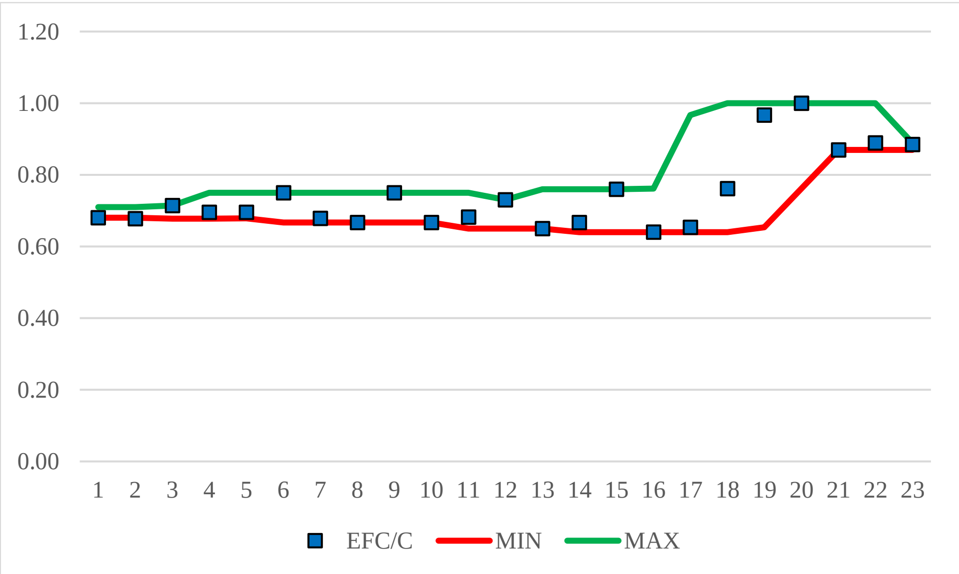


Figure 8.18 Accuracy (min-max graph)

In the first stage, the bandwidth of variability was narrow (from 0.67 to 0.75) from data point 1 to 16. In the second stage the bandwidth became wider (from 0.67 to 1.00) between data points 17 and 20. In the third stage, the bandwidth of variability was narrow again (from 0.87 to 1.00). However, the min-max graph clearly shows that the band moved up because the third stage was higher than the first stage. It might be speculated that the developmental peak at data point 20 might be statistically significant.

Table 8.2 shows that the p-value of the accuracy (EFC/C) index was 0.01. In other words, a statistically significant developmental peak was detected in the accuracy (EFC/C) index in Avarka's written data.

8.4 Interactions

The purpose of this section is to explore the dynamic interactions between lexical and syntactic complexity and accuracy. This section is subdivided into two parts: (1) interactions between general lexical and syntactic complexity and accuracy indices and (2) interactions between lexical variability, subordination, and accuracy indices.

8.4.1 Interactions Between General Lexical and Syntactic Complexity and Accuracy

Figure 8.19 shows (top) the normalised data of the general lexical complexity (AveWL) and general syntactic complexity (FVR) and the accuracy (EFC/C). The trajectories of the general lexical complexity (AveWL) and the general syntactic complexity (FVR) exhibit both alternating and parallel patterns. There were clear parallel shifts between data points 2 and 5 and between data points 10 and 13. Conversely, there were clear alternating shifts between data points 5 and 8. The trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) show both parallel and alternating variability patterns. Likewise, the trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) also show both alternating and parallel variability patterns.

The residual plot in Figure 8.19 (bottom) shows that the trajectories of the general lexical (AvewL) and syntactic complexity (FVR) indices were both parallel and alternating. Likewise, the trajectories of the general lexical complexity (AveWL) and accuracy (EFC/C) display both parallel and alternating shifts over the nine months. The trajectories of the general syntactic complexity (FVR) and accuracy (EFC/C) were also parallel and alternating over the investigated period.

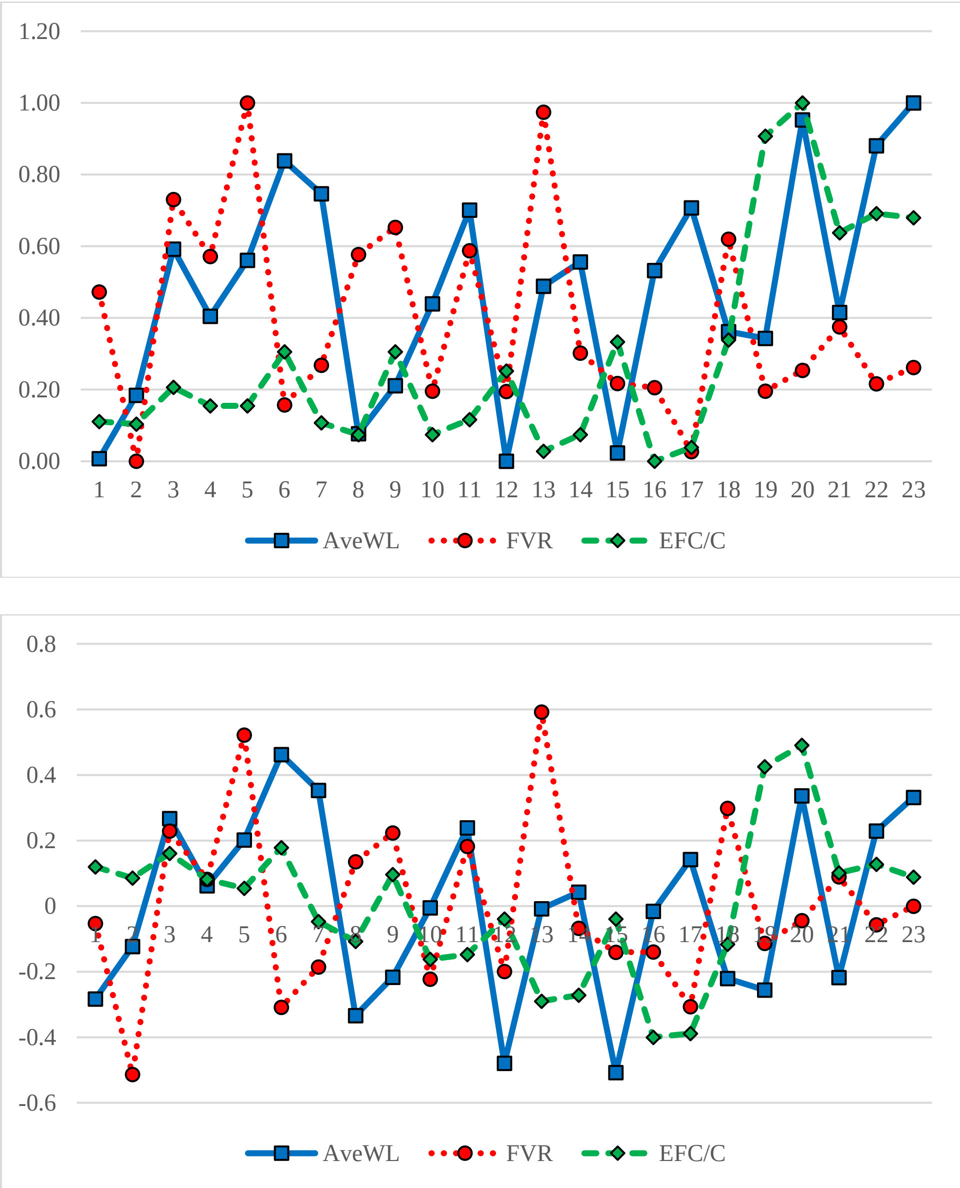


Figure 8.19 General lexical and syntactic complexity and accuracy (top: normalised data; bottom: residual plot)

Figure 8.20 shows the interactions between general lexical (AveWL) and syntactic complexity (FVR) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 8.20 shows that the general lexical complexity (AveWL)-general syntactic complexity (FVR) correlation shifted between strongly positive and strongly negative values. The general lexical complexity (AveWL)-accuracy (EFC/C) correlation shifted between strongly positive and moderately negative values over the nine months. The general syntactic complexity (FVR)-accuracy (EFC/C) correlation shifted between strongly positive and strongly negative values over the investigated period.

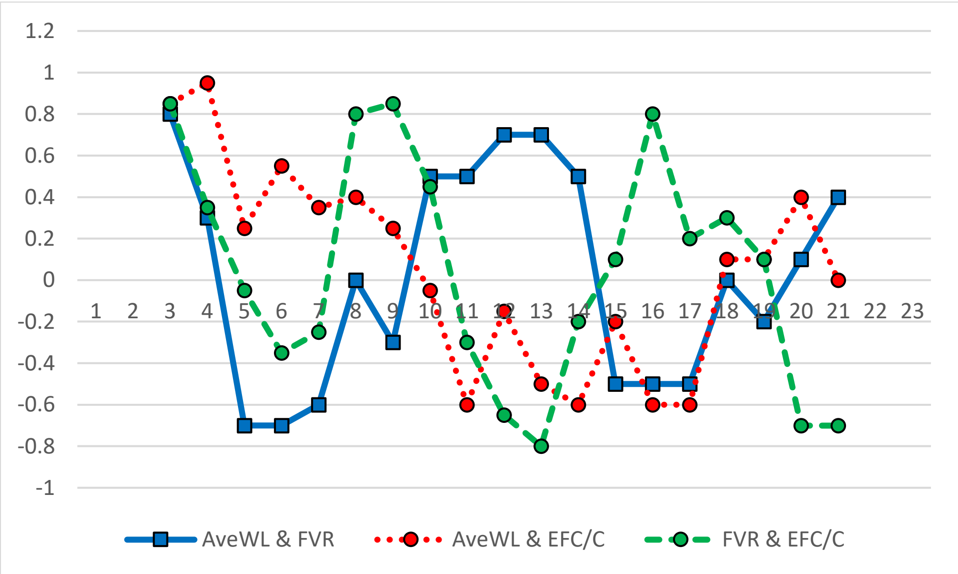


Figure 8.20 Moving correlation plot (General lexical and syntactic complexity and accuracy)

Table 8.3 shows the Spearman's rank correlation coefficient values between the general lexical (AveWL) and syntactic complexity (FVR) and accuracy (EFC/C) indices. The correlation between general lexical complexity (AveWL) and general syntactic complexity (FVR) was negative but very weak (-.033). The correlation between general lexical complexity (AveWL) and accuracy (EFC/C) was positive but also weak (.144). The correlation between general syntactic complexity (FVR) and accuracy (EFC/C) was also negative and extremely weak (-.010).

Table 8.3

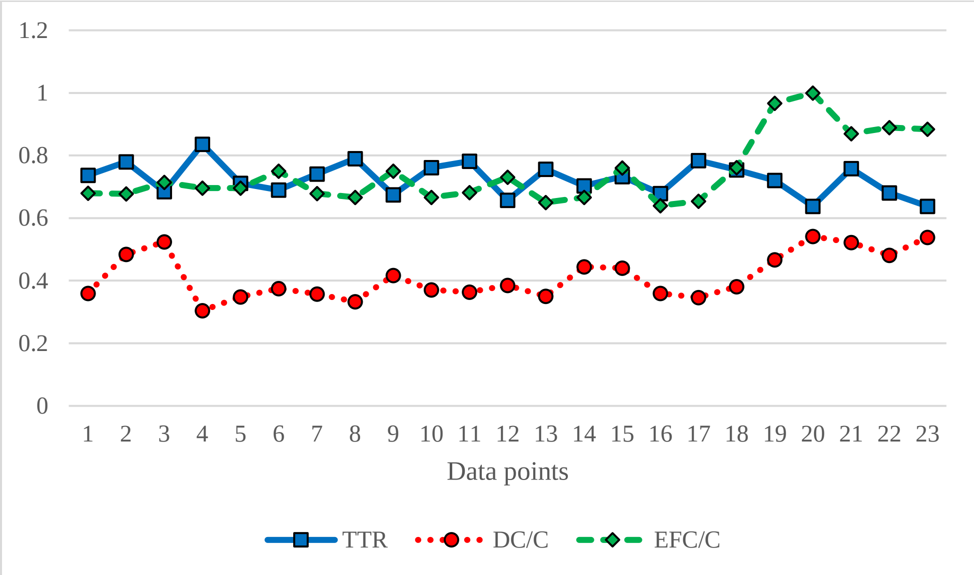
Correlation Matrix (General Lexical and Syntactic Complexity and Accuracy)

		AveWL	FVR	EFC/C
AveWL	ρ	1.000	-.033	.144
	Sig. (2-tailed)		.882	.513
	N	23	23	23
FVR	ρ	-.033	1.000	-.010
	Sig. (2-tailed)	.882		.962
	N	23	23	23
EFC/C	ρ	.144	-.010	1.000
	Sig. (2-tailed)	.513	.962	
	N	23	23	23

8.4.2 Interactions Between Lexical Variability, Subordination and Accuracy

Figure 8.21 (top) shows that the growth trajectories of lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices. The trajectories of the lexical variability (TTR) and subordination (DC/C) were predominantly alternating. Likewise, the trajectories of the lexical variability (TTR) and accuracy (EFC/C) indices were mainly alternating over the nine months. Conversely, the trajectories of the subordination (DC/C) and accuracy (EFC/C) were predominantly parallel over the investigated period. Interestingly, even the local fluctuations seem to correspond between data points 19 and 20.

Figure 8.21 (bottom) shows the residual plot of the lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices. The residual plot confirms the initial observation of the data plot. The trajectories of the lexical variability (TTR) and subordination (DC/C) were mainly alternating. Likewise, the trajectories of the lexical variability (TTR) and accuracy (EFC/C) were predominantly alternating. Conversely, the growth trajectories of the subordination (DC/C) and accuracy (EFC/C) were mainly parallel. For example, there were clear parallel variability patterns between data points 5 and 14 (except for the shift between data points 10 and 11).



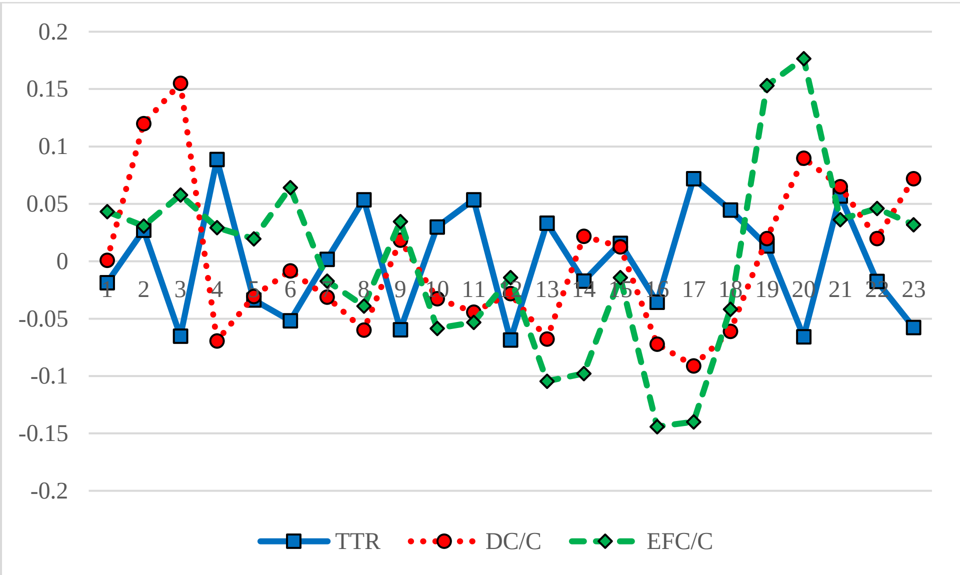


Figure 8.21 Data and residual plot (Lexical variability, subordination & accuracy)

Figure 8.22 shows the interactions between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) indices as a moving correlation in a window of 5 measurements. Figure 8.22 shows that the lexical variability (TTR)-subordination (DC/C) correlation was negative for the entire duration of the investigation. The lexical variability (TTR)-accuracy (EFC/C) correlation shifted between moderately negative and moderately positive values, with two peaks towards strongly negative values in the 7th window and the 19th window. The subordination (DC/C)-accuracy (EFC/C) correlation shifted between moderately and strongly positive values with one peak toward a weak negative value in the 3rd window.

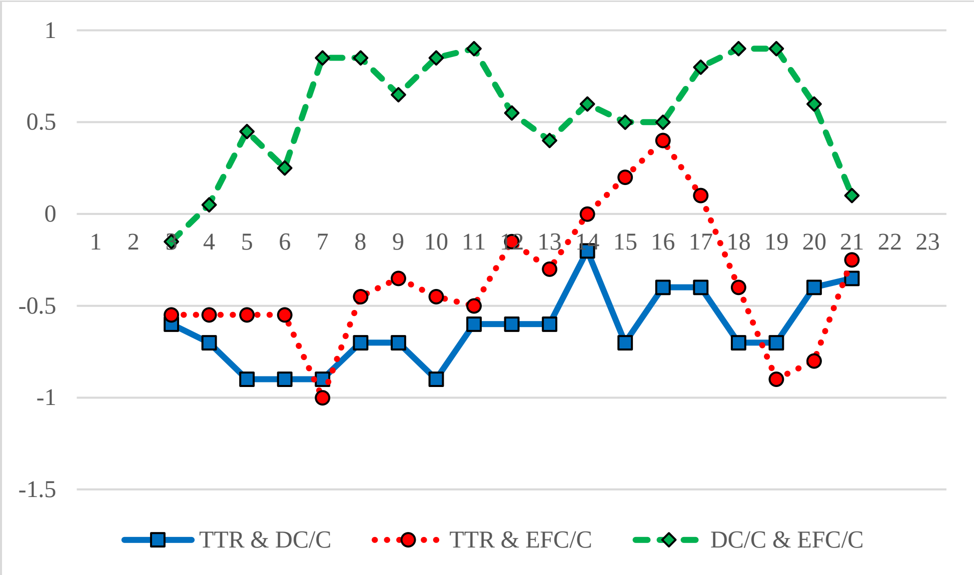


Figure 8.22 Moving correlation (Lexical variability, subordination, and accuracy)

Table 8.4 shows the correlation coefficient values between the lexical variability (TTR), subordination (DC/C) and the accuracy (EFC/C) indices. The correlation between lexical variability (TTR) and subordination (DC/C) was moderately negative (-.579) and statistically significant (p<0.01). The correlation between lexical variability (TTR) and accuracy (EFC/C) was also moderately negative (-.472) and statistically significant (p<0.05). The correlation between subordination (DC/C) and accuracy (EFC/C) was moderately positive (.658) and statistically significant (p<0.01).

Table 8.4

Correlation Matrix (Lexical Variability, Subordination, and Accuracy)

		TTR	DC/C	EFC/C
TTR	ρ	1.000	-.579**	-.472*
	Sig. (2-tailed)		.004	.023
	N	23	23	23
DC/C	ρ	-.579**	1.000	.658*
	Sig. (2-tailed)	.004		.001
	N	23	23	23
EFC/C	ρ	-.472*	.658*	1.000
	Sig. (2-tailed)	.023	.001	
	N	23	23	23

Note. *Correlation was significant at the 0.05 level (2-tailed);** Correlation was significant at the 0.01 level (2-tailed).

8.5 Modelling

The purpose of this section is to test the hypothesised interactions between lexical variability (TTR) and subordination (DC/C), between lexical variability (TTR) and accuracy (EFC/C), and between subordination (DC/C) and accuracy (EFC/C) by simulating the developmental processes of these three constructs. Based on the variability analyses (section 8.4.2) the following hypotheses were formulated:

1. a moderate competition between lexical variability (TTR) and subordination (DC/C),
2. a moderate competition between lexical variability (TTR) and accuracy (EFC/C),
3. a moderate support between subordination (DC/C) and accuracy (EFC/C).

To test the abovementioned hypotheses, a model of three connected growers was employed.

8.5.1 Model Setup

Three growers were specified in the model which corresponded with the three constructs under investigation in the previous section (Section 8.4.2) that is: (1) lexical variability (TTR), (2) subordination (DC/C), and (3) accuracy (EFC/C). Table 8.5 shows the numerical values that were assigned to the parameters in the model. The second step included the configuration of the relational parameters. Table 8.5 also shows the relational parameters used in Avarka's data.

Table 8.5

Property and Relational Parameters

		Grower A	Grower B	Grower C
Property parameters	Initial value	0.737	0.36	0.68
	Rate	-0.01	0.01	0.05
	Carrying capacity	1	1	1
Relational parameters		to A	to B	to C
	from A	-	-0.029	-0.023
	from B	-0.029	-	0.033
	from C	-0.023	0.033	-

After specifying the property and relational parameters the data were simulated. The outcome of 100 iterations of the configured equations was obtained by the data simulation. Figure 8.23 shows the result of the simulations after 100 iterations.

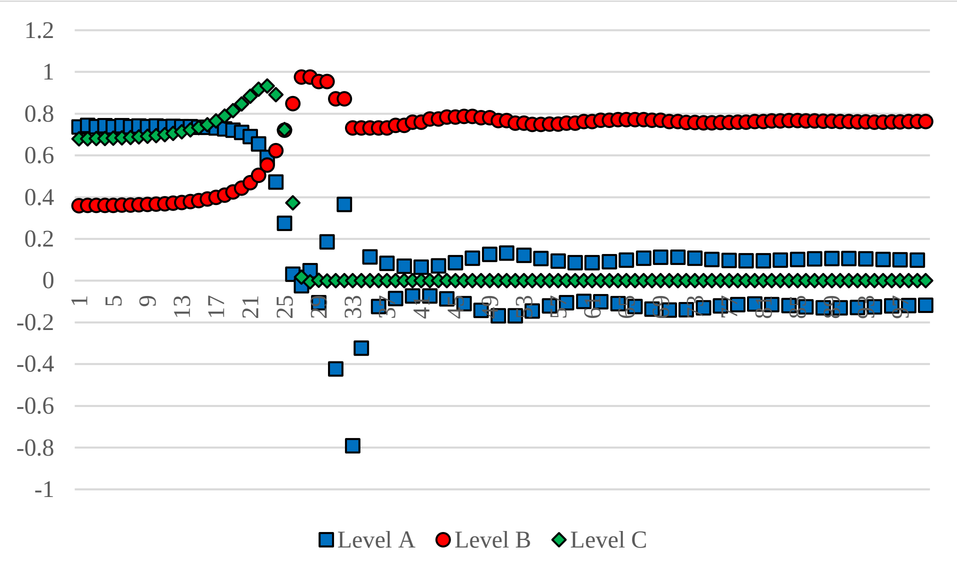


Figure 8.23 Growth model

8.5.2 Optimization

The purpose of the optimization was to minimize the outcome of the function, in this case the sum of squared differences between the model and the data. Table 8.6 shows the optimized parameter values.

Table 8.6

Optimized Parameter Values

		Grower A	Grower B	Grower C
Property parameters	Rate	-0.00995	0.01006	0.04994
Relational		to A	to B	to C
parameters	from A	-	-0.51695	0.60866
	from B	-4.80875	-	-0.64570
	from C	1.92806	0.35522	-

The optimization of the parameters made it possible to assess the overall fit of the model to the data. First, the sum of squared residuals between the model and the data was compared, which was 0.22272 in Avarka's data. Second, the model outcome was visually contrasted with the data. Figure 8.24 shows that model resembles the data reasonably well.

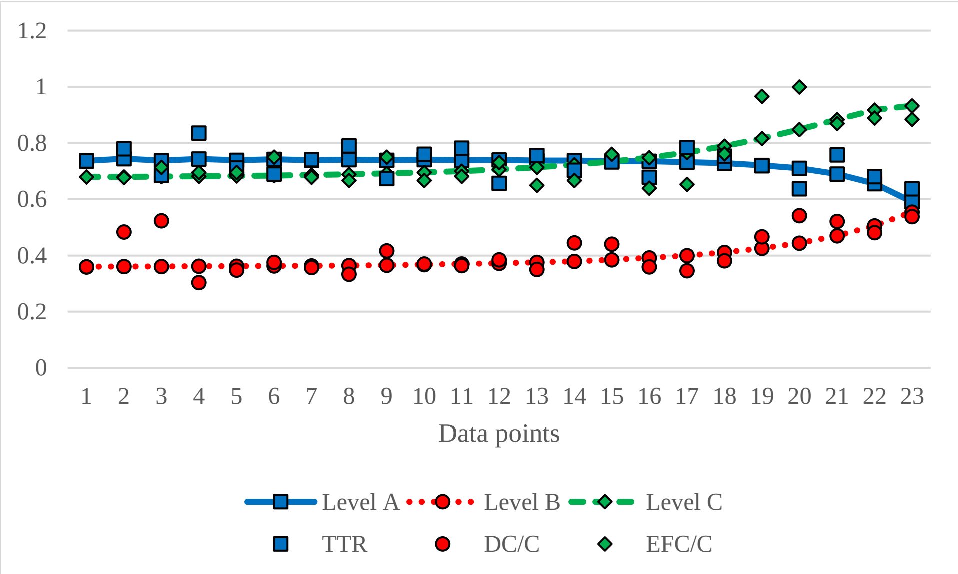


Figure 8.24 Data and model (fit = 0.22272)

The optimized parameter values confirm only one of the three hypotheses that informed the model setup. The aggregated negative parameter value (Table 8.6) between Level A and B confirms the hypothesised moderately competitive relationship between lexical variability (TTR) and subordination (DC/C). However, the optimized parameter values suggest that there was a strong competition between Level A and B since both directions were highly negative (from Level A to Level B and from Level B to Level A). The aggregated parameter value between Level A and C does not confirm the hypothesised negative relationship between lexical variability (TTR) and accuracy (EFC/C) since both directions (from Level A to C and from Level C to A) were positive. The aggregated parameter value between Level B and C does not confirm the hypothesised positive relationship between subordination (DC/C) and accuracy

(EFC/C). Although the parameter value from Level C to B was positive (0.35522), the parameter value from Level B to C was negative (-0.64570).

8.5.3 Evaluation of the Model

Figure 8.25 presents the data, the linear trends, the smoothing splines, and the model outcome. Visual inspection of these plots indicates that the Level A, Level B and Level C resemble a combination of the linear trends and the smoothing splines.

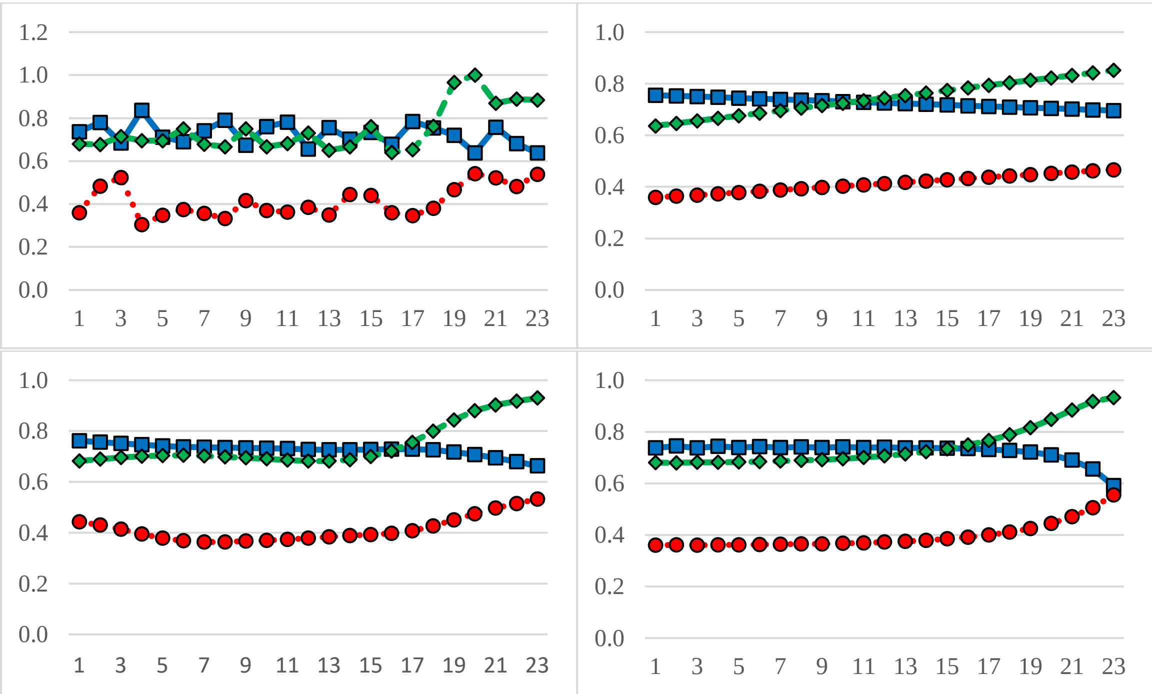


Figure 8.25 Data (top left), linear trends (top right), smoothing splines (bottom left) and model (bottom right)

The model was correlated with the data, the linear trends, and the smoothing splines. Table 8.7 shows that the model positively correlated at all three levels with the original data.

Table 8.7

Correlations Between the Model and the Data, Linear Trends, and Splines

			Model-data	Model-linear	Model-spline
Level A	TTR	ρ	.379	.820**	.767**
		Sig. (2-tailed)	.074	.000	.000
		N	23	23	23
Level B	DC/C	ρ	.423*	.999**	.515**
		Sig. (2-tailed)	.044	.000	.012
		N	23	23	23
Level C	EFC/C	ρ	.448*	.999**	.650**
		Sig. (2-tailed)	.032	.000	.001
		N	23	23	23

Note. * Correlation was significant at the 0.05 level (2-tailed);** Correlation was significant at the 0.01 level (2-tailed).

Correlation matrixes between the various categories in the data, the linear trends, the smoothing splines, and the model were compared to see whether the model replicated the surface level interactions of the data. Table 8.8 shows the correlation matrixes for the data, the linear trends, the smoothing splines, and the model.

Table 8.8

Correlation Matrixes

			Data	Linear	Spline	Model
TTR	DC/C	ρ	-.579**	-1.000**	-.467*	-.806**
		Sig. (2-tailed)	.004	.000	.026	.000
		N	23	23	23	23
TTR	EFC/C	ρ	-.472*	-1.000**	-.534**	-.833**
		Sig. (2-tailed)	.023	.000	.009	.000
		N	23	23	23	23
DC/C	EFC/C	ρ	.658*	1.000**	.540**	.997**
		Sig. (2-tailed)	.001	.000	.009	.000
		N	23	23	23	23

Note. *Correlation was significant at level 0.05 (2-tailed); ** Correlation was significant at level 0.01 (2-tailed).

Table 8.8 shows that correlation matrixes in the data, the linear trends, the smoothing splines and the model outcome show very similar patterns, particularly those of the linear trends and the model outcome. The polarity of the correlation coefficients in all combinations (TTR-DC/C, TTR-EFC/C and DC/C-EFC/C) in the data, the linear trend, the smoothing splines and the model were the same. However, the correlation coefficients in the model rather resemble the correlation coefficients in the linear trends than the data and the smoothing splines.

8.6 Motivation

The purpose of this section is to explore the changes in Avarka's motivational system. This section is subdivided into four parts: (1) goals for academic writing, (2) self-regulatory processes and (3) co-adaptation.

8.6.1 Goals for Academic Writing

During the initial interview, Avarka reported that she would write essays and research papers in English language during her future studies at university. She also added that she would like to publish research papers. Avarka's main goal was to gain a "confident knowledge "in English language and to study abroad. Avarka reported that she would write articles, essays and papers in her future career and occupation. Her usual method of writing includes collecting ideas, composing and revising. She also said that she usually re-writes her essays if she has time to do that. Avarka said that she would like to write in connection with psychology and sociology. She also added that she would like to improve her spelling and she would like to use more linkers and fillers in her writing. Moreover, she added that she would like to use conditional structures correctly. Avarka said that she would improve her vocabulary by reading and listening. She also revealed that her sister sometimes helps because she checks her essays at home. Avarka also said that she finds writing more difficult than speaking, listening and reading.

8.6.2 Self-Regulatory Processes

Figure 8.26 shows the evolution of Avarka's self-regulatory processes. Each value represents the number of themes identified in the interview data for each month. Figure 8.26 shows that Avarka focused on self-observation processes evenly over the investigated period. Conversely, an increase can be observed in the self-evaluation processes from data point 3. Figure 8.26 also shows that Avarka set one goal at data point 21.

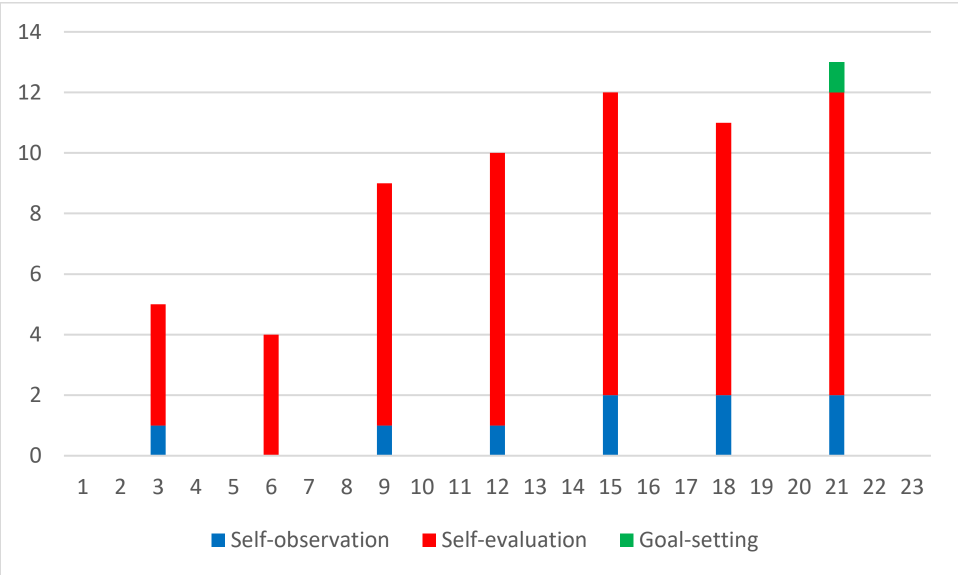


Figure 8.26 Self-regulatory processes

Figure 8.26 also shows that Avarka's focus shifted from self-observation to self-evaluation. For example, ten thematic units belonged to the self-evaluation processes category at data point 12, while only two thematic units belonged to self-observation processes category. The difference between the two types of self-regulatory processes remained constant between data points 12 and 21. Avarka set one goal over the nine months at data point 21.

Table 8.9 provides a breakdown of the range of Avarka's self-regulatory processes over the nine months. The number of self-observation thematic units was evenly distributed (ranging between 0 and 2) over time. However, the types of self-observation thematic units was not evenly distributed since Avarka mainly focused on her language and composing processes. The number of self-evaluation thematic units was not evenly distributed (ranging between 4 and 10) over time. The number of self-evaluation thematic units dramatically increased from data point 9. The types of self-evaluation thematic units were also not evenly distributed since Avarka mainly focused on her language. Table 8.9 also shows that Avarka set one goal at data point 21.

Table 8.9

Self-Regulatory Processes

		Data point							
		3	6	9	12	15	18	21	Total
Self-observation	Language	1	0	0	1	0	1	1	
	Composing processes	0	0	1	0	2	1	1	
	Content	0	0	0	0	0	0	0	
	Quality	0	0	0	0	0	0	0	
	Total	1	0	1	1	2	2	2	9
Self-evaluation	Language	2	3	4	4	5	5	6	
	Composing processes	1	1	1	2	1	2	1	
	Content	1	0	2	1	2	1	2	
	Quality	0	0	1	2	2	1	1	
	Total	4	4	8	9	10	9	10	54
Goal-setting	Language	0	0	0	0	0	0	1	
	Composing processes	0	0	0	0	0	0	0	
	Content	0	0	0	0	0	0	0	
	Quality	0	0	0	0	0	0	0	
	Total	0	0	0	0	0	0	1	1

The following examples (Interview excerpt 8.1-8.3) demonstrate that Avarka's focus shifted from self-observation to self-evaluation over the nine-month investigation. Interview excerpt 8.1 illustrates Avarka's impressionistic description of her action containing somewhat superficial description of her own writing at data point 3.

Interview excerpt 8.1 (Self-observation: language; data point 3)

Interviewer: How about the sentence constructions?
Avarka: I usually make the sentences too long. I add extra things to it.

However, Interview excerpt 8.2 and 8.3 contain evaluative comments that involved several cognitive processes such as reasoning, specifying and analysing. In addition, Avarka's specificity of her descriptions on L2 writing also shifted from a lack of tangibility to more elaborate and detailed explanations.

Interview excerpt 8.2 (Self-evaluation: language; data point 12)

Interviewer: Any grammatical structures?
Avarka: 'if sentences'… I think that's the only grammatical part that I am worried about because I don't really write very complicated sentences… it's risky if you use a sentence like this because it can decrease my, you know… if I write wrong… it shows that I can't use it … so I'm worried about using it.

Interview excerpt 8.3 (Self-evaluation: language; data point 18)

Interviewer: Were you thinking hard to find synonyms?
Avarka: I write 'regardless of the level' and it doesn't sound good. I had couple of sentences like this. They don't really sound natural. I usually don't write very bad sentences if I have a better idea but it's very short so I didn't have any ideas.

Interview excerpts 8.4-8.7 demonstrate that Avarka's focus also shifted from self-observation to self-evaluation and from self-evaluation to goal-setting. Interview excerpt 8.4 contains Avarka's impressionistic description of her own writing. Interview excerpt 8.4 does not include any specification, analysis, or reasoning.

Interview excerpt 8.4 (Self-observation; data point 9)

Avarka: I think how I write is how I speak.

However, Interview excerpt 8.5 and 8.6 contain evaluative comments that involved several cognitive processes such as reasoning and specifying.

Interview excerpt 8.5 (Self-evaluation: language; data point 21)

Avarka:	I was thinking about … to write less, like I would speak… because my sister used to tell me that she thinks that I write as I… speak. So it's really… so you write differently.

Interview excerpt 8.6 (Self-evaluation: language; data point 21)

Interviewer:	Are you happy with the level of vocabulary?
Avarka:	Not really. After I read the first version (Essay 3)… I think this one is better (Essay 3).
Interviewer:	Why do you think that?
Avarka:	I don't think that everything is correct in this… there is grammatical problems in this one (Essay 3)…. but I think the sentences and words are more complex (in Essay 3)… than this one (Essay 21)… I think I was trying harder… at the first time… but it was harder to put this whole text together, so I corrected… a lot more things than… this time (data point 21)… so it structures better (Essay 21) than this one (Essay 3) but I think this the old one (Essay 3) have more ideas and better sentences.
Interviewer:	Why do you think that?
Avarka:	Because now I was concentrating more on writing things correctly and less complicated… maybe that's better way to do this…

Interview excerpt 8.7 show that Avarka set a goal at data point 21. Her goal concerned the structuring of Essay 21 and accuracy. Avarka's goal also coincided with the developmental peak detected in the accuracy (EFC/C) index. Although the developmental peak was detected slightly earlier (at data point 16), it is quite likely that the developmental peak was triggered by a change in her motivational system as illustrated by Interview excerpt 8.7.

Interview excerpt 8.7 (Goal-setting: language; data point 21)

Interviewer:	Compared to the previous six pieces of writing… do you think this repetition was an improvement?
Avarka:	No… I think compared to the last one, so the one we wrote last week… it was an improvement… because I didn't have many ideas last week…. but compared to the ones, I think, it's OK but not the best… I learned things, and I remember things that I didn't when I came here, but… I realised that I have some mistakes or… I know things, some things incorrectly, I question more… I think wasn't… so afraid to use… things on the first week as much as I was now… I was trying really hard
Interviewer:	Why didn't you try really hard at the last essay?
Avarka:	because… I want to be long enough and… well-structured… and correct… so these were more important.

In Avarka's interview data no attractor states were identified. However, there was a strong indication of co-adaption of her linguistic and motivational systems.

8.6.3 Co-Adaptation

There is a strong indication of co-adaptation of linguistic and motivational systems in Avarka's data. Although her motivational system did not reside in an attractor state based on her interview data, it might be speculated that the lack of feedback caused system stability in her linguistic system. At data point 21, Avarka set a goal (see Interview excerpt 8.7) triggered by feedback arriving from her milieu (sister's feedback). Instead of concentrating on her lexical and syntactic systems, Avarka focused on her accuracy and mode. A change in her motivational system might have caused a change in her linguistic system, as expressed by a developmental peak detected in her accuracy (EFC/C) index at around data point 19.

8.7 Summary

All three constructs of linguistic complexity exhibited nonlinear developmental trends in Avarka's written data over the nine-month period. No developmental peaks were detected in the lexical and syntactic systems. Nonetheless, in the accuracy index (EFC/C) a developmental peak was detected indicating an improvement in accuracy. The growth trajectories along with the residual plots, and the moving correlation plots indicated that the interactions between lexical and syntactic complexity and accuracy were dynamic. Both the magnitude and the polarity of these relationships fluctuated over the nine months. Furthermore, correlation coefficients were also calculated to explore the surface interactions between the constructs. The model confirmed the hypothesised competitive association between lexical variability (TTR) and subordination (DC/C). However, the moderately competitive relationship between lexical variability (TTR) and accuracy (EFC/C) and the moderately supportive relationship between subordination (DC/C) and accuracy (EFC/C) were not confirmed. The evolution of Avarka's self-regulatory processes was also investigated and it was found that her focus shifted from self-observation to self-evaluation over the course of the investigation. One goal was also identified in Avarka's interview data indicating an improvement in her self-regulatory processes and consequently her L2 writing.

Chapter 9. Cross-Case Analysis

Chapter 9 compares the growth trajectories of the lexical and syntactic complexity and accuracy indices among the four participants' written data. There are three main sections: (1) lexical complexity, (2) syntactic complexity, and (3) accuracy.

9.1 Lexical Complexity

Figure 9.1 shows the growth trajectories of the general lexical complexity (AveWL) index in the four participants' written data. The four AveWL trajectories slightly oscillate between 4.00 and 5.00 characters between data points 1 and 15. However, the trajectory of the general lexical complexity (AveWL) index in Dalma's written data shows a peak between data points 16 and 17. Although smoothing algorithms are not added to the growth trajectories, upward trends can be observed in the four participants' data. Although the four trajectories fluctuate between 4.00 and 5.00 characters, Figure 9.1 also shows that the trajectory of the AveWL index, especially between data points 3 and 6, is slightly lower in Levente's data than the AWL trajectories in the other three learners' data.

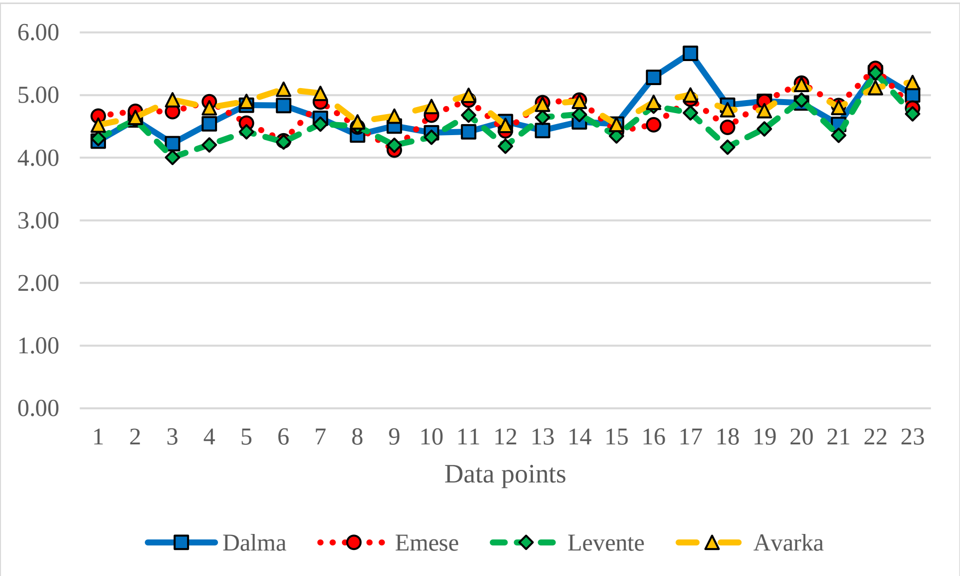


Figure 9.1 General lexical complexity (AveWL)

The trajectories of the lexical variability (TTR) index are shown in Figure 9.2. Although smoothing algorithms are not added to Figure 9.2, it is clearly visible that trajectories of lexical variability (TTR) display upward trends in Dalma and Emese's written data, while downward trends in Levente and Avarka's written data. Figure 9.2 also shows that the trajectory of the TTR index oscillates to a greater extent in Dalma's written data compared to the other three TTR trajectories in the other three participants' written data. Furthermore, the trajectory of the TTR index was relatively lower (below 0.70 TTR value) between data points 1 and 15 in Dalma's data than the TTR trajectories in the other three participants' data. Conversely, the TTR trajectory was relatively higher (above 0.70 TTR value) in Dalma's data between data points 16 and 23 than the TTR trajectories in the other three learners' data.

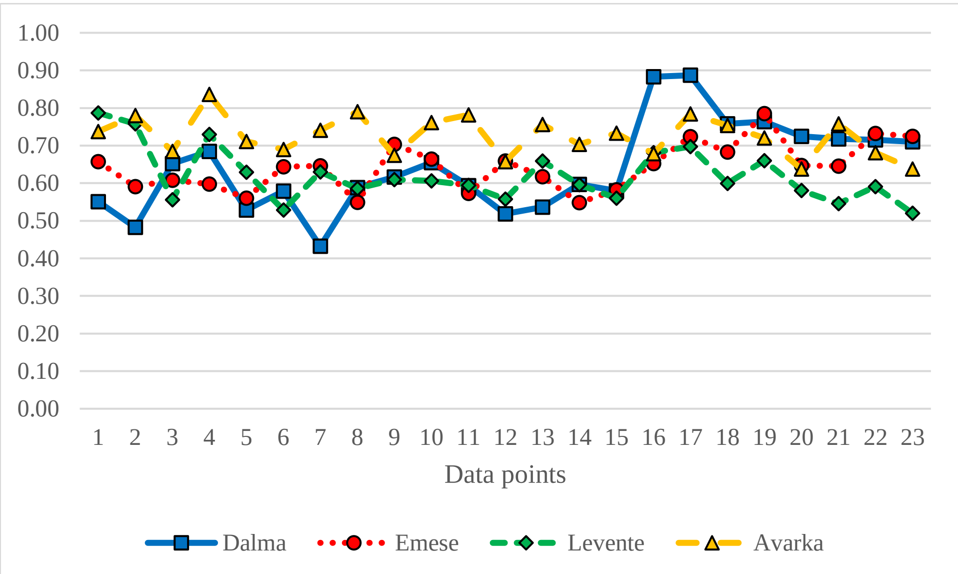


Figure 9.2 Lexical variability

Figure 9.2 also shows that the trajectory of the TTR index is relatively higher between data points 1 and 15 in Avarka's data than the TTR trajectories in the other three leaners' data which indicates that lexical items were varied to a greater extent in Avarka's data over the entire period investigation. For instance, the TTR trajectory never goes below 0.60 in Avarka's data over the nine-month period. Furthermore, TTR trajectories oscillate to a lesser extent in Emese and Avarka's data.

Figure 9.3 displays the trajectories of the lexical rarity (WRDFRQc) index in the four participants' written data. Although smoothing algorithms are not added to Figure 9.3, a downward trend can be observed in the WRDFRQc trajectory in Dalma's data. Figure 9.3 also shows that the trajectories of the lexical rarity (WRDFRQc) index vary to a similar extent in Dalma, Emese and Levente's data. Conversely, the WRDFRQc trajectory is relatively lower (below 1.20 WRDFRQc value) in Avarka's data than the WRDFRQc trajectories in the other three learners' data between data points 10 and 12 and at data point 14. In other words, Avarka used less frequent lexical items than the other three participants between data points 8 and 14.

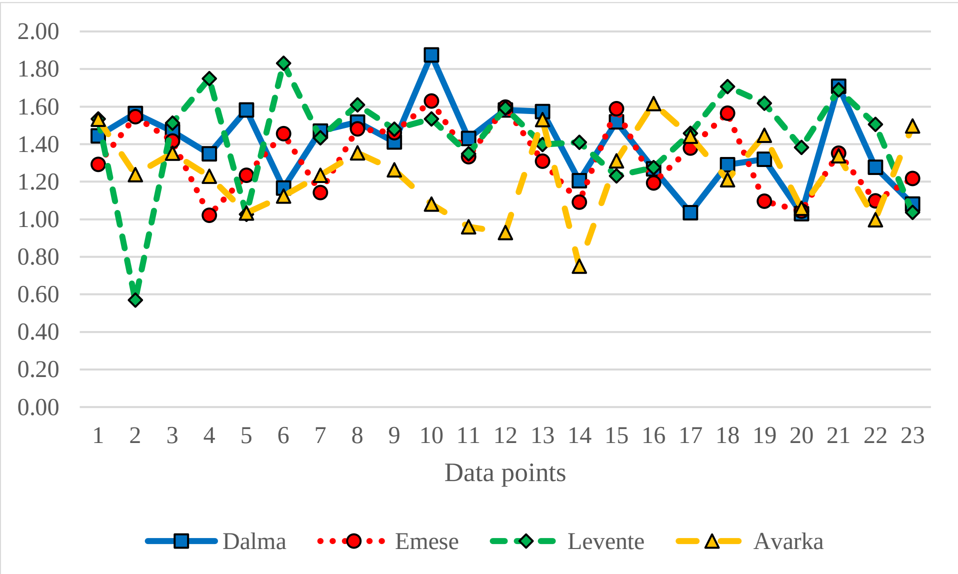


Figure 9.3 Lexical rarity (WRDFRQc)

Figure 9.4 displays the trajectories of the lexical disparity (LSA) index in the four learners' written data. Figure 9.4 clearly shows that the LSA indices oscillate to a greater extent than the trajectories of the AveWL, the TTR and the WRDFRQc indices. Due to the wild fluctuations, trends cannot be observed without the use of smoothing algorithms. Figure 9.4 also shows that the local peaks and dips overlap among the four participants between data points

19 and 21. The higher LSA values at data points 5, 6, and 7 mean that each sentence is conceptually more similar to every other sentence in Dalma's Essay 5, 6, and 7 than in the other three participants' corresponding essays.

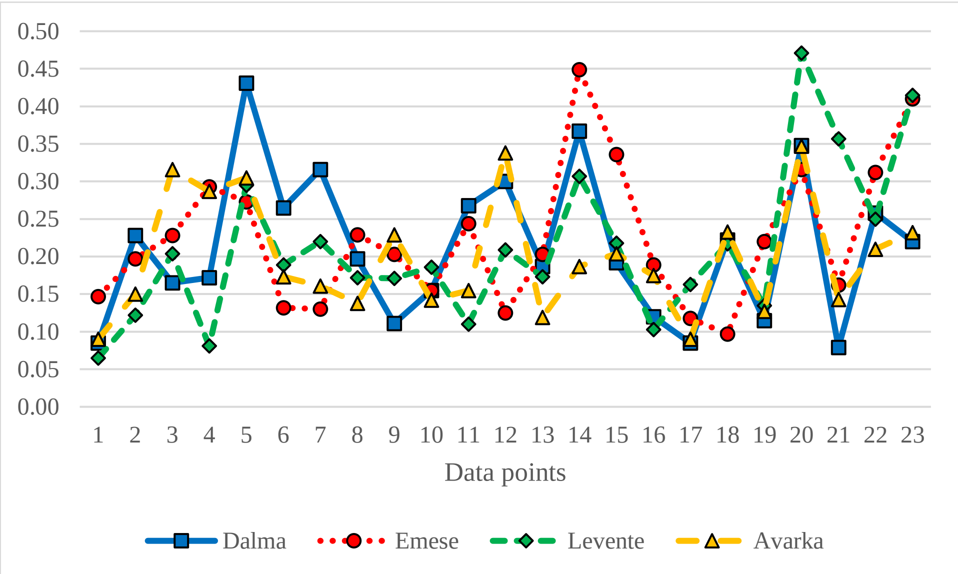


Figure 9.4 Lexical disparity (LSA)

The trajectories of the Academic Word List (AWL) index, displayed in Figure 9.5, show several similarities. First, the four trajectories of the AWL index show peaks at data points 7, 17, and 22. Conversely, the four AWL trajectories display dips at data points 6, 18, and 21. The correspondence between the four trajectories might be attributed to topic effects. In this study the four participants composed essays on 'language learning' at data point 3, 6, 9, 12, 15, 18, and 21, while the topics were not controlled at the remaining data points. Furthermore, the four participants' AWL trajectories are relatively lower (below 6%) at controlled data points than at uncontrolled data points. For instance, the trajectories of the AWL index in the four participants' data exceed 6% at data point 7 and 17 when the topic was 'work' and 'science and technology' respectively. This finding suggests that the topic affected word choice, especially words from the Academic Word List, in this study. Figure 9.5 also shows that the AWL trajectory was relatively higher (above 4%) in Avarka's data than the corresponding trajectories in the other three learners' data between data points 1 and 14 (except for data point 3, 6, 9 and 12).

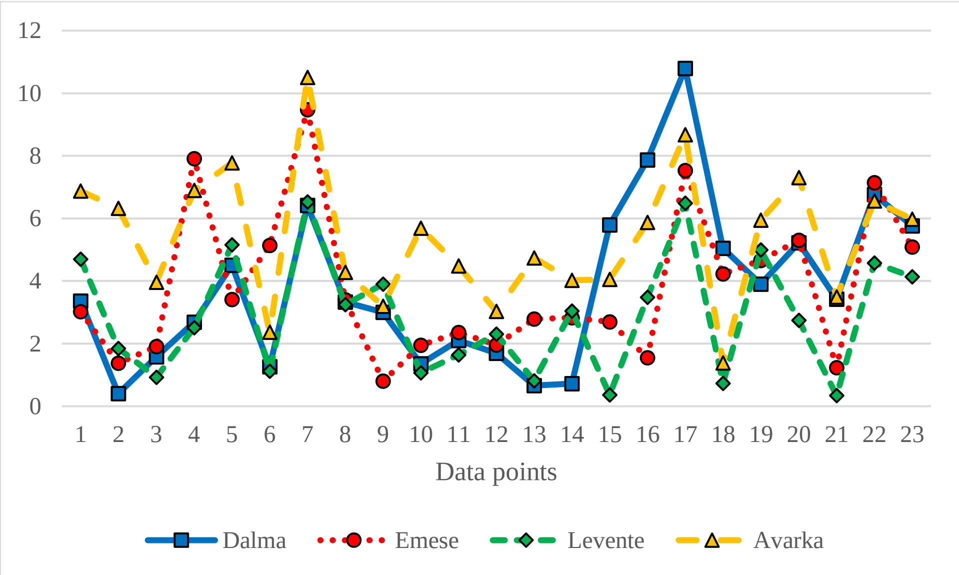


Figure 9.5 Academic Word List

9.2 Syntactic Complexity

Figure 9.6 shows the four trajectories of the general syntactic complexity (FVR) index. Although smoothing algorithms are not added to Figure 9.6, it is clearly visible that the FVR trajectories show upward trends in Dalma and Emese's written data. Figure 9.6 also shows that the FVR trajectory was relatively higher (above 8.00 FVR value) in Avarka's data than the corresponding trajectories in the other three participants' data. In other words, Avarka tended to use more words per finite verb in her essays than the other three learners from the beginning of the data collection. Conversely, the FVR trajectory was relatively lower (below 7.00 FVR values) in Dalma's data compared to the FVR trajectory in Avarka's data between data points 1 and 3. Furthermore, the FVR trajectory in Dalma's data displays the wildest fluctuations (ranging from 6.67 to 11.34 FVR values) among the four learners' trajectories.

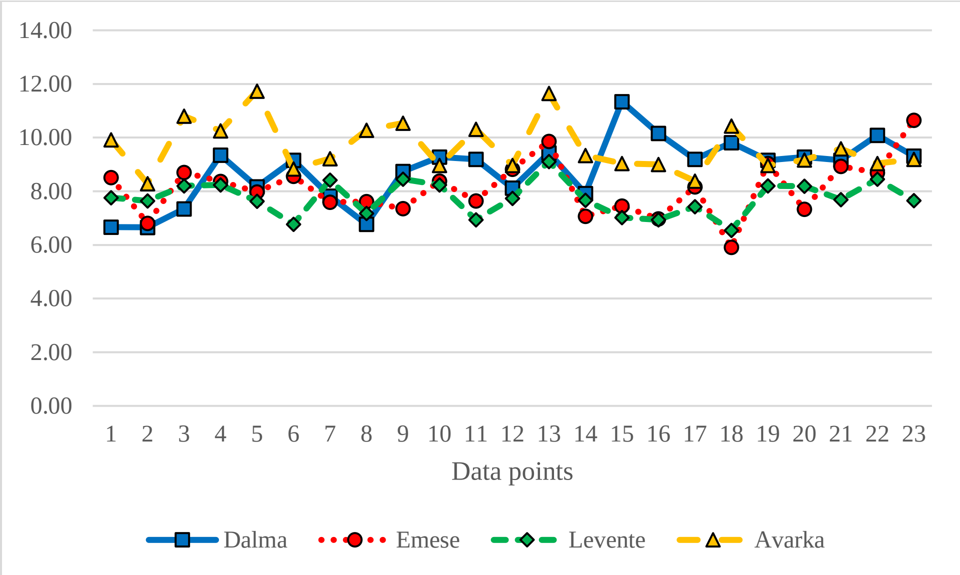


Figure 9.6 General syntactic complexity

Figure 9.7 shows the trajectories of the subordination (DC/C) index. Although smoothing algorithms are not added to the trajectories, it is clearly visible that the DC/C index increased in Emese, Levente, and Avarka's written data. Conversely, the DC/C index displays a downward trend in Dalma's written data which indicates that she tended to use fewer subordinate clauses in her essays over the nine months. Although the four trajectories mainly oscillate between 0.30 and 0.50 DC/C values between data points 1 and 14, the trajectory of the DC/C index displays lower than 0.30 DC/C values in Dalma's data. In other words, between data points 15 and 23, except for data points 18 and 22, every third clause was subordinate in Dalma's argumentative essays. In contrast, between data points 20 and 23, except for data point 22, every second clause was subordinate in Avarka's argumentative essays.

Figure 9.8 shows the trajectories of the clausal coordination (T/S) index. The four trajectories fluctuate between 1.00 and 1.60 T/S values. There are several similarities in the four learners' T/S trajectories. For example, there is a peak in the four T/S trajectory at data point 5, while local dips also correspond at data points 4 and 6.

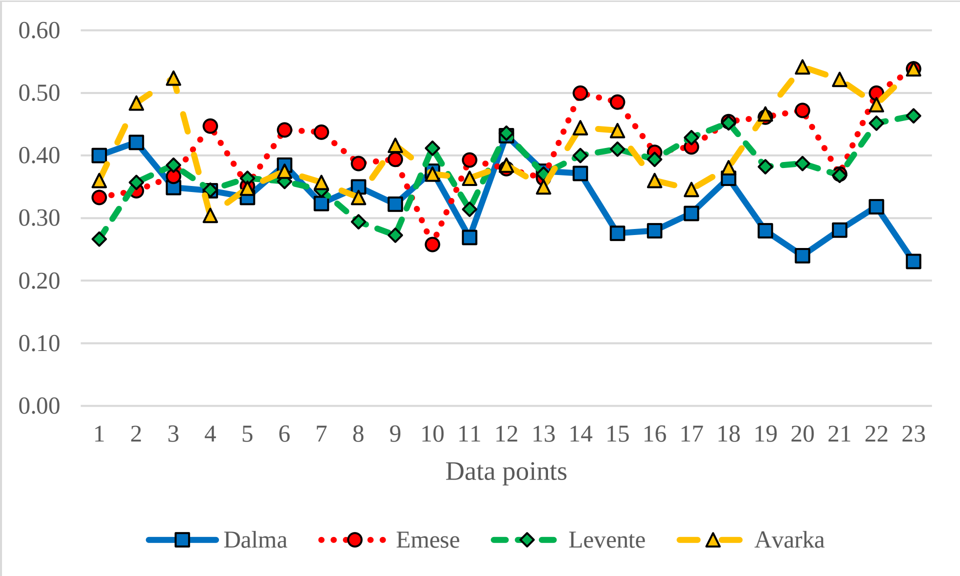


Figure 9.7 Subordination (DC/C)

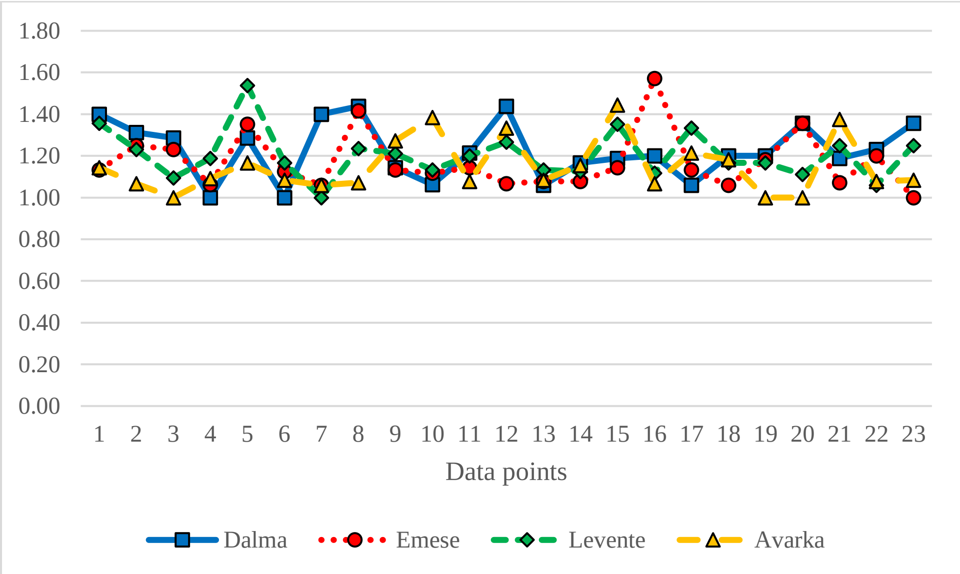


Figure 9.8 Clausal coordination (T/S)

Figure 9.9 shows the trajectories of the phrasal coordination (CP/C) index. Although smoothing algorithms are not added to the trajectories of the CP/C index, Figure 9.9 clearly shows upward trends in Dalma and Emese's written data. Figure 9.9 also shows that in Levente's data the CP/C values are lower than in the other three learners' data. The difference between the CP/C trajectory in Levente's data and the corresponding trajectories in the other three learners' data is clearly visible between data points 20 and 23. Figure 9.9 also shows that the CP/C trajectory in Dalma's data is above 0.20 between data points 16 and 23 (except for data point 21) which indicates that on average one coordinate phrase was used for every four clauses. In contrast, on average one coordinate phrase was used for every 10 clauses in Levente's argumentative essays. Figure 9.9 also shows a peak (0.58 CP/C value) in Avarka's data at data point 4.

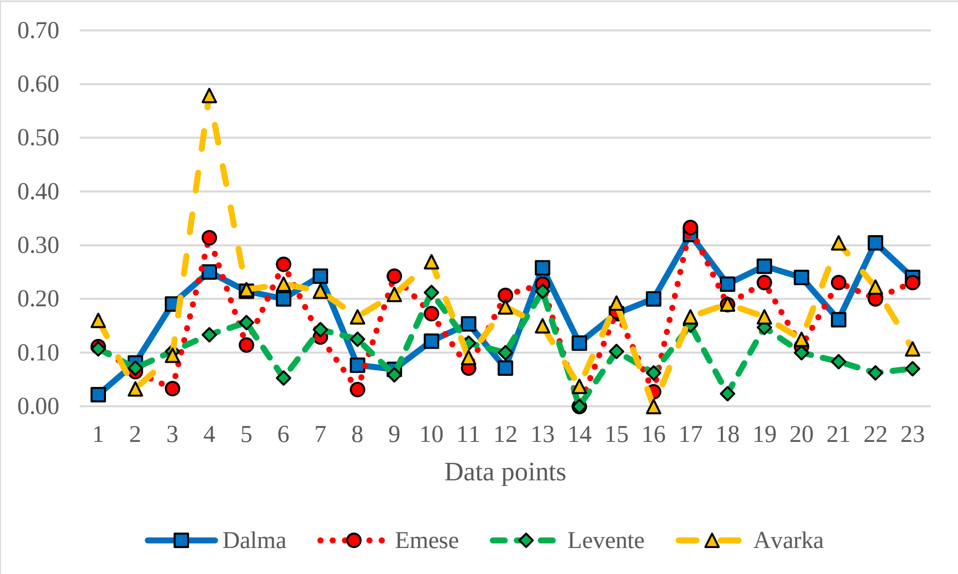


Figure 9.9 Phrasal coordination (CN/C)

Figure 9.10 shows the four trajectories of the phrasal complexity (CN/C) index. Although smoothing algorithms are not added to Figure 9.10, the trajectories of the CN/C index show clear upward trends in Dalma and Emese's written data. It is also important to note that the trajectories of the phrasal complexity (CN/C) index range between 0.40 and 1.14 CN/C values between data points 1 and 14 in Dalma, Emese and Levente's written data. Conversely, the CN/C trajectory shows higher CN/C values (ranging between 1.16 and 1.58) in Avarka's data between data points 1 and 9. In other words, Avarka used more complex nominals than the other three learners from the beginning of the data collection.

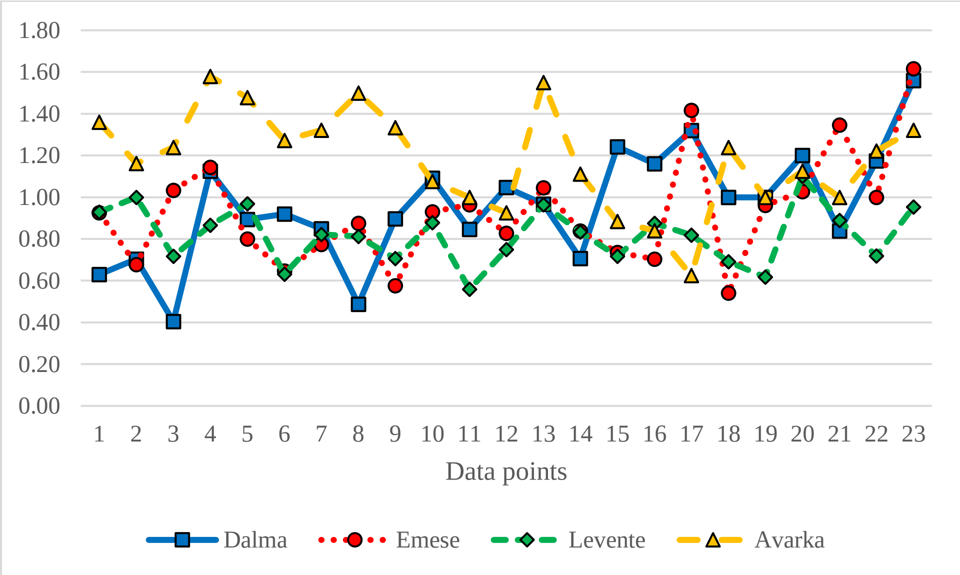


Figure 9.10 Phrasal complexity (CN/C)

9.3 Accuracy

Figure 9.11 shows the trajectories of the accuracy (EFC/C) index in the four participants' written data. The four trajectories of the EFC/C index range between 0.50 and 0.80 EFC/C values between data points 1 and 12.

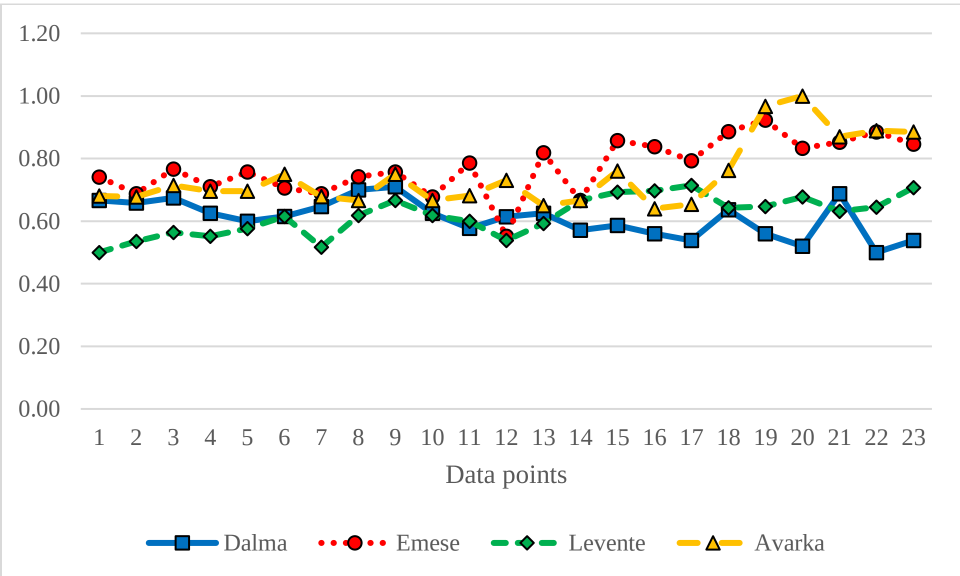


Figure 9.11 Accuracy (EFC/C)

However, from data point 13 the trajectories of the accuracy (EFC/C) index start to decrease in Dalma's data, while the EFC/C trajectory stagnates in Levente's data. Conversely, the trajectories of the accuracy (EFC/C) index start to increase in Emese's and Avarka's written data which suggests improvement in accuracy. The difference between the four trajectories is even clearer between data points 19 and 23. For instance, the trajectories show higher than 0.80 EFC/C values in Emese and Avarka's data from data point 19.

9.4 Summary

In Chapter 9 the lexical and syntactic complexity and accuracy indices were compared and contrasted among the four participants. Visual inspection of the complexity and accuracy trajectories shows that some indices were at different levels in the four learners' data. For example, the trajectories of the lexical variability (TTR) and the Academic Word List (AWL) indices were relatively higher in Avarka's data than the corresponding trajectories in the other three participants' data. Conversely, the trajectory of the lexical rarity (WRDFRQc) index was relatively lower in Avarka's data than the corresponding trajectories in the other three learners' data. As far as syntactic complexity is concerned, the trajectories of the syntactic complexity (FVR) and the phrasal complexity (CN/C) indices were relatively higher in Avarka's data than the corresponding trajectories in the other three learners' data.

Chapter 10. Summary and Discussion

Chapter 10 is a summary and discussion of the research findings of the four case studies (Dalma, Emese, Levente, and Avarka). I do not suggest that the research findings of this study are applicable beyond my own research context and data. Instead, this study makes "particular generalizations" (Gaddis, 2002, p. 62). Therefore, each case study is summarised and discussed in separate sections. Each section of Chapter 10 is divided into four sub-sections which answer the four research questions of this study respectively.

10.1 Dalma

Dalma was the first participant of this study, and she showed improvement, as indicated by the statistically significant developmental peaks, in some components of lexical complexity over the nine-month investigation. Although syntactic complexity and accuracy did not show improvements over time, both lexical and syntactic complexity indicated that Dalma's written data tended to display the characteristics of academic writing. The interview data showed that Dalma's self-regulatory processes did not evolve substantially over time. However, a shift between self-evaluation processes and goal-setting indicated a co-adaptation between Dalma's linguistic and motivational systems.

10.1.1 Developmental Trends

Lexical complexity indices developed nonlinearly in Dalma's written data over the nine-month investigation. The trajectories of the general lexical complexity (AveWL), the lexical variability (TTR), the lexical rarity (WRDFRQc), the lexical disparity (LSA) and the Academic Word List (AWL) indices showed ebbs and flows over time. The smoothing algorithms of the general lexical complexity (AveWL), the lexical variability (TTR) and the Academic Word List

(AWL) indices showed upward trends over time. However, the smoothing algorithms of TTR and the AWL indices showed inverse U-shaped curves between data points 15 and 20. The smoothing algorithm of the lexical rarity (WRDFRQc) index showed a downward trend, while the smoothing algorithm of the lexical disparity (LSA) index displayed a sideways trend over the nine-month investigation. However, the linear trend line of the LSA index showed a slightly downward trend in Dalma’s written data.

The upward trends in the general lexical complexity (AveWL) and the Academic Word List (AWL) indices, and the downward trend in the lexical rarity (WRDFRQc) index indicated that Dalma’s lexical complexity tended to display the characteristics of academic vocabulary over the nine-month investigation. Previous studies also found that the general lexical complexity (AveWL) and the Academic Word List (AWL) indices increased simultaneously in the written data of advanced learners over time (Penris & Verspoor, 2017; Verspoor et al., 2008, 2017). In addition, Penris and Verspoor (2017) found that the lexical rarity (WRDFRQc) index decreased in the written data of an advanced learner of English over 13 years. Likewise, Grant and Ginther (2000) found that as proficiency increases, L2 writers tend to use more unique and longer lexical items in writing. Schmid, Verspoor and MacWhinney (2011) also noted that academic words tend to be less frequent and longer. In addition, the upward trend in the lexical variability (TTR) index in Dalma’s written data indicated a more varied word use over time. Likewise, Verspoor et al. (2008) found an upward trend in the lexical variability (TTR) index in the written data of a Dutch advanced learner of English. Mazgutova and Kormos (2015) also found statistically significant increases in the lexical variability (MTLD) index in the data of an intermediate group. Likewise, Barkaoui (2016) found that the MTLD index increased over time. As far as word frequency is concerned, Bulté and Housen (2014) and Mazgutova and Kormos (2015) found overall decreases in word frequency over time. Although the smoothing algorithm of the lexical disparity (LSA) index showed a sideways trend, the linear trend indicated a slightly downward trend of the LSA index which suggested that Dalma's essays tended to become less cohesive. In other words, each sentence tended to become conceptually less similar to every other sentence in her texts. Mazgutova and Kormos (2015) also found that the LSA index stagnated in the upper-intermediate students' written data over one month. Likewise, Barkaoui (2016) did not detect statistically significant changes in the LSA index, except for the LSA overlap for adjacent paragraphs, over time. Overall increases in the Academic Word List (AWL) index were also found by studies adopting a two-wave research design (Knoch et al., 2015; Mazgutova & Kormos, 2015; Storch & Tapper, 2009).

Syntactic complexity indices developed nonlinearly in Dalma's written data over the nine-month investigation. The trajectories of the general syntactic complexity (FVR), subordination (DC/C), clausal coordination (T/S), phrasal coordination (CP/C), and the phrasal complexity (CN/C) indices showed fluctuations over time. The smoothing algorithms of the general syntactic complexity (FVR), phrasal coordination (CP/C) and phrasal complexity (CN/C and SYNNP) indices displayed upward trends, whilst the smoothing algorithm of the subordination (DC/C) index showed a downward trend. The smoothing algorithm of the clausal coordination (T/S) index showed a sideways trend. However, the linear trend of the T/S index indicated a downward trend.

The upward trends in the Academic Word List (AWL), the general syntactic complexity (FVR), the phrasal coordination (CP/C) and the phrasal complexity (CN/C and SYNPP) indices and the downward trend in the subordination (DC/C) index indicated that Dalma's written data tended to display the syntactic characteristics of academic writing over the nine months which corroborates previous findings in the field of corpus linguistics (Biber & Gray, 2010; Biber et al., 2011). The emergence of the features of academic writing in Dalma's written data can be explained by the following reason. The trajectories of the Academic Word List (AWL), the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices were at a lower value (e.g. percentage, ratio) in Dalma's written data than the corresponding trajectories in the other three learners' written data at the beginning of the data collection (Figure 9.5, Figure 9.6, and Figure 9.10). These differences between the trajectories might indicate that Dalma used the lexical and syntactic features of academic writing to a lesser extent than the other three participants at the beginning of the data collection. Furthermore, it might be speculated that Dalma was at a lower language proficiency level than the other three participants. According to the CDST, different subsystems of the language develop before others (Verspoor et al., 2012). For example, dependent clauses are more prevalent at B2 CEFR level and as the learners moves towards C1/C2 level more nominalisations might occur. Consequently, Dalma's data indicate that she moved from B2 to C1 CEFR level over the nine-month investigation. For example, the trajectory of the subordination (DC/C) index was at a higher value in Dalma's written data than the corresponding trajectories in the other three learners' written data at the beginning of the data collection (Figure 9.7). In other words, Dalma used more dependent clauses in her essays than the other three learners at the beginning of the data collection.

Previous studies found that the general syntactic complexity (FVR) index displayed upward trends in academic writing (Ma, 2012; Penris & Verspoor, 2017; Verspoor et al., 2008, 2017). In addition, Penris and Verspoor (2017) found simultaneous upward trends of the general syntactic complexity (FVR) and the average noun phrase length indices in the written data of an advanced Dutch learner of English. Mazgutova and Kormos (2015) found that the dependent clause per T-unit (DC/T) index decreased in the written data of an upper-intermediate group. Likewise, Hou et al. (2016) found statistically significant decreases in both the DC/C and the DC/T indices over time. Similarly, Knoch et al. (2015) found that the subordination (DC/C) index decreased over three years. As far as the clausal coordination (T/S) index is concerned, Hou et al. (2016) found that the T/S index stagnated over 18 months, while Vyatkina (2012) found that the coordinating conjunction (CC) index decreased in the written data of beginner learners of German. Overall increases in phrasal coordination (CP/C) are not uncommon in studies adopting a two-wave research design. For instance, Hou et al. (2016) found statistically significant increases in the coordinate phrase per clauses (CP/C) and the coordinate phrase per T-unit (CP/T) indices. Overall increases in phrasal complexity are also common in advanced learner's academic writing. For example, Bulté and Housen (2014) found that the mean length of noun phrase index increased over time. Likewise, Mazgutova and Kormos (2015) found statistically significant increases in the phrasal complexity (CN/C and SYNNP) index in the written data of an intermediate group. Upward trends in phrasal complexity were also found by studies adopting multi-wave research design. For example, Rosmawati (2016) found that the complex nominal index displayed an upward trend in a Korean learner's written data, whilst Penris and Verspoor (2017) found that the average noun phrase length increased in the written data of a Dutch learner of English over 13 years.

The proportion of simple and compound sentences displayed upward trends, while the proportion of complex and compound-complex sentences displayed downward trends in Dalma's written data. In other words, Dalma tended to use fewer complex or subordinate sentences in her writing. This finding is not surprising since the subordination (DC/C) index displayed a downward trend, while the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices showed upward trends in her data. In other words, Dalma tended to rely more on phrasal complexity than clausal complexity over time. Likewise, Penris and Verspoor (2017) found that the proportion of simple sentences increased in the written data of an advanced Dutch learner of English. Conversely, Verspoor et al. (2012) found that the proportion of simple sentences decreased, while the proportion of complex sentences increased across proficiency levels. However, Verspoor et al. (2012) investigated the development of sentence types from A1.1 to B1.2 CEFR levels.

The genre-specific syntactic complexity indices developed nonlinearly in Dalma's written data over the nine-month investigation. The trajectories of the normed rate of occurrence of the conditional clauses (ConC), the normed rate of occurrence of the infinitive clauses (InfC), the normed rate of occurrence of the relative clauses (RelC), and the incidence score of the prepositional phrases (DRPP) showed fluctuations over time. The smoothing algorithms of the ConC index showed a downward trend, while the InfC and the RelC indices displayed sideways trends over the nine-month investigation. However, the linear trend lines of the InfC and the RelC indices displayed downward trend over time. In other words, Dalma tended to use fewer conditional clauses, infinitive clauses and relative clauses over the nine-month investigation. Conversely, Mazgutova and Kormos (2015) found an increase in the usage of conditional clauses in both the intermediate and the upper-intermediate groups. The reduction in the usage of infinitive clauses suggested that Dalma tended to move in the direction of relying more on phrasal complexity in her writing than on post-modification. Likewise, Mazgutova and Kormos (2015) found statistically significant decreases in the use of infinitive clauses in the written data of an upper-intermediate group. The downward trend in the RelC index in Dalma's written data indicated that she tended to become less dependent on post-modification by relative clauses. Mazgutova and Kormos (2015) found mixed results in the usage of relative clauses. The frequency of the relative clauses decreased in the written data of the upper-intermediate group, while the frequency of the RelC index increased in the written data of the intermediate group. The smoothing algorithm of the prepositional phrases (DRPP) displayed an upward trend over time indicating that Dalma tended to use more prepositional phrases in her essays. Conversely, Mazgutova and Kormos (2015) found decreases in the usage of prepositional phrases in the written data of both groups.

The accuracy (EFC/C) index developed nonlinearly in Dalma's written data over the nine-month investigation. The trajectory of the accuracy (EFC/C) index displayed oscillations over time. The smoothing algorithm of the accuracy (EFC/C) index displayed a downward trend over time which suggested that Dalma started to make more linguistic errors over the nine-month investigation. Likewise, Storch (2009) found that the grammatical accuracy did not change significantly in the written data of 25 university students over 12 weeks. Although the ratio of errors per words increased, the ratio of error-free clause per clauses index (EFC/C) decreased over time in Storch's (2009) study. Storch and Tapper (2009) and Knoch et al. (2015) also found that the accuracy (EFC/C) index did not change significantly over time. As far as the error-types are concerned, formal errors (morphology and spelling) dominated Dalma's written data. Likewise, Verspoor et al. (2012) found that spelling errors were the most frequent type of errors at A1.1, A2, B1.1 and B1.2 CEFR levels.

10.1.2 Variability

Lexical and syntactic complexity and accuracy indices showed a great deal of variability in Dalma's written data over the nine-month investigation. Furthermore, the amount of variability constantly changed in the lexical and syntactic complexity and accuracy indices over time. First, the degree of variability in complexity and accuracy in Dalma's written data was visualised by min-max graphs. Second, Dalma's written data were tested for significance by data resampling and running Monte Carlo analyses. The lexical variability (TTR) and the Academic Word List (AWL) indices showed developmental peaks which proved to be statistically significant in the Monte Carlo analyses. However, there were no significant developmental peaks in the other complexity and accuracy indices. The developmental peaks in lexical variability (TTR) and Academic Word List (AWL) indicated that Dalma showed improvements in these two areas of lexical complexity. In other words, Dalma used a more varied vocabulary from data point 16 and used more words from the Academic Word List from data point 16.

Developmental peaks in advanced learners' data are uncommon in studies adopting a CDST perspective. For example, Rosmawati (2016) did not detect developmental peaks in her longitudinal study on academic writing development. However, Verspoor et al. (2008) found developmental peaks in the development of negative construction in a 13-year old Spanish learner of English, originally reported by Cancino, Rosansky, and Schumann (1978). Verspoor et al. (2018) also found developmental peaks (near significant) in the mean length of T-unit (MLTU) and the Guiraud's index in one Dutch learner's written data and in the MLTU index in another Dutch learner's written data. However, no developmental peaks were found in the written data of the other participants in Verspoor et al.'s (2018) study.

10.1.3 Interactions

The interactions between lexical and syntactic complexity and accuracy were dynamic over the nine-month investigation. The polarity of the interactions between lexical and syntactic complexity and accuracy changed from negative (competitive) to positive (supportive) and vice versa over time. Furthermore, the magnitude of the interactions fluctuated over time, ranging from weak to strong associations.

The surface interactions between the general lexical and syntactic complexity and accuracy showed statistically significant negative correlations between the general lexical complexity (AveWL) and the accuracy (EFC/C) and between the general syntactic complexity (FVR) and the accuracy (EFC/C) indices which indicated the asynchronous development of complexity and accuracy. Alternating variability patterns, as demonstrated by the residual and the moving window of correlation plots, might stem from dynamic precursor relations (Caspi, 2010; Verspoor et al., 2008).

The surface interactions between the specific lexical and syntactic complexity and accuracy showed statistically significant negative correlations between the lexical variability

(TTR) and the subordination (DC/C) which indicated that Dalma tended to use more varied vocabulary in her essays and at the same time she tended to become less dependent on clausal complexity. Furthermore, a statistically significant negative correlation was found between the lexical variability (TTR) and the accuracy (EFC/C) indices which suggested that lexical complexity and accuracy did not develop synchronously. Indeed, the trajectory of the TTR index showed an upward trend, whilst the trajectory of the EFC/C index displayed a downward trend. Moreover, a statistically significant positive correlation was found between the subordination (DC/C) and the accuracy (EFC/C) which indicated that Dalma became less dependent on clausal complexity but made more errors in her essays as expressed in the EFC/C index.

The hypothesised interactions between lexical variability (TTR), subordination (DC/C) and accuracy (EFC/C) in Dalma's written data were tested by mathematical modelling. The data simulation procedures demonstrated that growth is an iterative process. In other words, the current level of development is determined by the preceding level of development (van Geert, 1994). The stochastic models confirmed two of the three hypothesised interactions in Dalma's written data. The hypothesised moderately negative associations between lexical variability (TTR) and subordination (DC/C) and between lexical variability (TTR) and accuracy (EFC/C) were confirmed by the model. However, the hypothesised moderately supportive association between subordination (DC/C) and accuracy (EFC/C) was not confirmed by the model, which indicated a competitive relationship between subordination (DC/C) and accuracy (EFC/C).

10.1.4 Self-Regulatory Processes

Dalma's self-regulatory processes developed nonlinearly over the nine-month investigation. The development of self-regulatory processes showed progression and regression over time. Such phenomena substantiate the findings of previous longitudinal studies on the evolution of self-regulatory processes (Nitta & Baba, 2015, 2018; Sasaki et al., 2018). The interview data demonstrated that Dalma's focus constantly shifted from the performance phase (self-observation) to the self-reflection phase (self-evaluation) back and forth. However, Dalma's focus shifted from the self-reflection phase to the forethought (goal-setting) phase only once over the nine-month investigation at data point 12. In addition, Dalma's interview data showed that her focus shifted from one phase to the other phase on a monthly basis which confirms Nitta and Baba's (2015) claim that self-regulatory processes can be extended to longer periods of classroom learning.

The limited number of shifts between the three phases of self-regulation prompted further analysis of Dalma's interview data. It was found that an attractor state (anxiety), stemming from her instructional setting, dominated her L2 learning experience because between data points 1 and 12 she was writing her dissertation. This supports Waninge (2015) who found that anxiety was among the four attractor states that made up the participants L2 experience. In addition, Dalma's anxiety was amplified by a number of external factors. She reported that she did not have time to prepare for the first vocabulary test in January 2015 and she did not receive sufficient feedback from her teachers during the EAP course. Consequently, the combination of an attractor state (anxiety) and external factors (time to learn and lack of feedback) created a salient attractor state in Dalma's motivational system that caused system stability in her linguistic system between data points 1 and 15.

Dalma's interview data suggested that the linguistic and motivational systems are co-adaptive. At data point 12, a shift from self-reflection to forethought phase can be observed in the evolution of her self-regulatory processes since at this point, she set a goal for the first time during the investigation. As outlined in Chapter 4 (section 4.5), the four participants were required to complete vocabulary test starting from January 2015. After poor results on previous vocabulary tests during the EAP course, she decided to learn new lexical items. Furthermore, there were important changes in Dalma's external factors at the same time. She submitted her dissertation and passed her state examination. After these two events, Dalma's level of anxiety was reduced, and she also had more time to study for the second vocabulary test in February 2015. Furthermore, a developmental peak was detected in her lexical system at around data point 16 indicating an improvement in her lexicon. In dynamic parlance, the constant perturbations (vocabulary test) and changes in her external factors (time to learn) might have led to the weakening of the attractor state that Dalma's motivational system settled into. The change in her motivational system (goal-setting) might have caused a change in her linguistic system (lexicon) suggesting that the linguistic and motivational systems are co-adaptive.

These changes in the motivational systems also support Csizér et al.'s (2010) Model of the Nested Systems in Motivation (MNSM). According to the MNSM, influences stemming from the learners' instructional settings might affect learning goals. In Dalma's data the perturbations originating from her instructional setting resulted in goal-setting. Furthermore, Dalma's data suggested that the MNSM could be extended with the linguistic system as shown in Figure 10.1. A change, as demonstrated by the statistically significant developmental peak, in Dalma's lexical subsystem, positioned in the linguistic system, might have been triggered by a change (goal-setting for the first time) in the self-regulatory subsystem, placed in the motivational system. The linguistic and the motivational systems are both positioned in the learner's system which is then placed in the instructional setting.

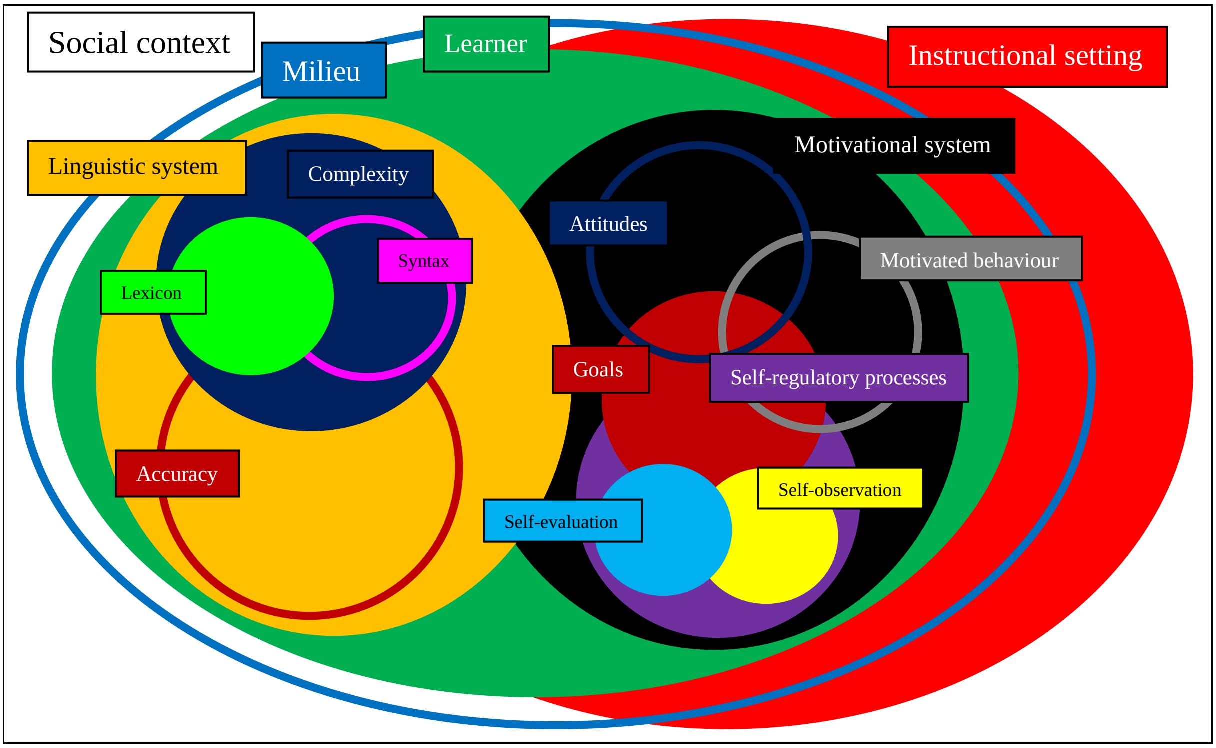


Figure 10.1 Co-adaptation in Dalma's data (activated systems and subsystems are filled)

10.2 Emese

Emese was the second participant of this study, and she did not show improvement, as demonstrated by the absence of developmental peaks, in lexical and syntactic complexity and accuracy indices over the nine-month investigation. Although the developmental peaks were not statistically significant, both lexical and syntactic complexity indices indicated that Emese's written data started to display the characteristics of academic writing. The interview data showed that Emese's self-regulatory processes did not evolve substantially over time. It was found that a salient attractor state – boredom – dominated her L2 learning experience over the nine-month investigation.

10.2.1 Developmental Trends

The smoothing algorithms of the general lexical complexity (AveWL) and the lexical variability (TTR) indices showed upward trends, while the smoothing algorithms of the lexical rarity (WRDFRQc), the lexical disparity (LSA) and the Academic Word List (AWL) indices displayed sideways trends over time. However, the linear trend lines of the WRDFRQc index showed a downward trend, while the LSA and the AWL indices showed upward trends over the nine months.

The upward trends in the general lexical complexity (AveWL) and the Academic Word List (AWL) indices and the downward trend in the lexical rarity (WRDFRQc) index indicated that Emese's lexical complexity started to demonstrate the features of academic vocabulary over the nine-month investigation. The developmental trends found in Emese's written data are identical to the developmental trends detected in Dalma's written data. The only difference is that the lexical disparity (LSA) index showed an upward trend in Emese's written data indicating that the sentences became more conceptually similar to every other sentence in her essays. In other words, Emese's essays became more cohesive over the nine-month investigation. Previous studies adopting a two-wave research design found that the LSA index increased over time. For example, Mazgutova and Kormos (2015) found statistically significant increases in the LSA index in the written data of the intermediate group. Likewise, Barkaoui (2016) found statistically significant increases in the LSA overlap for adjacent paragraphs index. The different directions in the developmental trends in the LSA index in Dalma's and Emese's written data confirmed the notion of idiosyncrasy, that is, there are no two learners who will go through the same developmental path despite the similarities in their learning context (de Bot & Larsen-Freeman, 2011; Lowie et al., 2009).

The smoothing algorithm of the general syntactic complexity (FVR) index displayed a sideways trend between data points 2 and 14 and then the smoother showed an upward trend. The smoothing algorithms of the subordination (DC/C) and the phrasal complexity (CN/C) indices displayed upward trends, whilst the clausal coordination (T/S) and the phrasal coordination (CP/C) indices showed sideways trend lines over time. The linear trend line of the

T/S index showed a downward trend, while the linear trend line of the CP/C index displayed an upward trend over the nine months.

The upward trends in the Academic Word List (AWL), the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices indicated that Emese's written data started to display the syntactic features of academic writing, which substantiates previous findings from the field of corpus linguistics (Biber & Gray, 2010; Biber et al., 2011). Nevertheless, the subordination (DC/C) index also demonstrated an upward trend in Emese's written data which suggested a simultaneous increase in both phrasal and clausal complexity. Likewise, Mazgutova and Kormos (2015) found simultaneous increases of the subordination (DC/T) and the phrasal complexity (CN) indices in the written data of an intermediate group. The different directions in the developmental trends in the subordination (DC/C) index in Dalma's and Emese's written data also confirmed the notion of idiosyncrasy. The stagnation of the clausal coordination (T/S) index is not surprising since clausal coordination is more prevalent in the written data of beginner learners (Bardovi-Harlig, 1992). However, the stagnation of the phrasal coordination index is unexpected since students at advanced proficiency levels tend to rely on phrasal instead of clausal complexity (Biber et al., 2011).

The proportion of simple and compound sentences displayed downward trends, whilst the proportion of complex sentences showed an upward trend over the nine-month investigation. The proportion of the compound-complex sentences showed a sideways trend over time. The trend lines of the sentence types suggested that Emese made her writing more complex by employing more complex (subordinate) sentences to the detriment of simple and compound sentences. These findings are in contrast with those of Dalma's written data which also confirmed the notion of idiosyncrasy.

The smoothing algorithm and the linear trend line of the ConC index showed a sideways trend over time which indicated that there were no changes in the usage of conditional clauses.

The smoothing algorithm of the infinitive clauses displayed a downward trend, as in Dalma's written data. The reduction in the infinitive clauses suggested that Emese tended to move in the direction of relying more on phrasal complexity in her essays than on pre- or post-modification. The smoothing algorithm of the relative clauses (RelC) showed an upward trend which indicated an increase in the usage of relative clauses over time. The smoothing algorithm of the prepositional phrases (DRPP) displayed an upward trend, as in Dalma's data, over time which also confirmed a shift towards phrasal complexity.

The accuracy (EFC/C) index developed nonlinearly in Emese's written data over the nine-month investigation. The trajectory of the accuracy (EFC/C) index showed oscillations over time. The smoothing algorithm of the accuracy (EFC/C) index displayed a sideways trend between data points 2 and 14 and then the smoother showed an upward trend. Upward trends were also identified by previous studies adopting a multi-wave research design. For example, Larsen-Freeman (2006) found that the EFT/T index increased over time. Likewise, Rosmawati (2016) found an overall increase over one academic year. The different directions of the accuracy (EFC/C) index also corroborate the notion of idiosyncrasy. As far as the error-types are concerned, formal errors (morphology and spelling) dominated Emese's written data, as in Dalma's written data. These findings support Verspoor et al. (2012) who also found that spelling errors were the most frequent type of errors.

10.2.2 Variability

Statistically significant developmental peaks were not detected in the lexical and syntactic complexity and accuracy indices which indicates that Emese did not show improvements in lexical and syntactic complexity and accuracy. The absence of developmental peaks in the written data of advanced learners is common. For example, Rosmawati (2016) did not find developmental peaks in her study on the academic writing development of advanced EFL learners.

10.2.3 Interactions

The surface interactions between the specific lexical and syntactic complexity and the accuracy indices showed a statistically significant positive correlation between the subordination (DC/C) and the accuracy (EFC/C) indices as in Dalma's written data. The hypothesised positive associations between the lexical variability (TTR) and the subordination (DC/C) indices and between the subordination (DC/C) and the accuracy (EFC/C) indices were confirmed by the stochastic models. However, the hypothesised supportive association between the lexical variability (TTR) and the accuracy (EFC/C) indices was not confirmed by the model which indicated a competitive relationship between lexical variability (TTR) and accuracy (EFC/C).

10.2.4 Self-Regulatory Processes

Emese's self-regulatory processes developed nonlinearly over the nine-month investigation. Such phenomena substantiate the findings of previous longitudinal studies on the evolution of self-regulatory, processes (Nitta & Baba, 2015; Sasaki et al., 2018). The interview data demonstrated that Emese's focus constantly shifted from the performance phase (self-observation) to the self-reflection phase (self-evaluation) back and forth. However, Emese's focus did not shift from the self-reflection phase to the forethought (goal-setting) phase over the nine-month investigation.

The limited number of shifts between self-observation and self-evaluation processes prompted further analysis of Emese's interview data. It was found that a salient attractor state (boredom) dominated her learning experience which aligns with Waninge (2015) who found that boredom was one of the four attractor states in students' L2 learning experience. Furthermore, Emese often expressed her negative attitude to writing during the retrospective interviews. Emese also reported that she did not receive sufficient feedback from her teachers during the EAP course. In dynamic parlance, a combination of attractor state (boredom), affect (negative attitude) and external factors (lack of feedback) caused system stability in her motivational system, hence in her linguistic system. Yashmia and Arano (2015) claimed that psychological components (cognition and affect) interact with context and might self-organise into attractor states that show stable behaviour. The perturbations, which triggered a phase shift in Dalma's lexical system, did not cause a phase transition in Emese's linguistic system which substantiates Henry (2015) who found that some perturbations did not lead to phase shifts.

Emese's attitude changed from negative to positive and then back to negative over the nine months. This result supports Csizér et al.'s (2010) claim that "attitudes and motivated behavior tend to fluctuate in the course of language learning" (p. 481) and Serafini (2017) who found that learners' attitude fluctuates over time at different time scales.

10.3 Levente

Levente was the third participant of this study, and he did not show improvement, as demonstrated by the absence of developmental peaks, in lexical and syntactic complexity and accuracy indices over the nine-month investigation. Furthermore, both lexical and syntactic complexity indices indicated that Levente's writing did not tend to display the features of academic writing over time. The interview data showed that Levente's self-regulatory processes did not evolve substantially over time. It was found that a salient attractor state – boredom – dominated his L2 learning experience over the nine-month investigation.

10.3.1 Developmental Trends

The smoothing algorithms of the general lexical complexity (AveWL) and the lexical disparity (LSA) indices showed upward trends over time (as in Emese's data) which indicated that Levente tended to use longer lexical items and each sentence tended to become conceptually more similar to every other sentence in his essays. The smoothing algorithm of the TTR index displayed a downward trend in Levente's written data which suggested that he tended to vary his vocabulary to a lesser extent over time. Likewise, Bulté and Housen (2014) found that the lexical diversity (D) index decreased over time. Although the smoothing algorithms of the lexical rarity (WRDFRQc) and the Academic Word List (AWL) displayed sideways trends, the linear trend lines of the WRDFRQc index displayed a slightly upward trend over time. However, the linear trend line of the AWL index shows an upward trend which indicates that there was a slight increase in the usage of academic words over time. The only similarities between the developmental trends detected in Dalma's, Emese's and Levente's written data were the upward trends in the general lexical complexity (AveWL) index which confirmed the notion of idiosyncrasy.

The smoothing algorithm of the subordination (DC/C) index displayed an upward trend which indicated that Levente tended to rely more on clausal complexity. The smoothing algorithms of the general syntactic complexity (FVR), the clausal coordination (T/S), the phrasal coordination (CP/C) and the phrasal complexity (CN/C) indices showed sideways trends over time. However, the linear trend lines of the FVR, the T/S and the CN/C indices showed downward trends over time. The upward trend in the DC/C and the downward linear trend line in the CN/C indices emphasised the notion of idiosyncrasy since the direction of the trends of these two indices (DC/C and CN/C) were different from Dalma and Emese's written data. The downward linear trend lines in the FVR and the CN/C indices confirmed the notion of idiosyncrasy since these two indices (FVR and CN/C) showed different directions in Dalma's and Emese’s written data. Furthermore, the linear trend lines of the FVR and the CN/C indices indicate that Levente’s written data did not start to show the features of academic writing.

The proportion of simple, compound and complex-compound sentences displayed downward trends, whilst the proportion of complex sentences showed an upward trend over the nine-month investigation. The trend lines of the sentence types suggested that Levente made his writing more complex by employing more complex (subordinate) sentences to the detriment of simple, compound and complex-compound sentences.

The smoothing algorithms of the normed rate of occurrence of the conditional clauses (ConC) and the normed rate of occurrence of the relative clauses (RelC) displayed sideways trends over time. The linear trend line of the RelC index showed an upward trend, while the linear trend line of the ConC index also showed a sideways trend. The smoothing algorithm of the InfC index displayed a downward trend, while the smoother of the prepositional phrases (DRPP) index displayed an upward trend. These results suggested that Levente tended to use fewer infinitive clauses, while more prepositional phrases and relative clauses in his essays over the nine months. However, in the usage of conditional clauses there were no changes.

The smoothing algorithm of the accuracy (EFC/C) index displayed an upward trend over time (as in Emese’s written data) which indicated that there were more error-free clauses in Levente’s essays. Similar to Dalma’s and Emese’s data, the most frequent error type belonged to the formal category (morphology and spelling).

10.3.2 Variability

Statistically significant developmental peaks were not detected in the lexical and syntactic complexity and accuracy indices which indicates that Levente did not show improvements in lexical and syntactic complexity and accuracy.

10.3.3 Interactions

A statistically significant positive correlation was found between the general lexical complexity (AveWL) and the accuracy (EFC) indices. Conversely, the AveWL-EFC/C interaction was negative in Dalma's written data. The positive correlation between the AveWL and the EFC/C indices suggested that lexical complexity and accuracy developed synchronously in Levente's written data over the nine months.

A statistically significant positive correlation was detected between the subordination (DC/C) and the accuracy (EFC/C) indices which indicated that Levente tended to rely more on clausal complexity and at the same time he made more error-free clauses in his essays. However, the model did not confirm the supportive relationship between the DC/C and the EFC/C indices. Likewise, the model did not confirm the competitive relationship between the TTR and the DC/C indices. However, the negative association between the TTR and the EFC/C indices was confirmed by the model.

10.3.4 Self-Regulatory Processes

Levente's self-regulatory processes developed nonlinearly over the nine-month investigation. The interview data demonstrated that Levente's focus constantly shifted from the performance phase (self-observation) to the self-reflection phase (self-evaluation) back and forth. However, his focus did not shift from the self-reflection phase to the forethought (goal-setting) phase over the nine-month investigation.

The restricted number of shifts between performance and self-reflection phases in the evolution of Levente's self-regulatory processes prompted further analysis of his interview data. It was found that an attractor state dominated his learning experience: boredom which supports Waninge (2015) who found that boredom was one of the four attractor states in students' L2 learning experience. Levente reported that he was frequently bored with the writing tasks during the EAP course. Furthermore, he also reported that the lack of feedback did not facilitate his development in writing. Therefore, it can be speculated that boredom formed a deeper attractor state in Levente's motivational system resulting in system stability in his linguistic system.

10.4 Avarka

Avarka was the fourth participant of this study, and she showed improvement, as indicated by the statistically significant developmental peak, in accuracy over the nine-month investigation. However, both lexical and syntactic complexity did not tend to display the features of academic writing over time. The interview data showed that Avarka's self-regulatory processes did not evolve substantially over time. However, a shift between self-evaluation processes and goal-setting indicated a co-adaptation between Avarka's linguistic and motivational systems.

10.4.1 Developmental Trends

The smoothing algorithm of the general lexical complexity (AveWL) showed an upward trend as in Dalma's, Emese's and Levente's written data. However, the smoothing algorithm of the lexical variability (TTR) index showed a downward trend in Avarka's data (as in Levente's data) which indicated that she tended to use less varied vocabulary in her essays. The smoothing algorithms of the lexical rarity (WRDFRQc), the lexical disparity (LSA) and the Academic Word List (AWL) indices showed sideways trends over the nine months. The linear trend line of the lexical rarity (WRDFRQc) index showed an upward trend over time which suggested that Avarka tended to use more frequent lexical items in her essays. The linear trend line of the lexical disparity (LSA) index also showed a sideways trend which implied that there were no changes in cohesion in Avarka's written data. The linear trend line of the AWL index showed a downward trend over the nine months which suggested that Avarka tended to use fewer words from the Academic Word List. The downward trend in the AWL index also confirmed the notion of idiosyncrasy since the AWL indices showed upward trends in Dalma's and Emese's data, whereas the AWL displayed a sideways trend in Levente's written data. Storch (2009) also found an overall decrease of the AWL index over time.

The smoothing algorithms of the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices displayed a downward trend, while the smoother of the subordination (DC/C) index showed an upward trend over the nine months which indicated that Avarka tended to rely on clausal complexity instead of phrasal complexity. The smoothing algorithms of the clausal (T/S) and phrasal coordination (CP/C) indices showed sideways trends over time. The linear trend lines of the T/S index showed an upward trend, while the linear trend line of the phrasal coordination (CP/C) index showed a downward trend over the nine months. The trend lines of the syntactic complexity indices suggested that Avarka's written data did not start to display the features of academic writing.

The proportion of simple and compound sentences displayed downward trends, whilst the proportion of complex sentences showed an upward trend over the nine-month investigation. Moreover, the proportion of the compound-complex sentences showed a sideways trend over time. The trend lines of the sentence types suggested that Avarka made her writing more complex by employing more complex (subordinate) sentences to the detriment of simple, compound and complex-compound sentences.

The smoothing algorithms of the infinitive clauses (InfC), relative clauses (RelC), and the prepositional phrases (DRPP) displayed downward trends over the nine months which suggested that Avarka tended to use fewer infinitive clauses, relative clauses, and prepositional phrases. The smoother of the conditional clauses (ConC) displayed a sideways trend over time.

However, the linear trend line of the ConC index showed an upward trend in Avarka's written data.

The smoothing algorithm of the accuracy (EFC/C) index displayed an upward trend over time (as in Emese's and Levente's written data) which indicated that there were more error-free clauses in Avarka's essays over time. Similar to Dalma's, Emese's and Levente's data, the most frequent error type belonged to the formal category (morphology and spelling).

The sideways trend in the Academic Word List (AWL) index and downward trends in the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices and the upward trends in the subordination (DC/C) index indicated that Avarka's written data did not tend to display the features of academic writing over the nine-month investigation. The lack of emergence of the characteristic of academic writing in Avarka's written data might be explained by the following. First, the trajectories of the Academic Word List (AWL), the general syntactic complexity (FVR) and the phrasal complexity (CN/C) indices were at a higher value at the beginning of the data collection in Avarka's written data than the corresponding trajectories in the other three participants' data (Figure 9.5, Figure 9.6, and Figure 9.10). These differences in the trajectories of the AWL, FVR and the CN/C indices might indicate that Avarka was already employing the lexical and syntactic features of academic writing at the beginning of the data collection. These differences between the trajectories of the AWL, FVR and the AWL indices might also indicate that Avarka was at a higher language proficiency level than the other three participants. Consequently, Avarka could not show as clear improvements in linguistic complexity as Dalma. Second, the lack of emergence of the features of academic writing might also be attributed to the type of writing task. Biber et al. (2011) claimed that phrasal complexity is more prevalent in academic texts than clausal subordination. However, in this study the four participants composed IELTS-type argumentative essays which may not have elicited the features of academic writing. Furthermore, the four participants' language proficiency level might have been at B2 CEFR level at the beginning of the EAP course since the minimum requirement was to possess a B2 CEFR level language certificate in English language to join the EAP course. Consequently, subordination was more prevalent than phrasal complexity in the four participants' argumentative essays.

10.4.2 Variability

The accuracy (EFC/C) index showed a developmental peak which proved to be statistically significant in the Monte Carlo analysis. However, there were no statistically significant developmental peaks in the lexical and syntactic complexity indices. The statistically significant developmental peak in the accuracy (EFC/C) index indicated that Avarka showed improvement in accuracy. In other words, there were significantly more error-free clauses in Avarka's essays from data point 19.

10.4.3 Interactions

The surface interactions between the specific lexical and syntactic complexity and accuracy showed statistically significant negative correlations between lexical variability (TTR) and subordination (DC/C) and between lexical variability (TTR) and accuracy (EFC/C) indices as in Dalma's written data. Conversely, in Avarka's written data, a statistically significant positive correlation was detected between subordination (DC/C) and accuracy (EFC/C) as in Dalma's written data. Interestingly, the polarity of the three interactions (TTR-DC/C, TTR-EFC/C, and DC/C-EFC/C) were the same in Avarka's and Dalma's written data.

The stochastic models confirmed only one of the three hypothesised interactions in Avarka's written data. The hypothesised negative association between lexical variability (TTR) and subordination (DC/C) was confirmed by the model. However, the hypothesised moderately competitive association between the lexical variability (TTR) and accuracy (EFC/C) and the moderately supportive association between subordination (DC/C) and accuracy (EFC/C) were not confirmed by the model. One of the modelling outcomes of the mathematical simulation in Avarka's data confirms the CDST tenet that sometimes there is chaotic variation in a dynamic system. The path of a dynamic system might be difficult to predict due to the interaction of variables over time and language learners might experience phases of fluency and disfluency for no apparent or discernible reason (de Bot & Larsen-Freeman, 2011). "Chaos refers simply to the period of complete randomness that complex nonlinear systems enter into irregularly and unpredictably" (Larsen-Freeman, 1997, p.143). This present study found that the model outcome in Avarka's data entered a phase of chaotic variation from data point 23. However, when the model of three connected growers was optimized separately (growth rates and relational parameters were optimized after one another) the outcome of the model did not display chaotic variation. Van Geert (1994) attributed chaotic oscillations to "deep intrinsic property of the processes themselves" and not to the consequence of external, independent factors (p. 84). In Avarka's data the trajectories of the three constructs lost their periodicity and moved into chaos. This finding is one of the most interesting results of this study since, to the best of my knowledge, no previous studies found chaotic oscillations in second language writing data.

10.4.4 Self-Regulatory Processes

The interview data demonstrated that Avarka's focus continuously shifted from the performance phase (self-observation) to the self-reflection phase (self-evaluation) back and forth. However, Avarka's focus shifted from the self-reflection phase to the forethought (goal-setting) phase only once over the nine-month investigation at data point 21. The three-phase cycle of self-regulatory processes supports Zimmerman's (2000) cyclical model of self-regulation. In addition, Avarka's interview data showed that her focus shifted from one phase to the other phase on a monthly basis which confirms Nitta and Baba's (2015) claim that self-regulatory processes can be extended to longer periods of classroom learning.

Although the limited number of shifts in self-regulatory processes prompted further analysis of Avarka's interview data, no attractor states were detected. In Avarka's interview data, there were no indications of boredom or anxiety as in Dalma's, Emese's and Levente's interview data. The system stability found in her linguistic and motivational systems between data points 1 and 21 appeared to be caused by the lack of feedback she reported.

There was a strong indication of co-adaptation between the linguistic and motivational systems in Avarka's data. At data point 21, in the evolution of her self-regulatory processes, a shift from self-reflection to forethought phase (goal-setting) could be observed. Although she did not receive sufficient feedback from her teachers, her sister continuously read and commented on the quality of her essays written at home. Avarka was motivated by her milieu (her sister's constant feedback) to focus on the genre and accuracy instead of lexicon and syntax. In addition, at data point 19, a statistically significant developmental peak was detected in the accuracy (EFC/C) index. In dynamic parlance, the perturbations (negative feedback) arriving from Avarka's milieu triggered a change in her motivational system (goal-setting) resulting in a change in her linguistic system (accuracy). These changes in her motivational system also substantiate the MNSM (Csizér et al., 2010) and call for an expansion of the model. Figure 10.2 shows the co-adaptation of the linguistic and motivational systems in Avarka's data. A change, as demonstrated by a statistically significant developmental peak, in Avarka's accuracy subsystem, positioned in the linguistic system, might have been triggered by a change (goal-setting for the first time) in the self-regulatory subsystem placed in the motivational system. The linguistic and the motivational systems are both positioned in the learner's system which then co-adapts with Avarka's milieu.

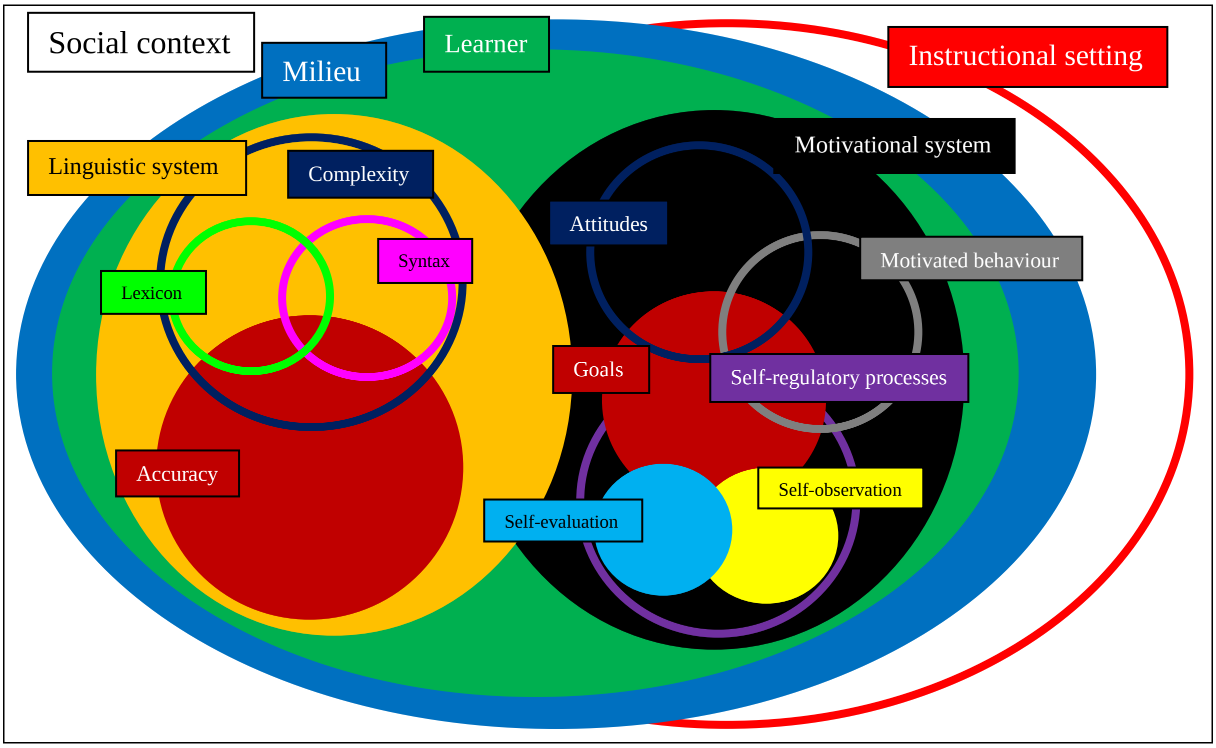


Figure 10.2 Co-adaptation in Avarka's data (based on Csizér, Kormos & Sarkadi, 2010)

10.5 Meta-Discussion

This section brings together the convergent, complementary, and divergent findings into a meta-discussion. Although there were some emerging tendencies and patterns in the four participants' data, I do not suggest any generalizations beyond my own research context and data.

Lexical and syntactic complexity and accuracy indices developed nonlinearly in the four participants' written data over the nine-month investigation. This finding substantiates previous studies on L2 writing development adopting a CDST perspective (Caspi, 2010; Chan et al., 2015; Larsen-Freeman, 2006; Ma, 2012; Penris & Verspoor, 2017; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2004, 2008, 2017, 2018; Vyatkina, 2012; Yang & Sun, 2015; Zheng, 2016). The directions of the trends were different for all lexical and syntactic complexity and accuracy indices in the four participants' written data. For example, the smoothing algorithm of the lexical variability (TTR) index displayed upward trends in

Dalma's and Emese's written data, while the smoother of the TTR index showed a downward trend in Levente's written data. Furthermore, the smoothing algorithm of the TTR index displayed a sideways trend in Avarka's written data. Moreover, the AWL index showed a sideways trend in Levente's written data. There was only one index, the general lexical complexity (AveWL) index, which displayed upward trends in the four participants' written data.

Lexical and syntactic complexity and accuracy indices demonstrated a great deal of variability in the four participants' written data over the nine-month investigation. This result supports previous studies on L2 writing development adopting a CDST perspective (Caspi, 2010; Chan et al., 2015; Larsen-Freeman, 2006; Ma, 2012; Penris & Verspoor, 2017; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2004, 2008, 2017, 2018; Vyatkina, 2012; Yang & Sun, 2015; Zheng, 2016). In addition, the amount of variability continuously changed in the lexical and syntactic complexity and accuracy indices over time. The degree of variability in the lexical and syntactic complexity and accuracy indices was also different in the four participants' written data.

The interactions between lexical and syntactic complexity and accuracy were dynamic over the nine-month investigation. The polarity of the interactions between lexical and syntactic complexity and accuracy changed from negative to positive and vice versa over time. Moreover, the magnitude of the interactions oscillated over time, ranging from weak to strong associations. This finding substantiates previous studies on L2 writing development adopting a CDST perspective (Caspi, 2010; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2008, 2017; Vyatkina, 2012). The polarity and the magnitude of the interactions between lexical and syntactic complexity and accuracy were different in the four participants' written data. The only similarity was the positive association between subordination (DC/C) and accuracy

(EFC/C) indices. However, the magnitude was different in the subordination (DC/C)-accuracy (EFC/C) interactions in the four participants' written data.

The self-regulatory processes developed nonlinearly over the nine-month investigation in the four participants' interview data. This findings substantiates previous studies on self-regulation (Forbes, 2018; Han & Hiver, 2018; Nitta & Baba, 2015, 2018; Sasaki et al., 2018). The self-regulatory processes evolved at different rates in the four participants' learning journey. The four participants' focus shifted from self-observation to self-evaluation processes at different points in time. In addition, only Dalma's and Avarka's focus shifted from self-evaluation to goal-setting over the nine-month investigation. This study supports Nitta and Baba's (2015, 2018) claim that self-evaluation processes play a more important role regarding L2 writing development than self-observation processes. In Dalma's and Avarka's interview data the ratio of self-observation to self-evaluation processes was 0.2 and 0.167 respectively (Table 5.9 and Table 8.9). In other words, the total number of self-evaluation processes were five and six times higher than the total number of self-observation processes in Dalma's and Avarka's interview data over the nine-month investigation. Conversely, in Emese's and Levente's interview data the ratio of self-observation to self-evaluation processes was 0.667 and 0.612 respectively (Table 6.9 and Table 7.9). In other words, the total number of self-evaluation processes was around the double of the total number of self-observation processes in Emese's and Levente's interview data. Zimmerman (2000) claimed that the quality and the quantity of a learner's self-regulatory processes are good indicators of the effectiveness of self-regulation. Although in this study the quality of self-regulatory processes was not measured, the quantity of the self-regulatory processes clearly shows that Dalma and Avarka were more effective in the utilisation of self-evaluation and more broadly self-regulatory processes than Emese and Levente. Therefore, it can be argued that the statistically significant developmental peaks detected in Dalma's and Avarka's written data might be linked to the more effective use of self-evaluation, and more broadly to the more effective use of self-regulation.

The findings of this study support the social cognitive perspective of self-regulation. First, the results of this study support Bandura's (1986) claim that self-regulation should be viewed as an interaction of personal, behavioural, and environmental triadic processes. According to the social cognitive perspective, self-regulation is defined in terms of context-specific processes that are used cyclically to attain personal goals. The triadic processes of self-regulation entail more than metacognitive knowledge and skill because they also include affective and behavioural processes, and a resilient sense of self-efficacy to monitor them. For example, Avarka made adjustments to her performance processes due to the feedback she received from her sister. In other words, changes in Avarka's environmental self-regulation, which refers to observing and adjusting environmental conditions or outcomes, might have motivated her to adjust performance processes.

Second, the findings of this study support Zimmerman's (2000) three-phase cyclical model of self-regulation. An important feature of the cyclical model is that it can explain dysfunctions in self-regulation as well as exemplary achievements. Dysfunctions in self-regulation are mainly due to the ineffective forethought (goal-setting) and performance control techniques (self-observation). The poorly self-regulated learners rely primarily on reactive methods instead of using proactive ones. Reactive methods of self-regulation are considered ineffective due to their inability to provide the necessary goal structure, strategic planning, and sense of personal agency for learners to progress consistently. In this study, dysfunctions in Emese and Levente's self-regulation could also be ascribed to the ineffective forethought and performance control techniques since they did not set any goals over the nine-month investigation. Furthermore, Zimmerman (2000) claimed that the presence of apathy and disinterest might also lead to dysfunctions in self-regulation. Indeed, salient attractor states (boredom) were identified in both Emese and Levente's interview data. In addition, Zimmerman (1995) claimed that self-regulation also depends on self-beliefs and affective reactions about specific performance contexts. In this study, Emese's inefficient use of self-regulation might be attributed to her negative attitude (affect) to writing while Levente's inefficient use of self-regulation might be ascribed to his low self-efficacy beliefs.

The differences in the directions of the developmental trends and in the degree of variability in lexical and syntactic complexity and accuracy indices and the differences in the polarity and magnitude of interactions between lexical and syntactic complexity and accuracy in the four participants' written data substantiate the notion of idiosyncrasy. In this study, there were no two participants who went through the same developmental path over the nine-month investigation despite the similarities in their learning context. This finding supports previous multiple-case studies on L2 writing development from a CDST perspective (Caspi, 2010; Chan et al., 2015; Larsen-Freeman, 2006; Ma, 2012; Romawati, 2016; Verspoor et al., 2017, 2018; Yang & Sun, 2015).

This study is important because it showed that second language writing development is idiosyncratic that is no two learners exhibited identical developmental paths. Although previous studies adopting a two-wave longitudinal research design could shed light on some general tendencies in L2 writing development (Barkaoui, 2016; Bulté & Housen, 2014; Crossley & McNamara, 2014; Hou et al., 2016; Knoch et al., 2015; Mazgutova & Kormos, 2015; Storch, 2009; Storch & Tapper, 2009; Verspoor et al., 2012; Vyatkina, 2012), these studies ignored the complex and dynamic nature of L2 writing development. Therefore, this study offers a complementary view on L2 writing development by adopting a multi-wave longitudinal case study research design and substantiates the findings of previous CDST studies (Caspi, 2010; Spoelman & Verspoor, 2010; Penris & Verspoor, 2017; Rosmawati, 2016; Verspoor et al., 2008). Furthermore, this study contributed to the findings of earlier CDST studies by adopting a multi-methods research design.

Chapter 11. Conclusions

Chapter 11 presents the significance, the implications, the limitations and the conclusions of this study. There are six sections: (1) theoretical significance, (2) methodological significance, (3) pedagogical significance, (4) implications for future research, (5) limitations and (6) conclusions.

11.1 Theoretical Significance

This study has contributed to the advancement of our knowledge of second language writing and showed how an interdisciplinary exploration might be fruitful for studies on second language development.

First, the findings support the CDST approach to second language development. Although the CDST perspective to applied linguistics (Larsen-Freeman & Cameron, 2008) and to second language acquisition (Verspoor et al., 2008) has received criticisms (Gregg, 2010), this study presented supportive evidence confirming the relevance of the CDST to study developmental phenomena. This study demonstrated that the development of complexity and accuracy in L2 writing reflected the main features of a dynamic system. The developmental trends of complexity and accuracy were nonlinear and idiosyncratic, and the degree of variability constantly changed over time. Furthermore, complexity and accuracy were completely interconnected, and change took place through self-organization and interaction with the environment. Complexity and accuracy were in constant change, with occasional chaotic variation. However, the systems temporarily settled into attractor states. The development of complexity and accuracy was an iterative process. Furthermore, L2 writing developed co-adaptively with self-regulatory processes. These findings support the applicability of the CDST to study second language development.

Second, the CDST is also relevant to explore the dynamic relationships between completely interconnected elements of language use, apart from studying second language development in essence. The present study demonstrated, by adopting a longitudinal multi-wave research design, that interactions between complexity and accuracy were dynamic. The moving correlations showed that the relationships between the constructs of complexity and accuracy did not remain static over time. Therefore, two-wave research designs only provide a still picture of the constantly changing interactions between complexity and accuracy resulting in disagreement in the descriptions. However, when the interactions are explored by a multi-wave research design, as offered by the CDST perspective, it becomes evident that the relationship between complexity and accuracy is dynamic. Therefore, a less static framework such as the CDST approach is more relevant to observe the relationships between the constructs. These results of the present study had important implications for hypothesising the nature of interactions between complexity and accuracy by the inclusion of the time factor in the analysis.

Third, the present study also demonstrated the importance of idiosyncrasy in second language development. The findings showed that there were no two writers with identical developmental paths. In other words, there were no two learners who went through the same developmental process despite the similar learning context. This result also presented valuable pedagogical implications (see Section 10.2). Although the general lexical complexity (AveWL) index increased in the four participants' written data over the nine-month investigation, the trajectories of the AveWL index were different from each other. For example, the AveWL index showed a peak at data point 17 in Dalma's written data, while the AveWL index showed a peak at data point 22 in Levente's data. However, there were no visible peaks in Emese's and Avarka's written data. A two-wave longitudinal research design could have demonstrated that the complexity and accuracy indices increased, decreased or stagnated in the four participants' written data over time. Nonetheless, a multi-wave research design made it possible to gain insight into the temporal changes of the complexity and accuracy indices in the four learners' written data. Therefore, the CDST approach can shed light on the idiosyncratic development of linguistic complexity in L2 writing.

Finally, the results of this present study provided an important contribution to the current body of research on L2 writing development and the relationship between L2 writing development and self-regulation. Previous studies primarily applied cross-sectional designs to explore the relationship between writing and motivation (De Bernardi & Antolini, 2007; Hidi et al., 2004), with a few exceptions in which longitudinal designs were employed (Bobbitt Nolen, 2007; Oldfather & Shanahan, 2007). Furthermore, previous studies adopting the CDST framework pointed to a possible co-adaptive development of second language writing and motivation (Caspi, 2010; Rosmawati, 2016). However, they did not explore it further. As a consequence, the present study bears theoretical significance in that it demonstrated the importance of integrating the time dimension in assessing the nature of the relationship between L2 writing and self-regulation. In addition, the present study also contributed to the limited number of studies on the co-adaptive nature of L2 writing and motivation (Nitta & Baba, 2015).

11.2 Methodological Significance

The present study explored L2 writing development from a CDST approach hence adopting the CDST principles and methodology. Firstly, this study showed how developmental profiling methods could be employed to explore the multifaceted nature of L2 writing development. This study also demonstrated how a smoothing technique, simple moving averages, could explore the developmental trends of the different measures over time.

Second, this study demonstrated how the degrees of variability could be displayed by min-max graphs. Furthermore, it was shown how data resampling and running Monte Carlo analyses could be applied to detect developmental peaks in the constructs under observation.

Although there are more modern techniques, such as the Hidden Markov Model, to detect changes in time-series data (Chan, 2015; Verspoor et al., 2017), the data resampling technique and the Monte Carlo analyses (van Geert & van Dijk, 2002) proved to be fruitful in the detection of developmental peaks in this current study. Consequently, by adopting the data resampling method this study contributed to the current research body on second language development (Zheng, 2016) and second language motivation (Piniel & Csizér, 2015).

This study explored two different types of interactions between the systems under observation: (1) surface and (2) underlying interactions. The surface interactions were calculated using an overall correlation coefficient. In this study, it was presumed that linguistic data might not be normally distributed, therefore the Spearman's rank correlation coefficient was applied to explore the surface interactions. The underlying interactions between complexity and accuracy were explored by the visual inspection of the growth trajectories and the de-trended or residual plots. The moving correlations demonstrated that the systems were in constant change over time. The magnitude and the polarity of the interactions were in constant change and none of the constructs remained static over time.

This study also applied advanced modelling procedures to test the hypothesised interactions between complexity and accuracy hence contributing to the limited number of studies which modelled the highly complex nature of L2 writing development (Caspi, 2010). While Caspi (2010) adopted the precursor model to explore both the hierarchy and interactions between the development of complexity and accuracy, this study adopted a model of three connected growers. The rationale behind the simplification of the modelling procedure was to concentrate only on one aspect: dynamic interactions. Moreover, it was presumed that upper-intermediate or advanced participants might not develop their lexicon and syntax in a hierarchy as is expected at the beginner or elementary proficiency levels.

Although the CDST methods and techniques such as developmental profiling, min-max graphs, data resampling, Monte Carlo analysis, data de-trending have been applied by previous studies on L2 writing development (Caspi, 2010; Rosmawati, 2016; Spoelman & Verspoor, 2010; Verspoor et al., 2008), this present study has contributed to the existing body of research by adopting a multi-methods research design. By conducting interviews with the participants, this study has explored the dynamic nature of self-regulation. Furthermore, this study has explored how the four participants' linguistic and motivational systems co-adapted over the nine-month investigation.

This study, by applying mathematical and statistical methods, offered a demonstration of how these analyses could be employed in the analysis of developmental data. Furthermore, this study exemplified how retro-diction might be useful to understand second language developmental processes instead of the more frequent method of prediction. The focus of this study shifted from predicting how writers would behave in the future to explaining what changes took place in their performance based on the collected data. By adopting an "explain after by before" (van Geert & Steenbeek, 2005, p. 1) perspective, this study contributed to the enrichment of the methods and techniques available in second language development research.

11.3 Pedagogical Significance

This study also presented valuable pedagogical implications for teaching and language learning. By applying profiling methods, this study provided detailed description of the writers' development over time and found two important findings.

First, the lack of some syntactic structures might imply that the participants avoided employing them and preferred using certain syntactic structures over others. This study, for instance, found that Dalma employed longer noun phrases and at the same time she also became less dependent on clausal subordination. Emese's written data showed an increase in both phrasal complexity and subordination synchronously suggesting that she made her sentences longer by using more subordinate clauses. Interestingly, Levente's and Avarka's written data displayed a decrease in phrasal complexity and an increase in subordination thus failing to display the features of academic writing. All four participants appeared to opt for a specific syntactic structure over other structures and might have overused it in their writing performance. The differences between the four participants' use of syntactic structures emerged due to the simultaneous use of both general and specific indices to measure the multifaceted constructs of syntactic complexity. Thus, the identification of different syntactic structures could be of a vital importance in understanding the development of L2 writers. It may be highly important for language teachers to detect such phenomena and discuss them with their students. Language learners might not be aware of their own syntactic structure preference thus causing the overuse of certain forms. Conversely, some second language writers might tend to avoid certain syntactic structures either deliberately or subconsciously.

Second, certain types of errors were consistent in the participants' written data over time. One of these persistent errors belonged to the formal category of errors, namely spelling errors. It might be presumed that these spelling errors fossilized in the participants' linguistic system and consequently had a snowballing effect on their written performance. Such phenomena support the theory that even advanced learners might make errors in their linguistic performance (Verspoor et al., 2012). According to the DIALANG scales of the Common European Framework of Reference, "spelling is accurate apart from occasional slips" at C1 level (Council of Europe, 2001, p. 241).

Third, this study has pedagogical implications for current language teaching practices. This study demonstrated that language development was dynamic and idiosyncratic. Language development is fundamentally the outcome of the self-organisation of the language systems. These findings may have important consequences for how languages can be taught in the classrooms. An item-processing approach cannot be fruitful since learning is not the transfer of knowledge from the teacher's head into the student's head (Lowie, 2012). According to constructivist approaches to learning, the language learner has to accomplish the task of learning and not the teacher. "Languages, in other words, cannot be taught; they can only be learned" (Lowie, 2012, p. 31). As a consequence, the teacher should provide the optimal conditions for learning under which learning can take place. The teacher cannot actively change the learners' language system as it was presumed in item processing models. Furthermore, as this study found, learning is an individual process and consequently the optimal conditions might differ for each learner.

Fourth, the findings of this research have important consequences for language testing. The dynamic nature of language development challenges the current language testing practices since any test result is just a snapshot of the language learner's skills. Instead, second language proficiency should be assessed by continuous test measures (Lowie, 2012). The Common European Framework of Reference (CEFR) is highly compatible with the CDST approach of language development since it values the multidimensionality of second language knowledge and skills, and it acknowledges the idiosyncratic nature of language development as a communicative process. Furthermore, the CEFR promotes the application of language portfolios hence acknowledges the longitudinal perspective of language development.

11.4 Implications for Future Research

This study has demonstrated that language development is a complex and dynamic process. The developmental trends were nonlinear, idiosyncratic, and displayed a constantly changing degree of variability over time. The linguistic systems were completely interconnected, and their development displayed the characteristics of an iterative process. The development of the linguistic systems was also interconnected with the motivational systems.

The present research has shown that complex and dynamic phenomena such as language development can only be captured by an organic approach such as the Complex Dynamic Systems Theory perspective. The CDST approach proved to be fruitful to gain a holistic insight into the multifaceted nature of language development. Therefore, future studies on second language development should aim to take an organic and comprehensive perspective in order to gain a full picture of language development.

One of the most important elements of the CDST approach to language development is the inclusion of the time dimension. Studies adopting the CDST framework are inherently longitudinal and apply multi-wave research designs. It would not have been possible to demonstrate the constantly changing degree of variability, and the interconnectedness of the systems without a longitudinal multi-wave research design. In other words, a two-wave research design cannot fully capture the dynamic nature of language development. Therefore, future studies should aim to devise a longitudinal language portfolio by applying a multi-wave research design (Lowie, 2012). Furthermore, second language learners do not write in one modality in a real-life setting. Instead, they write in many different modalities (for example, formal letters, essays, etc.) during their language learning development. Therefore, the language portfolio, apart from being longitudinal, should be multi-modal as well. In summary, this study suggests a multi-modal longitudinal language portfolio to create a comprehensive profile of the language learner's development. The result would be a detailed and rich data bank of the language learner's performance which could be used for analyses at many different levels and timescales. For example, a single written product could be used for microanalysis, while written products within a given modality could be used for meso-analysis and then the entire rich data bank could be used to build a comprehensive and detailed language profile of the development of the language learner in a macro-analysis.

A multimodal longitudinal language portfolio would provide the richest and the most holistic data bank of the development of the language learner. The longitudinal multi-wave data would make it possible to trace the changes in the systems. It would be possible to explore the nonlinearity, the continuously changing degree of variability and the interconnectedness of the linguistic systems. The multi-modal perspective would be the least reductionist and the most naturalistic approach to collect data.

This study has demonstrated that the development of the multifaceted constructs of lexical and syntactic complexity can only be explored by using a multidimensional measurement tapping into the sub-constructs of lexicon and syntax. In other words, a single measure would not be effective to tap into the multi-layered and complex nature of language development. This implication supports previous studies on second language writing (Norris & Ortega, 2009). Therefore, this study implies that both lexical and syntactic complexity should be explored by multiple measures tapping into the sub-constructs of the systems. Furthermore, this research proposes adjusting the choice of measures to the modality of the written texts (Lambert & Kormos, 2014; Mazgutova & Kormos, 2015). For example, formal letters and essays should be measured by different indices. This study suggests using two sets of measures: (1) at the global level and (2) at the specific level. For example, lexical complexity can be measured by the average word length index (Verspoor et al., 2017) at the macro level, while lexical complexity can be measured by the type-token ratio, the log frequency of content words, and the latent semantic analysis at the micro level. The measures at the global level could provide important information at the macro level and would be a good approximation of the construct under investigation, whereas the measures at the specific level could explore the fine-grained details at the micro level. The amalgam of these two sets of indices would capture both the general trend and the variability in the data set. This proposition is in agreement with the CDST, where variability is considered as a valuable source of information.

The results of this study showed changes in the development of the constructs of lexical and syntactic complexity by using a multidimensional measurement to tap into the sub-constructs of complexity. Both lexical and syntactic complexity were measured by five different indices including both global and genre-specific indices. However, future studies should aim to provide a fuller and more holistic description of complexity. A richer description would include the exploration of the development of morphological and phonological complexity (Bulté & Housen, 2014). The inclusion of the different levels of complexity would provide a fuller description of the language learner’s performances. Currently there is a dearth of research on phonological and morphological complexity. Consequently, there are no reliable and validated measures available. However, according to the Bottleneck Hypothesis (Slabakova, 2013), functional morphology is usually more problematic to acquire than linguistic domains such as lexicon and syntax hence it is often referred to as the bottleneck of L2 acquisition. Palotti (2015) also suggested the Morphological Complexity Index to measure the construct of morphological complexity. As a consequence, future studies might aim to validate this relatively novel index while exploring the longitudinal development of this construct.

11.5 Limitations

This study has several limitations in connection with its methodology. First, this study adopted a multiple-case study research design which provided a thick description of a “complex social issue embedded within a cultural context” (Dörnyei, 2007, p. 155). This study offered rich and in-depth insights into the dynamic nature of L2 writing development. Van Lier (2005) pointed out that “case study research has become a key method for researching changes in complex phenomena over time” (p. 195). Case study research designs are usually contrasted negatively with large-scale cross-sectional studies. However, the strength of one approach is the weakness of the other. Duff (2008) pointed out that the contrast between case studies and large-scale cross-sectional studies is unfair because the two approaches have different goals. Dörnyei (2007) recommended the multiple-case design due to the vulnerability of case studies in terms of "idiosyncratic unpredictability" (p. 155).

Second, the present study refrains from any generalisations beyond the focal points of this research. Therefore, it does not attempt to offer generalisations beyond the collected data set. Instead, it offers "particular generalisations" (Gaddis, 2002, p. 62) in line with the CDST approach. It is important to note that this study used data simulation in order to confirm the hypothesised interactions between complexity and accuracy and not to predict the development. Furthermore, Lowie and Verspoor (2018) claimed that differences between individual learners cannot be generalised beyond the individual learners under observation due to the ergodicity problem. According to the ergodicity principle, group statistics cannot be generalised to the individual and vice versa. Lowie and Verspoor (2018) concluded that two lines of research are necessary in applied linguistics: (1) group studies, and (2) single-case studies.

Third, this study took a semi-ecological approach and collected both naturalistic data (data produced not exclusively for the purpose of a specific study) and controlled data in the form of essays from the participants. The majority of the data set (16 written samples) consisted of the essays the participants produced for the EAP course they were enrolled in during the nine months. This research had no control over the topics of the naturalistic data. Although this study supplemented the naturalistic data with experimental data to control for the confounding effects of task characteristics, such a naturalistic data set has several limitations. First, the research could not monitor the writing time for each essay. The participants could spend as much time as they wished on the essays in their home environment. Therefore, the writing time could have had a possible effect on the quality of the participants' essays. Second, the naturalistic data set was elicited by writing prompts on various topics. It could be hypothesised that different topics might elicit different types of lexical items and syntactic structures. However, the 16 writing prompts in the naturalistic data set were carefully selected by the teacher of the EAP course from the same course book. Therefore, it can be presumed that there was a certain degree of uniformity in the topics and the degree of difficulty. Third, the participants might have been familiar with certain topics, while less familiar with some others. Therefore, it can be assumed that a topic on education might have been more familiar to them than a topic on work since they were all university students. Fourth, the participants had access to writing help and resources such as word processing and auto-correction software programs. Therefore, the naturalistic data set might not be the most authentic reflection of the participants' ability. As a consequence, there might be certain concerns about the adequacy of the naturalistic data in representing the underlying constructs. Since the participants had access to auto-correction tools, their error-free clause ratio might be lower for each participant. Furthermore, the participants might have used thesaurus in order to vary their vocabulary in writing resulting in the inflation of the type-token ratio and the Academic Word List indices.

Therefore, the present study abandoned the full-ecological approach and supplemented the naturalistic data set with an experimental data set of seven written samples. However, such a research design has several inherent drawbacks. First, the participants composed the seven written samples in an exam-like situation which might have caused unwarranted stress for them. Such anxiety might have exerted deleterious effects on the performance of the participants (Wong, Mahar, Titchner, & Freeman, 2013). Second, an experimental research design discounts the resources that the participants might have access to in a real-life situation. Consequently, such a design is unable to capture the language performance in its natural settings.

Fourth, the present study was limited to the exploration of only lexical and syntactic complexity instead of a holistic analysis of linguistic complexity for two reasons. First, there have not been reliable and validated indices available to measure morphological and phonological complexity. Second, the manual analysis of such a rich data bank (92 texts) would have been incredibly time-consuming.

Fifth, both the naturalistic and the experimental data sets were not lemmatised in this study. Treffers-Daller (2013) showed that data lemmatisation might be fruitful before calculating any lexical indices. Nonetheless, a lemmatisation procedure was not employed for several reasons. First, it was presumed that argumentative essays might infrequently contain different versions of a lemma since the arguments were most frequently in the same tense. Furthermore, lemmatisation might distort the possibility to explore the participants' ability to employ a wide range of derivatives that show the learners' familiarity with many different morphological complexity levels. As a consequence, this study did not apply a lemmatisation procedure of the data sets.

Finally, conducting interviews is a natural and socially acceptable way of collecting information and it can yield in-depth data (Dörnyei, 2007). However, one of the limitations of interviews is that the respondents mighty try to display himself or herself in a better than real light. Furthermore, interviewees might be too shy and inarticulate to produce enough data for the research project. Conversely, respondents can be verbose which might generate a lot of less-than-useful data. For example, in this study Emese and Levente was found to be inarticulate, while Dalma and Avarka turned out to be verbose. Furthermore, the interview questions used during the retrospective interviews might have not elicited the self-regulatory processes sufficiently. First, the participants of this study might have not been able to verbalize their thoughts and actions, although they were encouraged to switch back to their mother tongue whenever they felt it necessary. Second, direct questions were not used to elicit goal-setting in order not to direct the participants' attention to goals for improving their L2 writing development. Consequently, the low number of goals might be attributed to the nature of retrospective interviews.

11.6 Conclusions

This study demonstrated that language development was a dynamic and idiosyncratic process. Language learning can no longer be seen as a linear and finite process. Instead, it should be accepted that language learners' performance is nonlinear and shows a constantly changing degree of variability over time. Language development might show ebbs and flows over time. Furthermore, no two language learners go through the same developmental paths. The results of this research will contribute to the recognition of idiosyncrasy in language development, hence facilitating the learners' language development.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scl.59.15ai

Ainley, M. (2006). Connecting with learning: Motivation, affect and cognition in interest processes. *Educational Psychology Review*, *18*(4), 391–405. doi: https://doi.org/10.1007/s10648-006-9033-0

Anderson, J. E. (1937). An evaluation of various indices of linguistic development. *Child Development*, *8*, 62–68.

Árva, V. (2007). *The instruction of English language writing skills at secondary schools in Hungary* [Unpublished doctoral dissertation]. Eötvös Loránd University.

Assiri, M. S. (2015). Development of a rating scale for elementary EFL writing. *International Journal of Social, Management, Economics and Business Engineering*, *9*(1), 91–99. doi: urn: http://10.1999/1307-6892/10000149

Baker, R. S. J. D., D'Mello, S. K., Mercedes, M., Rodrigo, T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. doi: https://doi.org/10.1016/j.ijhcs.2009.12.003

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*(2), 390–395. doi: https://doi.org/10.2307/3587016

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, *11*(1), 17–34. doi: https://doi.org/10.1017/S0272263100007816

Barkaoui, K. (2007). Teaching writing to second language learners: Some insights from theory and research. *TESL Reporter*, *40*(1), 35–48.

Barkaoui, K. (2016). *What changes and what doesn't? An examination of changes in the linguistic characteristics of IELTS repeaters' Writing Task 2 scripts* (IELTS Research Report Series, No. 3). Retrieved from https://www.ielts.org/teaching-and-research/research-reports

Barros, R. (2007). *The nature of trends: Strategies and concepts for successful investing and trading*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bazzanella, C. (2011). Redundancy, repetition, and intensity in discourse. *Language Sciences*, *33*(2), 243–254. doi: https://doi.org/10.1016/j.langsci.2010.10.002

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(1), 1–26. doi: https://doi.org/10.1111/j.1467-9922.2009.00533.x

Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, *22*(2), 185–200. doi: https://doi.org/10.1007/s11145-007-9107-5

Benton, S. L., Corkill, A. J., Sharp, J. M., Downey, R. G., & Khramtsova, I. (1995). Knowledge, interest, and narrative writing. *Journal of Educational Psychology*, *87*(1), 66–79. doi: http://dx.doi.org/10.1037/0022-0663.87.1.66

Berman, R., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, *43*(2), 79–120. doi: https://doi.org/10.1080/01638530709336894

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: John Benjamins Publishing Company.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

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. doi: https://doi.org/10.5054/tq.2011.244483

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Bobbitt Nolen, S. (2007). The role of literate communities in the development of children's interest in writing. In P. Boscolo & S. Hidi (Eds.), *Writing and motivation* (pp. 241–255). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_014

Boekaerts, M. (1996). Personality and the psychology of learning. *European Journal of Personality*, *10*(5), 377–404. doi: https://doi.org/10.1002/(SICI)1099-0984(199612)10:5<377::AID-PER261>3.0.CO;2-N

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds). *Handbook of self-regulation: Directions and challenges for future research* (pp. 417–450). San Diego, CA: Academic Press.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlation. *Psychometrika*, *65*(1), 23–28. doi: http://dx.doi.org/10.1007/BF02294183

Boscolo, P., Del Favero, L., & Borghetto, M. (2007). Writing on an interesting topic: Does writing foster interest? In P. Boscolo & S. Hidi (Eds.), *Writing and motivation* (pp. 73–92). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_006

Boscolo, P., & Hidi, S. (2007). The multiple meanings of motivation to write. In P. Boscolo & S. Hidi (Eds.), *Writing and motivation* (pp. 1–14). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_002

Brown, R. (1973). *A first language: The early stages.* Cambridge, Mass: Harvard University Press.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.02bul

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42–65. doi: https://doi.org/10.1016/j.jslw.2014.09.005

Bybee, J. (1998). The emergent lexicon. In M. C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *CLS 34: The panels* (pp. 421–435). University of Chicago: Chicago Linguistic Society.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). London: Routledge.

Byrnes, H. (2012). Conceptualizing FL writing development in collegiate settings: A genre-based systemic functional linguistic approach. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 191–220). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781934078303.191

Byrnes, H., & Sinicrope, C. (2008). Advancedness and the development of relativization in L2 German: a curriculum-based longitudinal study. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 109–138). New York: Routledge/Taylor & Francis.

Cancino, H., Rosansky, E., & Schumann, J. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. M. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 207–230). Rowley, MA: Newbury House.

Carlisle, R. (1989). The writing of Anglo and Hispanic elementary school students in bilingual, submersion, and regular programs. *Studies in Second Language Acquisition*, *11*(3), 257–280. doi: https://doi.org/10.1017/S0272263100008135

Carroll, J. B. (1938). Diversity of vocabulary and the harmonic series law of word-frequency distribution. *The Psychological Record*, *2*, 379–386.

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

Casanave, C. P. (2012). Heading in the wrong direction: A response to Porte and Richards. *Journal of Second Language Writing*, *21*(3), 296–297. doi: https://doi.org/10.1016/j.jslw.2012.05.007

Caspi, T. (2010). *A dynamic perspective on second language development* [Unpublished doctoral dissertation]. University of Groningen. Retrieved from: https://www.rug.nl/research/portal/files/2617896/DigitalVersionCaspiDissertation.pdf

Caspi, T., & Lowie, W. (2013). The dynamics of L2 vocabulary development: A case study of receptive and productive knowledge. *Revista Brasileira de Linguística Aplicada*, *13*(2), 437–462. doi: http://dx.doi.org/10.1590/S1984-63982013005000002

Chan, H. (2015). *A dynamic approach to the development of lexicon and syntax in a second language* [Unpublished doctoral dissertation]. University of Groningen. Retrieved from: https://www.rug.nl/research/portal/files/16144166/Complete_dissertation.pdf

Chan, H., Verspoor, M., & Vahtrick, L. (2015). Dynamic development in speaking versus writing in identical twins. *Language Learning*, *65*(2), 298–325. doi: https://doi.org/10.1111/lang.12107

Chomsky, N. (1980). On cognitive structures and their development: A reply to Piaget. In M. Piattelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, Mass: Harvard University Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use.* New York, NY: Praeger.

Cobb, T. (1994). Web Vocabprofile. Retrieved from http://www.lextutor.ca/vp/

Cook, G. (2009). Use of translation in language teaching. In M. Baker & G. Saldanha (Eds.), *Routledge Encyclopaedia of translation studies* (pp. 112–115). New York: Routledge.

Cook, V. (2001). Using the first language in the classroom. *The Canadian Modern Language Review*, *57*(3), 402–423.

Cook, V. (2005). Written language and foreign language teaching. In V. Cook & B. Bassetti (Eds.), *Second language writing systems* (pp. 424–441). Clevedon, UK: Multilingual Matters. https://doi.org/10.21832/9781853597954-019

Cook, V. (2008). *Second language learning and language teaching.* London: Hodder Education.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, *5*(4), 161–170. doi: http://dx.doi.org/10.1515/iral.1967.5.1-4.161

Council of Europe. (2001). *Common European framework of reference for languages: Language, teaching, assessment.* Cambridge: Cambridge University Press.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, *32*(2), 1–16. doi: https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S. A., & McNamara, D. S. (2013). Applications of text analysis tools for spoken response grading. *Language Learning and Technology*, *17*(2), 171–192. doi: http://www.lltjournal.org/item/2815

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, *26*(4), 66–79. doi: https://doi.org/10.1016/j.jslw.2014.09.006

Crossley, S. A., Weston, J. L., Sullivan, S. T. M., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, *28*(3), 282–311. doi: http://dx.doi.org/10.1177/0741088311410188

Csapó, B., & Nikolov, M. (2001, July). *Hungarian students' performances on English and German tests.* Paper presented at the European Language Testing in a Global Context Conference. Universitat Pompeu Fabra, Barcelona, Spain.

Csapó, B., & Nikolov, M. (2009). The cognitive contribution to the development of proficiency in a foreign language. *Learning and Individual Differences*, *19*(2), 203–218. doi: https://doi.org/10.1016/j.lindif.2009.01.002

Csizér, K., Kormos, J., & Sarkadi, Á. (2010). The dynamics of language learning attitudes and motivation: lessons from an interview study of dyslexic language learners. *The Modern*

Language Journal, *94*(3), 470–487. doi: https://doi.org/10.1111/j.1540-4781.2010.01054.x

Csizér, K., & Tankó, G. (2017). English majors' self-regulatory control strategy use in academic writing and its relation to L2 motivation, *Applied Linguistics*, *38*(3), 386–404. https://doi.org/10.1093/applin/amv033

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, *39*(l), 81–141. doi: https://doi.org/10.1111/j.1467-1770.1989.tb00592.x

Cumming, A. (2001). Learning to write in a second language: two decades of research. *International Journal of English Studies*, *1*(2), 1–23. doi: http://revistas.um.es/ijes/article/view/48331/46301

Cumming, A. (Ed.) (2006). *Goals for academic writing: ESL students and their instructors.* Amsterdam: John Benjamins Publishing Company.

Cumming, A. (2009). The contribution of studies of foreign language writing to research, theories and policies. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 209–231). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691859-012

Cumming, A. (2010). Theories, frameworks, and heuristics: Some reflections on inquiry and second language writing. In T. Silva & P. K. Matsuda (Eds.), *Practicing theory in second language writing* (pp. 48–71). West Lafayette, IN: Parlor Press.

Cumming, A. (2012). Goal theory and second-language writing development, two ways. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 135–164). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781934078303.135

Cumming, A. (2016). Theoretical orientations to L2 writing. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 65–88). Boston/Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614511335-006

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System*, *26*(2), 163–174. doi: https://doi.org/10.1016/S0346-251X(98)00001-3

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, *24*(2), 197–222. doi: http://dx.doi.org/10.1093/applin/24.2.197

De Bernardi, B., & Antolini, E. (2007). Fostering students' willingness and interest in argumentative writing: an intervention study. In P. Boscolo & S. Hidi (Eds.), *Writing and motivation* (pp. 183–202). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_011

de Bot, K. (2017). Complexity Theory and Dynamic Systems Theory. In L. Ortega & Z. Han (Eds.), *Complexity Theory and language development* (pp. 51–58). Amsterdam: John Benjamins Publishing Company. doi: https://doi.org/10.1075/lllt.48.03deb

de Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 5–23). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.01deb

de Bot, K., Lowie, W., & Verspoor, M. (2005a). *Second language acquisition - An advanced resource book.* London, New York: Routledge. doi: https://doi.org/10.1075/sl.30.4.08kun

de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21. doi: https://doi.org/10.1017/S1366728906002732 de Bot, K., Verspoor, M., & Lowie, W. (2005b). Dynamic systems theory and applied linguistics: the ultimate 'so what'? *International Journal of Applied Linguistics*, *15*(1), 116–118. doi: https://doi.org/10.1111/j.1473-4192.2005.0083b.x

Dewaele, J. M. (2011). Reflections on the emotional and psychological aspects of foreign language learning and use. *Anglistik: International Journal of English Studies*, *22*(1), 23–42.

Dirven, R., & Verspoor, M. (Ed.) (2004). *Cognitive exploration of language and linguistics*. Amsterdam: John Benjamins Publishing Company. doi: https://doi.org/10.1075/clip.1

Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.) *Individual differences and instructed language learning* (pp. 137–158). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.10dor

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Oxford: Oxford University Press.

Dörnyei, Z. (2009a). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Bristol: Multilingual Matters. https://doi.org/10.21832/9781847691293-003

Dörnyei, Z. (2009b). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning*, *59*(1), 230–248. doi: https://doi.org/10.1111/j.1467-9922.2009.00542.x

Dörnyei, Z. (2009c). *The psychology of second language acquisition.* Oxford: Oxford University Press.

Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation. In E. Macaro (Ed.), *Continuum companion to second language acquisition* (pp. 247–267). London: Continuum.

Dörnyei, Z., Ibrahim, Z., & Muir, C. (2015). Directed motivational currents: Regulating complex dynamic systems through motivational surges. In Z. Dörnyei, P. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 95–105). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783092574-012

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*(3), 275–300. doi: https://doi.org/10.1177/136216880000400305

Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics*, *4*, 43–69.

Dörnyei, Z., & Tseng, W. T. (2009). Motivational processing in interactional tasks. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction* (pp. 117–134). London: Routledge.

Duff, P. (2008). *Case study research in applied linguistics.* New York: Lawrence Erlbaum Associates.

Durrant, P., & Schmitt, N. (2009). To what extent do native and nonnative writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, *47*, 157–177. doi: https://doi.org/10.1515/iral.2009.007

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap.* New York, NY: Chapman and Hall.

Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219–238). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.11.14eld

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, *48*(4), 631–664. doi: https://doi.org/10.1111/0023-8333.00063

Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In F. Meunier & S. Grainger (Eds.), *Phraseology in language learning and teaching* (pp. 1–13). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.138.02ell

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford: Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition.* Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford: Oxford University Press.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59–84. doi: https://doi.org/10.1017/S0272263104026130

Elman, J. (1995). Language as a dynamical system. In R. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 195–225). Cambridge, MA: The MIT Press.

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *4*(2), 139–155. doi: https://doi.org/10.1016/1060-3743(95)90004-7

European Commission. (2012). Europeans and their Languages. Brussels: Eurobarometer. Retrieved from http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_243_en.pdf

Evans, N. W., Hartshorn, K. J., Cox, T. L., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing*, *24*, 33–50. doi: https://doi.org/10.1016/j.jslw.2014.02.005

Fang, Z., Schleppegrell, M. J., & Cox, B. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, *38*(3), 247–273. doi: https://doi.org/10.1207/s15548430jlr3803_1

Forbes, K. (2018). The role of individual differences in the development and transfer of writing strategies between foreign and first language classrooms. *Research Papers in Education*, 1–20. doi: https://doi.org/10.1080/02671522.2018.1452963

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition*, *18*(3), 299–324. https://doi.org/10.1017/S0272263100015047

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36,* 98–116. doi: https://doi.org/10.1017/S0267190515000082

Frogner, E. (1933). Problems of sentence structure in pupils' themes. *English Journal*, *22*(9), 742–749. doi: https://doi.org/10.1207/s15548430jlr3803_1

Frydrychova Klimova, B. (2014). Constraints and difficulties in the process of writing acquisition. *Procedia Social and Behavioural Sciences*, *122*, 433–437. doi: https://doi.org/10.1016/j.sbspro.2014.01.1367

Gaddis, J. L. (2002). *The landscape of history*. Oxford: Oxford University Press.

Gaies, S. J. (1980). T-unit analysis in second language research: Applications, problems and limitations. *TESOL Quarterly*, *14*(1), 53–60. https://doi.org/10.2307/3586808

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Massachusetts: Newbury House.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah, N. J.: Lawrence Erlbaum Associates.

Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, *10*(1), 9–27. doi: https://doi.org/10.1080/15434303.2011.642040

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Good, P. I. (1999). *Resampling methods: A practical guide to data analysis.* Boston, MA: Birkhauser.

Göy, N. (2017). An action research on the development of self-regulated writing strategies of Turkish EFL students. *Eurasian Journal of Applied Linguistics*, *3*(2), 191–204. doi: http://www.ejal.eu/index.php/ejal/article/view/150

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. doi: http://dx.doi.org/10.3102/0013189X11413260.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. doi: http://dx.doi.org/10.3758/BF03195564

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, *9*(2), 123–145. doi: http://dx.doi.org/10.1016/S1060-3743(00)00019-9

Gray, B. (2015). *Linguistic variation in research articles: when discipline tells only part of the story.* Amsterdam: John Benjamins Publishing Company.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach.* London: Chapman and Hall.

Gregg, K. R. (2010). Review article: Shallow draughts: Larsen-Freeman and Cameron on complexity. *Second Language Research*, *26*(4), 549–560. doi: https://doi.org/10.1177/0267658310366582

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire: Essai de méthodologie*. Paris: Presses universitaires de France.

Hall, G., & Cook, G. (2012). Own-language use in language teaching and learning. *Language Teaching*, *45*(3), 271–308. doi: https://doi.org/10.1017/S0261444812000067

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Halliday, M. A. K., & Martin, J. R. (1993/1996). *Writing science: Literacy and discursive power.* London, England: Falmer Press.

Halliday, M. A. K., & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition.* London: Cassell.

Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing*, *40*, 44–59. doi: https://doi.org/10.1016/j.jslw.2018.03.001

Hawking, S. (2000, January 23). "Unified Theory" is getting closer, Hawking Predicts. Interview in *San Jose Mercury News,* 29A.

Heatley, A., & Nation, P. (1994). Range. Victoria University of Wellington, NZ. [Computer program]. Retrieved from http://www.vuw.ac.nz/lalsl

Heckhausen, J., & Dweck, C. S. (Eds.). (1998). *Motivation and self-regulation across the life span.* Cambridge: Cambridge University Press.

Hedgcock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing*, *1*(3), 255–276. doi: https://doi.org/10.1016/1060-3743(92)90006-B

Henry, A. J. (2015). The dynamics of L3 motivation: A longitudinal interview/observation-based study. In Z. Dörnyei, P. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 83–94). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783092574-021

Hidi, S., & Boscolo, P. (Eds.) (2007). *Writing and motivation.* Amsterdam: Elsevier.

Hidi, S., & McLaren, J. (1990). The effect of topic and theme interestingness on the production of school expositions. In H. Mandl, E. De Corte, N. Bennett, & H. Freidrich (Eds), *Learning and instruction: European research in an international context* (pp. 295–308). Oxford: Pergamon Press.

Hidi, S., & McLaren, J. (1991). Motivational factors and writing: The role of topic interestingness. *European Journal of Psychology of Education*, *6*(2), 187–197. https://doi.org/10.1007/BF03191937

Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion and cognition* (pp. 89–115). Mahwah, NJ: Lawrence Erlbaum Associates.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319–340. doi: http://dx.doi.org/10.1037/0033-295X.94.3.319

Hiver, P. (2015). Attractor states. In Z. Dörnyei, P. Mac-Intyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 20–28). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781783092574-005

Hiver, P., & Al-Hoorie, A. H. (2016). A Dynamic ensemble for second language research: Putting Complexity Theory into practice. *The Modern Language Journal*, *100*(4), 741–756. doi: https://doi.org/10.1111/modl.12347

Hou, J. (2017). *The dynamics of English writing development in advanced Chinese learners* [Unpublished doctoral dissertation]. University of Groningen. Retrieved from: https://www.rug.nl/research/portal/files/44214133/Complete_thesis.pdf

Hou, J., Verspoor, M., & Loerts, H. (2016). An exploratory study into the dynamics of Chinese L2 writing development. *Dutch Journal of Applied Linguistics*, *5*(1), 65–96. https://doi.org/10.1075/dujal.5.1.04loe

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–473. doi: http://hdl.handle.net/11245/1.323877

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (National Council of Teachers of English Research Report No. 3). National Council of Teachers of English, Urbana.

IELTS Test taker performance 2016. (2016). Retrieved from https://www.ielts.org/teaching-and-research/test-taker-performance or https://intellilab.enz.govt.nz/document/415-2016-ielts-test-taker-peformance

Ionin, T. (2007). DST vs. UG: Can DST account for purely linguistic phenomena? *Bilingualism: Language and Cognition*, *10*(1), 27–29. doi: https://doi.org/10.1017/S1366728906002756

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, *4*(1), 51–70. doi: https://doi.org/10.1016/1060-3743(95)90023-3

Jakeman, V., & MacDowell, C. (1997). *Cambridge practice tests for IELTS 1*. Cambridge: Cambridge University Press.

Jakeman, V., & MacDowell, C. (2000). *Cambridge practice tests for IELTS 2*. Cambridge: Cambridge University Press.

Jakeman, V., & MacDowell, C. (2002). *Cambridge practice tests for IELTS 3*. Cambridge: Cambridge University Press.

Jakeman, V., & MacDowell, C. (2005). *Cambridge practice tests for IELTS 4*. Cambridge: Cambridge University Press.

Jakeman, V., & MacDowell, C. (2006). *Cambridge practice tests for IELTS 5*. Cambridge: Cambridge University Press.

Jakeman, V., & MacDowell, C. (2007). *Cambridge practice tests for IELTS 6*. Cambridge: Cambridge University Press.

Jakeman, V., & MacDowell, C. (2009). *Cambridge practice tests for IELTS 7*. Cambridge: Cambridge University Press.

James, C. (1998). *Errors in language learning and use*. London: Routledge.

Jarvis, S. (2002). Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing*, *19*(1), 57–84. doi: http://dx.doi.org/10.1191/0265532202lt220oa

Jarvis, S. (2013a). Capturing the diversity in lexical diversity. *Language Learning*, *63*(1), 87–106. doi: https://doi.org/10.1111/j.1467-9922.2012.00739.x

Jarvis, S. (2013b). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 13–45). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.47.03ch1

Johnson, W. (1939). *Language and speech hygiene: An application of general semantics*. Ann Arbor, MI: Edwards Brothers.

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs*, *56*(2), 1–15. doi: http://dx.doi.org/10.1037/h0093508

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2). doi: https://doi.org/10.1177/1558689806298224

Jordan, R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, *1*(1), 69–78. https://doi.org/10.1016/S1475-1585(02)00004-8

Jucker, A. H. (1992). *Social stylistics: Syntactic variation in British newspapers*. New York: Mouton de Gruyter.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, *102*(1), 120–141. doi: https://doi.org/10.1111/modl.12447

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, *28*, 39–52. doi: https://doi.org/10.1016/j.jslw.2015.02.005

Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English medium university? *Assessing Writing*, *21*(1), 1–17. doi: https://doi.org/10.1016/j.asw.2014.01.001

Kobayashi, H., & Rinnert C. (2012). Understanding L2 writing development from a multicompetence perspective: Dynamic repertoires of knowledge and text construction. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 101–134). Boston, MA: De Gruyter Mouton. https://doi.org/10.1515/9781934078303.101

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*(2), 148–161. doi: https://doi.org/10.1016/j.jslw.2011.02.001

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403. doi: https://doi.org/10.1016/j.jslw.2012.09.003

Kormos, J., & Csizér, K. (2014). The interaction of motivation, self-regulatory strategies, and autonomous learning behavior in different learner groups. *TESOL Quarterly*, *48*(2), 275–299. doi: https://doi.org/10.1002/tesq.129

Kormos, J., & Dörnyei, Z. (2004). The interaction of linguistic and motivational variables in second language task performance. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, *9*(2). http://zif.spz.tu-darmstadt.de/jg-09-2/beitrag/

Kovacsik, Á. (2016, May 22). Virágzik a szakdogagyártás. *Magyar Nemzet Online*. Retrieved from: https://mno.hu/belfold/oriasi-biznisz-a-szakdoga-gyartas-1342662

Krashen, S. (1977). Some issues related to the Monitor Model. In H. D. Brown, C. Yorio, & R. Crymes (Eds.) *On TESOL '77: Teaching and learning English as a second language: Trends in research and practice*. Washington: TESOL.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 140–154). Cambridge: Cambridge University Press.

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 261–284. doi: https://doi.org/10.1515/iral.2007.012

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *17*(1), 48–60, https://doi.org/10.1016/j.jslw.2007.08.003

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333–349. doi: https://doi.org/10.1111/modl.12468

LaBrant, L. L. (1933). A study of certain language developments of children in grades four to twelve, inclusive. *Genetic Psychology Monographs*, *14*, 387–491.

Lam, R. (2015). Understanding EFL students' development of self-regulated learning in a process-oriented writing course. *TESOL Journal*, *6*(3), 527–553. doi: https://doi.org/10.1002/tesj.179

Lambert, C., & Kormos, J. (2014). Complexity, accuracy and fluency in task-based second language research: toward more developmentally-based measures of second language acquisition. *Applied Linguistics*, *35*(5), 607–614. doi: https://doi.org/10.1093/applin/amu047

Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. W. (2008). Cognitive grammar as a basis for language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 66–88). London: Routledge.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, *18*(2), 141–165. doi: https://doi.org/10.1093/applin/18.2.141

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*, 590–616. doi: https://doi.org/10.1093/applin/aml029

Larsen-Freeman, D. (2015). Ten 'lessons' from Complex Dynamic Systems Theory: What is on offer. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in* *language learning* (pp. 11–19). Bristol: Multilingual Matters Ltd. https://doi.org/10.21832/9781783092574-004

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics.* Oxford: Oxford University Press.

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language, Learning & Technology*, *19*(2), 50–68. doi: http://dx.doi.org/10125/44417

Lee, I. (2016). EFL writing in schools. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 113–139). Boston/Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614511335-008

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417. doi: https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Lennon, P. (1991). Error: some problems of definition and identification. *Applied Linguistics*, *12*(2), 180–195. doi: https://doi.org/10.1093/applin/12.2.180

Lens, W., & Vansteenkiste, M. (2008). Promoting self-regulated learning: A motivational analysis. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research and applications* (pp. 141–168). New York, NY: Lawrence Erlbaum.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85–102. doi: https://doi.org/10.1016/j.jslw.2009.02.001

Li, X. (2018). Language enhancement for translation undergraduates: an evaluation of students' reactions to and effectiveness of a comparable and parallel text analysis course. *Journal of Multilingual and Multicultural Development*, *39*(6), 475–490. doi: https://doi.org/10.1080/01434632.2018.1439948

Li, Y. (2000). Linguistic characteristics of ESL writing in task-based e-mail activities. *System*, *28*, 229–245. doi: https://doi.org/10.1016/S0346-251X(00)00009-9

Little, L. A. (2011). *Trend qualification & techniques to identify the best trends to trade*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Little, L. A., & Farley, A. (2012). *Trend trading set-ups: Entering and exiting trends for maximum profit.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. (Research Report No. 18). National Council of Teachers of English.

Lock, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year Odyssey. *American Psychologist*, *57*(9), 705–717. doi: http://dx.doi.org/10.1037/0003-066X.57.9.705

Long, D. (1998). Poor writing will harm a web site. *Computer Dealer News*, *14.* 17.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego: Academic Press.

Lowie, W. (2012). The CEFR and the dynamics of second language learning: trends and challenges. Language learning in higher education. *Journal of the European Confederation of Language Centres in Higher Education*, *2*(1), 17–34. doi: https://doi.org/10.1515/cercles-2012-0002

Lowie, W. (2017). Emergentism: wide ranging theoretical framework or just one more meta-theory? *Recherches en didactique des langues et des culture. Les cahiers de l'Acedle*, *14*(1).

Lowie, W., de Bot, K., & Verspoor, M. (2009). A dynamic view of second language development. In K. de Bot & R. W. Schrauf (Eds.), *Language development over the lifespan* (pp. 125–145). New York, London: Routledge.

Lowie, W., & Verspoor, M. (2018). Individual differences and the Ergodicity Problem. *Language Learning*, *0*(0). doi: https://doi.org/10.1111/lang.12324

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. doi: https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. doi: https://doi.org/10.5054/tq.2011.240859

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. doi: https://doi.org/10.1177/0265532217710675

Lu, X., & Ai, H. (2011). Synlex: Lexical complexity analyzer. L2 syntactical complexity analyzer. [Computer software]. Retrieved from http://aihaiyang.com/software/

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, *29*, 16–27. doi: https://doi.org/10.1016/j.jslw.2015.06.003

Ma, L. (2012). *A dynamic perspective on the variability of lexicon and syntax in second language development* [Unpublished master's dissertation]. University of Groningen.

Macaro, E. (2006). Codeswitching in the L2 classroom: A communication and learning strategy. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges, and contributions to the professions* (pp. 63–84). Amsterdam: Springer. doi: http://dx.doi.org/10.1007/0-387-24565-0_5

MacIntyre, P. D., & Serroul, A. (2015). Motivation on a per-second timescale: Examining approach-avoidance motivation during L2 task performance. In Z. Dörnyei, P. D.

MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 109–138). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783092574-013

MacWhinney, B. (2008). A unified model. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341–371). London: Routledge.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58–71). Clevedon, UK: Multilingual Matters.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment.* New York: Palgrave Macmillan.

Manchón, R. M. (2001). Trends in the conceptualizations of second language composing strategies: A critical analysis. *International Journal of English Studies*, *1*(2), 47–70. doi: http://revistas.um.es/ijes/article/view/48081/46061

Manchón, R. M. (2009a). Introduction. Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 102–129). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691859-004

Manchón, R. M. (2009b). Individual differences in foreign language writing: The dynamics of beliefs about L2 writing. *Revista Española de Lingüística Aplicada*, *22*, 245–268.

Manchón, R. M. (2011). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3–14). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.31.03man

Manchón, R. M. (2012). Multiple perspectives on L2 writing development: An introduction. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 1–16). Boston, MA: De Gruyter Mouton. https://doi.org/10.1515/9781934078303.1

Mancón, R. M. (2016). Introduction: Past and future of L2 writing research. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 1–15). Boston/Berlin: De Gruyter. https://doi.org/10.1515/9781614511335-003

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 102–129). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691859-008

Matthews, G., & Zeidner, M. (2004). Traits, states, and the trilogy of mind: an adaptive perspective on intellectual functioning. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion and cognition* (pp. 143–171). Mahwah, NJ: Lawrence Erlbaum Associates.

Mazgutova, D. (2015). *Linguistic and cognitive development of L2 writing during an intensive English for Academic Purposes (EAP) programme* [Unpublished doctoral dissertation]. Lancaster University.

Mazgutova, D., & Kormos J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, *29*, 3–15. doi: http://dx.doi.org/10.1016/j.jslw.2015.06.004

McCarter, S., & Ash, J. (2003). *IELTS testbuilder*. Oxford: Macmillan.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* [Unpublished doctoral dissertation]. University of Memphis.

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459–488. doi: http://dx.doi.org/10.1177/0265532207080767

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. doi: http://dx.doi.org/10.3758/BRM.42.2.381

McNamara, D. S., Graesser, A. C., Cai, Z., & Kulikowich, J. M. (2011). *Coh-Metrix easability components: Aligning text difficulty with theories of text comprehension.* Paper presented at the annual meeting of the American Educational Research Association New Orleans, LA April.

Mehrabi, M., Kalantarian, S. R., & Boshrabaid, A. M. (2016). The interplay between self-regulation strategies, academic writing achievement and gender in an Iranian L2 context. *Journal of Applied Linguistics and Language Research*, *3*(2), 230–239. doi: http://www.jallr.com/index.php/JALLR/article/view/288

Mellow, D., Reeder, K., & Forster, E. (1996). Using time-series research designs to investigate the effects of instruction on SLA. *Studies in Second Language Acquisition*, *18*(3), 325–350. doi: https://doi.org/10.1017/S0272263100015059

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 241–259. doi: https://doi.org/10.1515/iral.2007.011

Mihaljević Djigunović, J., Nikolov, M., & Ottó, I. (2008). A comparative study of Croatian and Hungarian EFL students. *Language Teaching Research*, *12*(3), 433–452. doi: https://doi.org/10.1177%2F1362168808089926

Molnár, K. (2007). Néhány gondolat a szakdolgozatok elkészítésének színvonaláról a Rendőrtiszti Főiskolán. *Matarka*, *7(1)*, 38–44. doi:

http://www.nyelvilektoralas.hu/publikacio/nehany_gondolat_a_szakdolgozatok_elkeszitesenek_szinvonalarol_rtf.pdf

Nami, Y., Enayati, T., & Ashouri, M. (2012). The relationship between self-regulation approaches and learning approaches in English writing tasks on English foreign language students. *Procedia Social and Behavioural Sciences*, *47*, 614–618. doi: https://doi.org/10.1016/j.sbspro.2012.06.705

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge, UK: CUP.

Nelder, J. A., & Mead, R. (1965). A Simplex method for function minimization. *Computer Journal*, *7*(4), 308–313. doi: https://doi.org/10.1093/comjnl/7.4.308

Nicolás-Conesa, F., Roca de Larios, J., & Coyle, Y. (2014). Development of EFL students' mental models of writing and their effects on performance. *Journal of Second Language Writing*, *24*(2), 1–19. doi: https://doi.org/10.1016/j.jslw.2014.02.004

Nikolov, M. (2001). Minőségi nyelvoktatás – a nyelvek európai évében. *Iskolakultúra, 8*, 3–11. doi: http://www.iskolakultura.hu/ikultura-folyoirat/documents/2001/8/tanulm2001-8.pdf

Nikolov, M. (2002). *Issues in English Language Education.* Bern: Peter Lang.

Nippold, M. A. (2004). Research on later language development. The school-age and adolescent years. In R. Berman (Ed.), *Language development across childhood and adolescence* (pp. 1–8). Amsterdam: John Benjamins Publishing Company.

Nitta, R., & Baba, K. (2015). Self-regulation in the evolution of the ideal L2 self: A complex dynamic systems approach to the L2 motivational self system. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 374–404). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783092574-023

Nitta, R., & Baba, K. (2018). Understanding benefits of repetition from a complex dynamic systems perspective. In M. Bygate (Ed.), *Learning language through task repetition* (pp.

279–309). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tblt.11.11nit

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. doi: https://doi.org/10.1093/applin/amp044

Norris, J. M., & Manchón, R. M. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 221–244). Boston, MA: De Gruyter Mouton. https://doi.org/10.1515/9781934078303.221

Oldfather, P., & Shanahan, C. H. (2007). A cross-case study of writing motivation as empowerment. In S. Hidi & P. Boscolo (Eds.), *Writing and motivation* (pp. 257–279). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_015

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218–233. doi: https://doi.org/10.1016/j.jslw.2010.10.003

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology: Theory & Practice*, *8*(5), 375–387. doi: http://dx.doi.org/10.1080/13645570500402447

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492–518. doi: https://doi.org/10.1093/applin/24.4.492

Ortega, L. (2012). Foreword. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. x–xii). Boston, MA: De Gruyter Mouton. https://doi.org/10.1515/9781934078303.x

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45. doi: https://doi.org/10.1017/S0267190505000024

Oxenden, C., Latham-Koenig, C., & Storton, L. (2006). *New English File*. Oxford: Oxford University Press.

Oyama, S., Griffiths, P., & Gray R. (2001). *Cycles of contingency, development systems and evolution*. Cambridge, MA: The MIT Press.

Oyserman, D., & James, L. (2009). Possible selves: From content to process. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 373–394). NY: Psychology Press.

Pajares, F., Valiante, G., & Cheong, Y. F. (2007). Writing self-efficacy and its relation to gender, writing motivation, and writing competence: a developmental perspective. In P. Boscolo & S. Hidi (Eds.), *Writing and motivation* (pp. 141–159). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_009

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590–601. doi: https://doi.org/10.1093/applin/amp045

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117–134. doi: https://doi.org/10.1177/0267658314536435

Parks, S. (2016). Workplace writing: From text to context. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 223–241). Boston/Berlin: De Gruyter. https://doi.org/10.1515/9781614511335-013

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531–549. doi: http://dx.doi.org/10.1037/a0019243

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology*, *106*(3), 696–710. doi: http://dx.doi.org/10.1037/a0036006

Penris, W., & Verspoor, M. (2017). Academic writing development: a complex, dynamic process. In S. Pfenniger & J. Navracsics (Eds.), *Future research directions for applied linguistics* (pp. 215–242). Bristol; Tonawanda, NY; North York, Ontario: Multilingual Matters Ltd. https://doi.org/10.21832/9781783097135-012

Petneki, K. (2009). *Az idegen nyelvek oktatása az ezredfordulón.* Szeged, Hungary: JATE-Pressz.

Petrić, B. (2012). Legitimate textual borrowing: Direct quotation in L2 student writing. *Journal of Second Language Writing*, *21*(2), 102–117. doi: https://doi.org/10.1016/j.jslw.2012.03.005

Piniel, K., & Csizér, K. (2015). Changes in motivation, anxiety and self-efficacy during the course of an academic writing seminar. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 187–201). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783092574-015

Pintrich, P. R., & De Groot, E. V. (1990). Motivation and self-regulated learning components of academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. doi: http://dx.doi.org/10.1037/0022-0663.82.1.33

Plakans, L. (2013). Writing scale development and use within a language program. *TESOL Journal*, *4*(1), 151–163. doi: https://doi.org/10.1002/tesj.66

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, *47*(1), 101–143. doi: https://doi.org/10.1111/0023-8333.31997003

Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching*, *50*(2), 261–275. doi: https://doi.org/10.1017/S0261444817000015

Polio, C., & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research.* New York: Routledge.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. doi: https://doi.org/10.1016/j.jslw.2014.09.003

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Reppen, R. (1994). *Variation in elementary student language: A multi-dimensional perspective* [Unpublished doctoral dissertation]. Northern Arizona University.

Rinnert, C., & Kobayashi, H. (2009). Situated writing practices in foreign language settings: The role of previous experience and instruction. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 23–48). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781847691859-005

Robinson, P. (2001). *Cognition and second language instruction.* Cambridge: Cambridge University Press.

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.012

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57. doi: https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Malden, Ma: Blackwell.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance.* Amsterdam: John Benjamins Publishing Company.

Roca, J. (1996). Linearization strategies in L2 writing: Some observations. *Lenguaje y Texlos*, *8*, 191–208.

Rosmawati, R. (2016). *Dynamic development and interactions of complexity, accuracy, and fluency in ESL academic writing* [Unpublished doctoral dissertation]. University of Sydney. Retrieved from: https://ses.library.usyd.edu.au/handle/2123/15901

Rumelhart, D., & McClelland, J. (1987). Learning the past tense of English verbs: Implicit rules or parallel processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 195–248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. doi: http://dx.doi.org/10.1037/0003-066X.55.1.68

Samanian, S., & Roohani, A. (2018). Effects of self-regulatory strategy development on EFL learners' descriptive writing and reflective thinking. *Research in English Language Pedagogy*, *6*(1), 95–116. doi: http://relp.khuisf.ac.ir/article_538782.html

Sansone, C., & Smith, J. L. (2000). The "how" of goal pursuit: Interest and self-regulation. *Psychological Inquiry*, *11*(4), 306–309. https://doi.org/10.1207/S15327965PLI1104_02

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, *9*(3), 259–291. doi: https://doi.org/10.1016/S1060-3743(00)00028-X

Sasaki, M. (2009). Changes in English as a foreign language students' writing over 3.5 years: A sociocognitive account. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 49–76). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691859-006

Sasaki, M., Mizumoto, A., & Murakami, A. (2018). Developmental trajectories in L2 writing strategy use: A self-regulation perspective. *The Modern Language Journal*, *102*(2), 292–309. doi: https://doi.org/10.1111/modl.12469

Schmid, M. S., Verspoor, M., & MacWhinney, B. (2011). Coding and extracting data. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 39–54). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.03sch

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. doi: https://doi.org/10.1093/applin/11.2.129

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.

Scott, C. M. (1988). Spoken and written syntax. In M. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 49–95). Boston: Little, Brown.

Seker, M. (2016). The use of self-regulation strategies by foreign language learners and its role in language achievement. *Language Teaching Research*, *20*(5), 600–618. doi: https://doi.org/10.1177/1362168815578550

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, *10*(1–4), 209–241. doi: http://dx.doi.org/10.1515/iral.1972.10.1-4.209

Serafini, E. J. (2017). Exploring the dynamic long-term interaction between cognitive and psychosocial resources in adult second language development at varying proficiency.

The Modern Language Journal, *101*(2), 369–390. doi: https://doi.org/10.1111/modl.12400

Skehan, P. (1996). A Framework for the implementation of task-based instruction, *Applied Linguistics*, *17*(1), 38–62. doi: https://doi.org/10.1093/applin/17.1.38

Skehan, P. (1998). *A Cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2003). Task based instruction. *Language Teaching*, *36*(1), 1–14. doi: https://doi.org/10.1017/S026144480200188X

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. doi: https://doi.org/10.1093/applin/amp047

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185–211. doi: https://doi.org/10.1177/136216889700100302

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.009

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–216). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.11.12ske

Slabakova, R. (2013). What is easy and what is hard to acquire in a second language. In M. García Mayo, M. Gutierrez Mangado, & M. Martínez Adrián (Eds.), *Contemporary approaches to second language acquisition* (pp. 5–28). Amsterdam/Philadelphia: John Benjamin Publishing Company. https://doi.org/10.1075/aals.9.04ch1

Smith, L. B., & Thelen, E. (1993). *A dynamic systems approach to development: Applications.* Cambridge, MA, US: The MIT Press.

Smith, Z. R., & Wells, C. S. (2006, October 18–20). *Central Limit Theorem and Sample Size.* Retrieved from: The annual meeting of the North-eastern Educational Research Association: http://www.umass.edu/remp/Papers/Smith&Wells_NERA06.pdf

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*(4), 532–553. doi: https://doi.org/10.1093/applin/amq001

Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 443–466). Thousand Oaks, CA.: Sage Publications Ltd.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149–183. doi: https://doi.org/10.1177/0741088316631527

Stern, H. (1992). *Issues and options in language teaching*. Oxford: Oxford University Press.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, *14*(3), 153–173. doi: https://doi.org/10.1016/j.jslw.2005.05.002

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, *18*, 103–118. doi: http://dx.doi.org/10.1016/j.jslw.2009.02.003

Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, *8*, 207–223. doi: http://dx.doi.org/10.1016/j.jeap.2009.03.001

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning*, *62*, 1110–1133. doi: https://doi.org/10.1111/j.1467-9922.2012.00720.x

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Tardy, C. M. (2012). A rhetorical genre theory perspective on L2 writing development. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 165–190). De Gruyter Mouton. https://doi.org/10.1515/9781934078303.165

Tarone, E., Downing, B., Cohen, A., Gillette, S., Murie, R., & Dailey, B. (1993). The writing of Southeast Asian-American students in secondary school and university. *Journal of Second Language Writing*, *2*(2), 149–172. doi: https://doi.org/10.1016/1060-3743(93)90015-U

Templin, M. (1957). *Certain language skills in children.* Minneapolis, MN: University of Minneapolis Press.

Tang, J. (2002). Using L1 in the English classroom. *English Teaching Forum*, *40*(1), 36–43. doi: http://exchanges.state.gov.forum/vols/vol40/no1/p36.pdf

Teng, L. S., & Zhang, L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal*, *100*(3), 674–701. doi: https://doi.org/10.1111/modl.12339

Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic Dialogues*, *15*(2), 255–283. doi: http://dx.doi.org/10.1080/10481881509348831

Thelen E., Corbetta D., & Spencer J. P. (1996). The development of reaching during the first year: The role of movement speed. *Journal of Experimental Psychology: Human*

Perception and Performance, *22*(5), 1059–1076. doi: http://dx.doi.org/10.1037/0096-1523.22.5.1059

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, Massachusetts: MIT Press.

Thelen, E., & Ulrich, B. D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development*, *56*(1), 1–107. doi: http://dx.doi.org/10.2307/1166099

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: An exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 79–104). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.47.05ch3

Ushioda, E. (1996). *Learner autonomy 5: The role of motivation.* Dublin, Ireland: Authentik.

Ushioda, E. (2003). Motivation as a socially mediated process. In D. Little, J. Ridley, & E. Ushioda (Eds.), *Learner autonomy in the foreign language classroom: Teacher, learner, curriculum and assessment* (pp. 90–102). Dublin, Ireland: Authentik.

Ushioda, E. (2006). Motivation, autonomy and sociocultural theory. In P. Benson (Ed.), *Learner autonomy 8: Insider perspectives on autonomy in language learning and teaching* (pp. 5–24). Dublin, Ireland: Authentik.

van Dijk, M., Verspoor, M., & Lowie, W. (2011). Variability and DST. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 55–85). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.04van

van Geert, P. (1991). A dynamic systems theory model of cognitive and language growth. *Psychological Review*, *98*(1), 3–53. doi: http://dx.doi.org/10.1037/0033-295X.98.1.3

van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos.* London: Harvester Wheat sheaf.

van Geert, P. (1995). Growth dynamics in development. In R. F. Port & T. J. Van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 313–338). Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/4622.003.0012

van Geert, P. (2003). Dynamic systems approaches and modelling of developmental processes. In J. Valsiner & K. Connoly (Eds.), *Handbook of developmental psychology* (pp. 640–672). London: SAGE Publications Inc.

van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal*, *92*(2), 179–199. doi: https://doi.org/10.1111/j.1540-4781.2008.00713.x

van Geert, P., & Steenbeek, H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, *25*(3–4), 408–442. doi: http://dx.doi.org/10.1016/j.dr.2005.10.003

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior & Development*, *25*(4), 340–374. doi: http://dx.doi.org/10.1016/S0163-6383(02)00140-6

van Geert, P., & van Dijk, M. (2003). Ambiguity in child language: The problem of inter-observer reliability in ambiguous observation data. *First Language*, *23*(3), 259–284. doi: https://doi.org/10.1177/01427237030233001

van Gelder, T., & Port, R. (Eds.) (1995). *Mind as motion: Explorations in the dynamics of cognition.* MIT Press.

van Lier, L. (2005). Case study. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 195–208). Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B., & Benati, A. G. (2010). *Key terms in second language acquisition.* London: Continuum.

Verspoor, M. (2015). Initial conditions. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 38–46). Bristol: Multilingual Matters Ltd. https://doi.org/10.21832/9781783092574-007

Verspoor, M., & Behrens, H. (2011). Dynamic systems theory and a usage-based approach to second language development. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 25–38). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.02ver

Verspoor, M., de Bot, K., & Lowie, W. (2004). Dynamic systems theory and variation: a case study in L2-writing. In M. Hannay, H. Aertsen, & R. Lyall (Eds.), *Words in their places: A Festschrift for J. Lachlan* (pp. 407–421). Amsterdam: Free University Press.

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 culture. Les cahiers de l'Acedle, 14*(1), 1–27. doi: http://rdlc.revues.org/1450

Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in L2 development from a dynamic systems perspective. *The Modern Language Journal*, *92*(2), 214–231. doi: https://doi.org/10.1111/j.1540-4781.2008.00715.x

Verspoor, M., Lowie, W., van Geert, P., van Dijk, M., & Schmid, M. S. (2011). How to sections. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language* *development: Methods and techniques* (pp. 129–200). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.08ver

Verspoor, M., Lowie, W., & Wieling, M. (2018). L2 developmental measures from a dynamic perspective. In M. Paquot & B. LeBruyn (Eds.), *Learner Corpora and Second Language Acquisition.* Cambridge: Cambridge University Press.

Verspoor, M., & Sauter, K. (2000). *English sentence analysis: An introductory course.* Amsterdam: John Benjamins Publishing Company.

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage-based perspective on L2 writing. *Journal of Second Language Writing*, *21*(3), 239–263. doi: https://doi.org/10.1016/j.jslw.2012.03.007

Verspoor, M., & van Dijk, M. (2011). Visualizing interaction between variables. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 85–98). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.05ver

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, *96*(4), 576–598. doi: https://doi.org/10.1111/j.1540-4781.2012.01401.x

Waninge, F. (2015). Motivation, Emotion and Cognition: Attractor States in the Classroom. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 195–213). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783092574-016

Waninge, F., de Bot, K., & Dörnyei, Z. (2014). Motivational dynamics in language learning: Change, stability, and context. *The Modern Language Journal*, *98*(3), 704–723. doi: https://doi.org/10.1111/modl.12118

Warchal, K. (2010). Moulding interpersonal relations through conditional clauses, consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes*, *9*, 140–150. doi: http://dx.doi.org/10.1016/j.jeap.2010.02.002.

Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, *84*(2), 171–184. doi: https://doi.org/10.1111/0026-7902.00060

Wigfield, A., Hoa, L. W., & Klauda, S. L. (2008). The role of achievement values in the self-regulation of achievement behaviors. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 169–195). New York, NY: Lawrence Erlbaum.

Wigglesworth, G., & Foster, P. (2008, April). *Measuring accuracy in second language performance.* New York City, New York: Paper presented at the TESOL Convention.

Wigglesworth, G., & Storch, N. (2012). Feedback and writing development through collaboration: A socio-cultural approach. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 69–100). De Gruyter Mouton. https://doi.org/10.1515/9781934078303.69

Wind A. M. (2014). Second language writing development from a Dynamic Systems Theory perspective. In B. P. Olmos-Lopez, J. Huang, & J. Almeida (Eds.), *Proceedings of the 8th Lancaster University Conference in Linguistics & Language Teaching* (pp. 90–123). Lancaster: University of Lancaster. https://www.lancaster.ac.uk/fass/events/laelpgconference/papers/v08/Attila-M-Wind.pdf

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity.* Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Wong, I., Mahar, D., Titchener, K., & Freeman, J. (2013). The impact of anxiety on processing efficiency: Implications for the Attentional Control Theory. *The Open Behavioural Science Journal*, *6*, 7–15. doi: https://doi.org/10.2174/1874230001307010007

Wyatt, R. (2001). *Check your English vocabulary for IELTS*. Bloomsbury Information.

Yang, H. C., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, *46*(1), 80–103. doi: https://doi.org/10.1002/tesq.6

Yang, W., & Sun, Y. (2015). Dynamic development of complexity, accuracy, and fluency in multilingual learners' L1, L2 and L3 writing. *Theory and Practice in Language Studies*, *5*(2), 298–308. doi: http://dx.doi.org/10.17507/tpls.0502.09

Yashima, T., & Arano, K. (2015). Understanding EFL learners' motivational dynamics: A three-level model from a dynamic systems and sociocultural perspective. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 285–314). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781783092574-020

Yin, R. K. (1984). *Case study research: Design and methods.* Beverly Hills, California: Sage Publications.

Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.

Zhang, S. (1987). Cognitive complexity and written production in English as a second language. *Language Learning*, *37*(4), 469–481. doi: https://doi.org/10.1111/j.1467-1770.1987.tb00580.x

Zhao, J. (2017). Native speaker advantage in academic writing? Conjunctive realizations in EAP writing by four groups of writers. *Ampersand*, *4*, 47–57. doi: https://doi.org/10.1016/j.amper.2017.07.001

Zheng, Y. (2012). Exploring long-term productive vocabulary development in an EFL context: the role of motivation. *System*, *40*(1), 104–119. doi: https://doi.org/10.1016/j.system.2012.01.007

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, *56*, 40–53. doi: https://doi.org/10.1016/j.system.2015.11.007

Zhou, A., Bush, M., Gentil, G., Eouanzoui, K., & Cumming, A. (2006). Students' goals for ESL and university courses. In A. Cumming (Ed.), *Goals for academic writing: ESL students and their instructors* (pp. 29–49). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.15.05zho

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *11*(4), 307–313. doi: https://doi.org/10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (1990a). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17. doi: https://doi.org/10.1207/s15326985ep2501_2

Zimmerman, B. J. (1990b). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, *2*(2), 173–201. doi: http://dx.doi.org/10.1007/BF01322178

Zimmerman, B. J. (1995). Self-regulation involves more than metacognition: A social cognitive perspective. *Educational Psychologist*, *30*(4), 217–221. doi: https://doi.org/10.1207/s15326985ep3004_8

Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1–19). New York: Guilford Press.

Zimmerman, B. J. (2000). Attainment of self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds), *Self-regulation: Theory, research, and applications* (pp. 13–39). Orlando, FL: Academic Press.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*(4), 845–862. doi: https://doi.org/10.3102/00028312031004845

Zimmerman, B. J., & Kitsantas, A. (2007). A writer's discipline: the development of self-regulatory skill. In P. Boscolo & S. Hidi (Eds.), *Writing and motivation* (pp. 51–69). Oxford: Elsevier. https://doi.org/10.1163/9781849508216_005

Appendix 1. Writing Prompts (Pilot Study)

1.	Science	Some people believe that computers are more a hindrance than a help in today’s world. Others feel they are such indispensable tools that they would not be able to live or work without them. In what ways are computers a hindrance? What is your opinion?
2.	Work	Some people feel that certain workers like nurses, doctors and teachers are undervalued and should be paid more, especially when other people like film actors or company bosses are paid huge sums of money that are out of proportion to the importance of the work that they do. How far do you agree? What criteria should be used to decide how much people are paid?
3.	Media	Many newspapers and magazines feature stories about the private lives of famous people. We know what they eat, where they buy their clothes and who they love. We also often see pictures of them in private situations. Is it appropriate for a magazine or newspaper to give this kind of private information about people?
4.	Society	People in all modern societies use drugs, but today's youth are experimenting with both legal and illegal drugs, and at an increasingly early age. Some sociologists claim that parents and other members of society often set a bad example. Discuss the causes and some effects of widespread drug use by young people in modern day society. Make any recommendations you feel are necessary to help fight youth drug abuse.

Appendix 2. Writing Prompts (Main Study)

Code	Data point	Topic	Writing prompt
WP1	1.	Education	The idea of having a single career is becoming an old fashioned one. The new fashion will be to have several careers or ways of earning money and further education will be something that continues throughout life.
WP2	2.	Education	In many countries schools have severe problems with student behaviour. What do you think are the causes of this? What solutions can you suggest?
CWP1		Language learning	You cannot learn a foreign language unless you visit the country where the language is spoken. To what extent do you agree?
WP3	4.	The media	News editors decide what to broadcast on television and what to print in newspapers. What factors do you think influence these decisions? Do we become used to bad news? Would it be better if more good news was reported?
WP4	5.	The media	Compare the advantages and disadvantages of three of the following as media for communicating information. State which you consider to be the most effective. Comics, books, radio, television, film, theatre.
CWP2		Language learning	A native language teacher is always better than a non-native one. To what extent do you agree?
WP5	7.	Work	As most people spend a major part of their adult life at work, job satisfaction is an important element of individual wellbeing. What factors contribute to job satisfaction? How realistic is the expectation of job satisfaction for all workers?
WP6	8.	Work	In many countries children are engaged in some kind of paid work. Some people regard this as completely wrong, while others consider it as valuable work experience, important for learning and taking responsibility.
CWP3		Language learning	Obtaining a degree (from a Hungarian university) should not be dependent on passing a foreign language exam. To what extent do you agree?
WP7	10.	Politics	Should wealthy nations be required to share their wealth among poorer nations by providing such things as food and education? Or is it the responsibility of the governments of poorer nations to look after their citizens themselves?
WP8	11.	Healthcare	"Prevention is better than cure." Out of a country's health budget, a large proportion should be diverted from treatment to spending on health education and preventative measures.
CWP4		Language learning	The best way to learn a language is by speaking and not by learning grammar. To what extent do you agree?
WP9	13.	Travel	In some countries young people are encouraged to work or travel for a year between finishing high school and starting university studies. Discuss the advantages and disadvantages for young people who decide to do this.

WP10	14.	Crime & the law	Without capital punishment (the death penalty) our lives are less secure and crimes of violence increase. Capital punishment in essential to control violence in society.
CWP5		Language learning	The older you get, the most difficult it is to learn a foreign language. To what extent do you agree?
WP11	16.	Social tensions	The position of women in society has changed markedly in the last twenty years. Many of the problems young people now experience, such as juvenile delinquency, arise from the fact that many married women now work and are not at home to care for their children.
WP12	17.	Science & technology	When a country develops its technology, the traditional skills and ways of life die out. It is pointless to try and keep them alive.
CWP6		Language learning	It is better if you speak one foreign language at an advanced level than speaking more (2-3) languages at medium levels. To what extent do you agree?
WP13	19.	Children & the family	"Fatherhood ought to be emphasised as much as motherhood. The idea that women are solely responsible for deciding whether or not to have babies leads on to the idea that they are also responsible for bringing the children up."
WP14	20.	The arts	There are many different types of music in the world today. Why do we need music? Is the traditional music of a country more important than the International music that is heard everywhere nowadays?
CWP7		Language learning	The writing prompt was different for the four participants.
WP15	22.	Men & women	Universities should accept equal numbers of male and female students in every subject.
WP16	23.	Global problems	Popular events like the football World Cup and other international sporting occasions are essential in easing international tensions and releasing patriotic emotions in a safe way.

Note. WP = writing prompt; CWP = controlled writing prompt; Controlled data points are highlighted. Controlled writing prompts were composed at different data points by the four participants in this study.

Appendix 3. Interview Questions (Initial Interview)

1. What kinds of writing in English do you expect to do in your future studies at university?
2. What goals do you have for improving your writing for your future studies at university?
3. What kinds of writing in English do you expect to do in your future career or occupation?
4. What goals do you have for improving your writing?
5. Are there specific types of writing that you are trying to improve?
6. What is your usual method of writing? So, what do you do first, second?
7. What is your usual method of checking or rewriting your writings or compositions?
8. Who do you write for? So, do you have any audience?
9. Where do you get your information for writing?
10. Are there any special types of writing that you want to do?
11. Are there any special topics that you want to write about?
12. Are you trying to improve your grammar and vocabulary?
13. How do your teachers help you to write?
14. What tools do you use to help you write? So, dictionaries, books, computer softwares.
15. Are you trying to improve or change the way you use these?
16. Does reading influence how you write?
17. How do you feel when you write in English?
18. Do you have a specific identity or “voice” when you write in English?

Appendix 4. Retrospective Interview Questions

Self-observation and self-evaluation processes

1. Did you try out any new words or phrases in this essay that you haven't used much before?
2. Were there any points when you were searching for a particular word? If yes, what word(s)?
3. Did you try out any new grammatical structures in this essay that you haven't used much before?
4. How did you compose this essay?
5. Can you tell me how you approached this piece of writing?
6. Did you have any difficulties in structuring what you were writing? Why?
7. What kind of genre were your trying to write in?
8. Was it easy for you to think of ideas for the topic? Why?
9. Did you feel confident while you were writing? Why?
10. Are you happy with the level of vocabulary? Why?
11. Are you happy with the level of grammar? Why?
12. Compared to the previous pieces of writing do you think it was an improvement? Why?

Goal-setting

13. Do you feel that this piece of writing is a good representation of you view on this topic? If, not. What would you change?
14. Do you think this piece of writing is a good representation of how well you can write? If, not. What would you change?

Appendix 5. Interview Questions (Final Interview)

Writing in general

1. Do you like writing in English?
2. Do you like writing in Hungarian?
3. Do you find it easy to write in English?
4 Do you think writing is an important element of language learning?
5 Do you think writing in English is important in your daily life?

Writing prompts

6 What difficulties did you have during the completion of these writing tasks?
7 Do you think 40-minute time frame is enough to complete these writing tasks?
8 Would you be able to write better essays if you had different writing prompts? What topic would you write about if you had the chance to choose?
9 Do you feel any of these writing prompts are a bit unnatural or forced?

Writing and motivation

10 Do you think your level of motivation had an effect on the quality of your essays you were writing?
11 Do you think if you had been more motivated you could have written better essays?
12 Do you think the obligation to write had an effect on the quality of your essays you were writing?
13 Do you have a different attitude to writing after completing the six writing tasks?

Improvement

14 Do you think your writing improved over the last six months? If so, how?
15 Do you think you have become a better second language (L2) writer?
16 Do you think the completion of these writing tasks helped you become a better L2 writer?
17 Do you like trying out new words? Do you consider yourself as an experimentalist? Do you like taking risks at exams?
18 Do you avoid using new words which you are not sure about their meaning, spelling or form?
19 Do you tend to avoid using new grammatical structures which you are not sure about?

Rhetoric and genres

20 Do you like writing argumentative essays?
21 Do you feel confident in writing essays?
22 Do you think you could improve in using the discourse markers (linkers and fillers) of the essay?

Composing processes

23 Did you have problems with composing (planning, drafting, editing, and revising a text) before completing the six writing tasks?
24 Do you still have problems with composing after completing the six writing tasks?
25 Did you reread your essays before submitting them?

Ideas and knowledge

26 Is it easy for you to think of ideas for an IELTS-type writing prompt?
27 Was it easy for you to think of ideas for these six writing tasks?

	Affective states
28	How do you usually feel when you are writing in English?
29	How did you feel when you were composing these six writing tasks?
30	Do you feel different (e.g. less/more stressed) during writing after completing these six writing tasks?
31	Does it make you feel nervous if you see that other students are writing more words and more fluently?
	Language and transfer
32	Do you think you have improved in transforming your knowledge into writing?
	Ideal essay
33	How would you describe your ideal essay?
34	Do you think you could write a better essay if you knew more words or grammatical structures?
	Complexity in writing
35	You must have heard of complexity in writing but what does that mean to you?
36	Do you think if an essay is grammatically complex it is always better than a less complex essay?
37	Do you think your essays are comparable? How would you compare them?
	Discussing assessments
38	Do you think these assessments are an accurate reflection of how well you can write in English?
39	The scores for your essays show no/little/a lot of improvement over time – do you think you writing has improved as indicated here?

Appendix 6. Example of Participants' Essay

(Participant: Dalma; Essay 9)

Name:

Lancaster University

Date: 12. 19.

WRITING TASK 4

You should spend about 40 minutes on this task.

Write about the following topic:

> **The best way to learn a language is by speaking and not by learning grammar. To what extent do you agree?**

You should give reasons, any relevant examples from your experience or knowledge for your answer.

You should write at least 250 words.

Nowdays one of the most important question is, how ~~can~~ could people acquire a foreign language in an easier way. My opinion is that learn a language by speaking and not by learning grammar has more benefits than somebody knows the grammar better than
44
speaking.

At first I am going to introduce my boyfriend's sister's experiance. She could not speak English in the ~~right~~ correct way, she ~~could not~~ did not know the grammar as well. But she has never had any problems with her personality, she was not shy in speaking with foreign people. When she moved to ~~the~~ USA, she ~~could~~ ~~can put into~~ could get a no job, she was brave to not take into consideration not have has the right grammar. Less than a half year she already learnt in English. Her situation can provide that, if somebody is in a society, where the others can speak in the correct form, can manage to learn a foreign
51

language easier ~~Bb~~.

At the previous example I showed a lucky way to gain skills learning languages. ~~B~~ But this theme is not as light as Dori's case. If somebody 80 Can speak a language very well, but does not care the grammer, it can ~~to~~ lead ~~to~~ to ~~s~~ some problems. These people show 200 out from the others, and ~~it~~ it will be clear, they can not speak in the suitable form. Based on this opinion, these people could not get jobs, where the officialy and correct language is inevitable.

All in all, it is an easier way to learn languge by speaking, but people ~~to~~ should not ~~an~~ forget the grammar neither, because for example they can lose a better ~~the~~

277

opportunity of job.

Appendix 7. Example of Manual Coding 1

(Participant: Dalma; Essay 9)

9			
The best way to learn a language is by speaking and not by learning grammar. To what extent do you agree?			
Total words: 267		DC/C: 0.3225	WP4

		Type	FV	DC
1	Nowadays one of the most important question is, how could people acquire a foreign language in an easier way.	Cx	2	1
2	My opinion is that learn a language by speaking and not by learning grammar has more benefits than somebody knows the grammar better than speaking.	Cx	5	3
3	At first I am going to introduce my boyfriend's sister's experience.	Si	6	
4	She could not speak English in the correct way, (and) she did not know the grammar as well.	Co	8	
5	But she has never had any problems with her personality, (and) she was not shy in speaking with foreign people.	Co	10	
6	When she moved to USA, she could get a job, (and) she was brave to not take into consideration not have the right grammar.	CoCx	13	4
7	Less than a half year she has already learnt in English.	Si	14	
8	Her situation can provide that, if somebody is in a society, where the others can speak in the correct form, can manage to learn a foreign language easier.	Cx	18	6
9	At the previous example I showed a lucky way to gain skills learning languages.	Si	19	
10	But this theme is not as light as Dóri's case.	Si	20	
11	If somebody can speak a language very well, but does not care the grammar, it can lead to some problems.	CoCx	23	7
12	These people show out from the others, and it will be clear, (that) they can not speak in the suitable form.	CoCx	26	8
13	Based on this opinion, these people could not get jobs, where the official and correct language is inevitable.	Cx	28	9
14	All in all, it is an easier way to learn language by speaking, but people should not forget the grammar neither, because for example they can lose a better opportunity of job.	CoCx	31	10

Note. FV = Finite verb; DC = Dependent clause; Si = Simple sentence; Co = Compound sentence; Cx = Complex sentence; CoCx = Compound-complex sentence.

Appendix 8. Example of Manual Coding 2

(Participant: Dalma; Essay 9) (academic genre)

9			
The best way to learn a language is by speaking and not by learning grammar. To what extent do you agree?			
Total words: 267	DC/C: 0.3225		WP4

1	Nowadays one of the most important question is, how could people acquire a foreign language in an easier way.
2	My opinion is that learn a language by speaking and not by learning grammar has more benefits than somebody knows the grammar better than speaking.
3	At first I am going to introduce my boyfriend's sister's experience.
4	She could not speak English in the correct way, (and) she did not know the grammar as well.
5	But she has never had any problems with her personality, (and) she was not shy in speaking with foreign people.
6	When she moved to USA, she could get a job, (and) she was brave to not take into consideration not have the right grammar.
7	Less than a half year she has already learnt in English.
8	Her situation can provide that, if somebody is in a society, where the others can speak in the correct form, can manage to learn a foreign language easier.
9	At the previous example I showed a lucky way to gain skills learning languages.
10	But this theme is not as light as Dóri's case.
11	If somebody can speak a language very well, but does not care the grammar, it can lead to some problems.
12	These people show out from the others, and it will be clear, (that) they can not speak in the suitable form.
13	Based on this opinion, these people could not get jobs, where the official and correct language is inevitable.
14	All in all, it is an easier way to learn language by speaking, but people should not forget the grammar neither, because for example they can lose a better opportunity of job.

	Total
Infinitive clause (InfC)	5
Conditional clause (ConC)	2
Relative clause (RelC)	2

Appendix 9. Example of Manual Coding 3

(Participant: Dalma; Essay 9) (accuracy)

9
The best way to learn a language is by speaking and not by learning grammar. To what extent do you agree?
Total clauses: 30

1	Nowadays one of the most important question is,	1	GNN
2	how could people acquire a foreign language in an easier way.	2	WO
3	My opinion is		
4	that learn a language by speaking and not by learning grammar has more benefits	3	FM
5	than somebody knows the grammar better than speaking.		
6	At first I am going to introduce my boyfriend's sister's experiance.	4	FS
7	She could not speak English in the correct way,		
8	(and) she did not know the grammar as well.		
9	But she has never had any problems with her personality,		
10	(and) she was not shy in speaking with foreign people.		
11	When she moved to USA,		
12	she could get a job,		
13	(and) she was brave to not take into consideration not have the right grammar.	5	WO FM
14	Less than a half year she has already learnt in English.		
15	Her situation can provide that,	6	LS
16	if somebody is in a society,		
17	can manage to learn a foreign language easier.		
18	where the others can speak in the correct form,		
19	At the previous example I showed a lucky way to gain skills learning languages.		
20	But this theme is not as light as Dóri's case.		
21	If somebody can speak a language very well,		
22	but does not care (about) the grammar,	7	WM
23	it can lead to some problems.		
24	These people show out from the others,		
25	and it will be clear,		
26	(that) they can not speak in the suitable form.		
27	Based on this opinion, these people could not get jobs,		
28	where the official and correct language is inevitable.		
29	All in all, it is an easier way to learn (a) languge by speaking,	8	GA FS
30	but people should not forget the grammar neither,	9	XCONJCO
31	because for example they can lose a better opportunity of job.		

Note. FM = morphology; GA = article; XCONJCO = conjunction complementation; WO = word order; FS = spelling; GNN = noun, number; LS = lexical single; WM = word missing.