A longitudinal multidimensional analysis of L2 academic writing

Submitted by Sangeun Kim, to the University of Exeter as a thesis

for the degree of Doctor of Philosophy in Education, May 2024.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis/dissertation may be published without proper acknowledgement. I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Abstract

This longitudinal study investigates the development of English language writing skills in adult learners within an ESL context. Focusing on syntactic complexity (SC) and syntactic sophistication in verb-argument constructions (VACs), the research examines linguistic variations in academic writing across five semesters. Employing corpus linguistics methods, including multidimensional (MD) analysis and collostructional analysis, the study explores a usage-based perspective of syntactic development in learner writing development.

The research findings suggest the nature of L2 academic writing development is multi-dimensional, showcasing increased salience of phrasal features associated with literate and informational writing functions over time. Notable nonlinear patterns are observed across functional dimensions. Analysing constructions associated with stance functions reveals shifts toward syntactically sophisticated writing in verb-argument constructions.

This study contributes empirical evidence of linguistic development, highlighting form-function mapping as crucial in identifying register-specific language use in L2 writing. While statistically non-significant, increases in noun- and adjective-clause constructions were observed, emphasising the need for more hybrid methodology by incorporating qualitative inspection to complement quantitative findings.

The research underlines the significance of understanding individual and situational influences on syntactic variation in L2 writing, providing implications for developing syllabi and materials tailored to learners' communicative goals. Limitations, such as the lack of individual trajectory tracing and granularity in linguistic measures, suggest avenues for further research. Overall, this dissertation offers essential insights into L2 writing development, contributing to a deeper understanding of L2 adult learners' language use and its implications for pedagogy.

1. Table of Contents

2. Tables

3. Figures

Definitions

L1: first language; used to refer to speakers of English as a first language

L2: second language; used to refer to speakers of English as a second language

Abbreviations

AI: artificial intelligence

AIC: Akaike Information Criterion

AMOD: adjectival modifiers per nominal

ATHC: adjectives controlling that-complement clauses

BAWE: the British Academic Written English corpus

CEFR: the Common European Framework of Reference

Chat-GPT: Chat Generative Pre-trained Transformer

CNC: complex noun phrases per clause

ECDF: empirical cumulative distribution function

EFA: exploratory factor analysis

EFL: English as a foreign language

ESL: English as a second language

FL: formulaic language

ICNALE: the International Corpus Network of Asian Learners of English (Ishikawa, 2019)

IEP: Intensive English Programme

KMO: Kaiser-Meyer-Olkin

L2SCA: the L2 Syntactic Complexity Analyzer (Lu, 2010, 2011)

LOB corpus: Lancaster-Oslo-Bergen Corpus of British English

LOO-CV: leave-one-out cross-validation

MAT: Multidimensional analysis tagger (Nini, 2019)

MCMC: Markov Chain Monte Carlo

MD analysis, MDA: multidimensional analysis (Biber, 1988)

MEM: mixed-effects model

Michigan EPT: the Michigan English Placement Test

MICUSP: The Michigan Corpus of Upper-Level Student Papers

MLC: mean clause length

MLE: maximum likelihood estimation

NTHC: nouns controlling that-complement clauses

PCA: Principal component analysis

PELIC: The University of Pittsburgh English Language Institute Corpus (Juffs et al., 2020)

POS: part-of-speech

Prep: prepositional dependents per nominal

SC: syntactic complexity

SPIN: split infinitive

TAASSC: Tool for the automatic analysis of syntactic sophistication and complexity (Kyle, 2016)

THAC: that-adjective clause

THATD: subordinator *that* deletion

THNC: that-noun clause

THVC: that-verb clause

TO: to-infinitive

TOBJ: that-relative clauses on object position

TOEFL: Test of English as a Foreign Language

TTR: type-token ratio

VAC: verb-argument construction

Chapter 1 Introduction

The personal motivation for this research is rooted in my interest in learning writing in an English as a foreign language (EFL) school context. This study aims to analyse how English writing evolves under guided instruction in an English as a second language (ESL) context, where students are given significantly more target language input and motivation for communication is likely to be higher than in EFL contexts, as English is the genuine communication tool. Students’ written texts in an intensive English language programme provide an advantage in tracing English development as they have been produced with genuine motivation for learning the language and through ample classroom language input. The scope of this thesis is confined to human writing before the universal use of Chat Generative Pre-trained Transformer (Chat-GPT) and similarly powered artificial intelligence (AI) tools. Even in the time of AI-assisted learning and teaching writing becoming ever more possible, with the advent of Chat-GPT and similar machine learning algorithms, the importance of using English as a core of academic skills has not diminished, rather, it has grown. AI tools can assist with grammar and mechanics, but critical thinking, analysis, and the ability to construct clear and concise arguments remain essential (Amante-Nochefranca et al., 2023). English, as a medium for clear and nuanced communication, remains a cornerstone of developing these abilities. In light of this, this thesis aims to follow the human writers’ production in a shared educational context over time.

The development of academic writing skills is of significant educational concern, particularly at the university level, where assignment writing has been recognised as the core of university education assessment (e.g., Zhu 2004; Nesi & Gardner, 2006). However, for students from other language backgrounds, academic writing in English may provide additional challenges that are related to language skills rather than subject knowledge. Empirical studies suggest that target language proficiency is a significant predictor of academic success in tertiary education (e.g., Trenkic & Warmington, 2019). Moreover, studies suggest that using the linguistic form appropriate for academic writing strongly predicts L2 academic writing quality (Biber & Gray, 2013, p.10). In light of this, linguistic features typically found in academic discourses have been treated as development markers in academic writing contexts (Gardner, Nesi & Biber, 2019, p.647). Thus it is of research value to learn about how L2 students' language use becomes appropriate to academic writing in university education.

There are two main linguistic constructs of interest in this thesis. The first construct, syntactic complexity (SC), has often been conceptualised as formal or structural quality observed within sentence level. As syntactic patterns are a foundational element of linguistic production, supporting writing development by accommodating the functional needs to deliver complex ideas (Ortega, 2015, p.87). In light of this, SC has been considered a potential index for writing development, and therefore, it has been extensively studied in L2 English writing development.

Recent research has indicated that the phrasal level of SC has a significant explanatory power to explain syntactic development in English written registers (e.g., Biber et al., 2011; Kyle & Crossley, 2018). Building upon this, this study takes a hypothesis-driven approach by adopting a usage-based perspective to explore SC in L2 writing development. In a usage-based approach, language consists of constructions as form-meaning pairings, each construction defined by its use in communicative situational contexts (Ellis, 2002). Researchers have noted the usefulness of exploring the pairings of grammatical forms and syntactic functions performed by the forms, such as subordination or prepositions because of their potential pedagogical value (e.g., Grant & Ginther, 2000; Biber et al., 2021a). Drawing on this notion of form-function pairing, SC is operationalised as a measure of frequencies of constructions associated with syntactic complexity. The aim is to find evidence of more frequent use of SC constructions associated with written registers in L2 writing over time.

The second construct of focus is syntactic sophistication in verb-argument constructions (VACs); certain pair of verb-argument constructions that are likely to be used more frequently in a specific register, and therefore more strongly associated with the register (Kyle & Crossley, 2017). As seen from this operational definition of syntactic sophistication, it is conceived as a relative construct to a specific register in this study.

The key communicative function explored in terms of its sophistication in this thesis is the stance function. Stance is a communicative function performed by the language of evaluation, and it has been extensively researched under various frames and terms. This thesis focus is confined to attitudinal or personal stance conveyed by a specific construction (a headword controlling *that* clause), which has been found to be a unique marker of student register (Gardner et al., 2019).

As Biber et al. (2004) noted, more frequently used combinations of lexico-grammatical patterns are likely to be learned and used in the early stages of language learning. Drawing on this observation, it was hypothesised that the formulaic (conventional and fixed) pattern of a specific construction may decrease over time and become replaced with more infrequent or sophisticated associations, using more variety of headwords within the construction appropriate to the academic written discourses.

Exploring these two constructs in learner corpora enables hypotheses which draw on usage-based approaches to be tested; that more exposure leads to the acquisition and therefore will be used at earlier developmental stages. Students' texts enable us to investigate how student writing progresses to show linguistic features that are more elaborated and complex (Staples et al., 2016, p.152). Considering that L2 students in English medium higher education contexts often need language support for their writing, the writing data collected in such a language learning context can be used as a proxy of these students' writing development in the target language. In this regard, an intensive English programme (IEP) is committed to developing such language performance and therefore is a useful addition to our understanding of whether the L2 students are developing the necessary language use appropriately in a university education context. For example, the patterns of frequency of the linguistic features associated with academic writing can provide evidence of a gap between the writer's production and what is expected of them by an academic discourse community (Hardy & Römer, 2013, p.184).

Many of the previous studies exploring syntactic development in L2 writing have used cross-sectional data while longitudinal linguistic analysis tended to adopt qualitative methods. As Rankin notes, corpus linguistics methods can be a practical methodological resource in second language research (Rankin, 2015, p.234). However, only a few studies used large corpora to explore learner-written texts longitudinally (e.g. Crosthwaite, 2016; Römer & Berger, 2019; Gray et al., 2019). Longitudinal studies with a larger size of data may contribute to advancing our knowledge of L2 development.

To fill this gap, the series of longitudinal analyses discussed in this thesis explore syntactic variation in academic English writing of adult learners of English as a second or another language (L2) in a non-experimental setting. The current study uses a learner corpus of academic writing, collected in an intensive English programme at a U.S. university throughout at least three consecutive academic semesters. The research aims to find differences in a grammatical form associated with syntactic complexity between drafts on personally chosen topics. The corpus linguistic methods employed include a multidimensional (MD) analysis (Biber, 1988) and collostruction analysis (Gries et al., 2005), to capture the patterns of syntactic forms indicative of communicative functions commonly associated with academic written registers in the L2 learners' writing.

This paper consists of seven chapters. Following this introduction, Chapter 2 reviews the literature regarding L2 linguistic development and some significant findings concerning academic written register variations in both L1 and L2 speakers of English. The review intends to provide a general overview of the previous studies and define the scope of the research interest of this thesis, by including summaries of the previous methodological approaches and research findings of the primary L2 writing data used for this thesis. Chapter 3 depicts the research design proposed in the study, consisting of six sub-sections. The philosophical assumptions and methodologies to address research questions are introduced and justified. The primary research methods comprise MD analysis, collostructional analysis, and mixed-effect modelling. The linguistic measures relevant to answering the research questions are identified. The procedures to be followed in selecting and analysing data are also stated. Chapters 4, 5, and 6 provide details on each of the literature reviews of more directly relevant previous studies, the methods, the findings resulting from the analysis and the discussion regarding the research questions. Finally, Chapter 7 summarises the study's findings and how they can be interpreted focusing on both the pedagogical and research significance of the study.

Chapter 2 Literature Review

2.1 Introduction

This chapter aims to assess the current state of understanding in grammatical complexity development in L2 academic writing. It reviews previous literature under two themes of research interests. The first theme of this review is the notions of *register* applied to corpus linguistic studies exploring L2 writing development. This approach treats the collection of written text (corpus) created in a specific communicative situation (shared communicative context) as an instance of an L2 register. The register framework becomes the lens through which we understand the kind of written data being analysed. In other words, analysing the linguistic patterns in L2 writing enables us to view how learners use their L2 language in similar contexts. This chapter discusses the general points of these frameworks to justify r*egister* as a main framework of the analysis adopted in this thesis. The first section of the literature review focuses on the empirical findings drawing on the studies exploring L2 learner registers.

Section 2.2 introduces *linguistic development* and addresses general issues in exploring learner corpora, focusing on academic register variation. It also introduces some of the studies using the MDA method, which illustrates some of the significant findings of learner registers. The subsequent sub-section presents a form-function mapping (usage-based perspective) on SC, which the current study has used as an overarching framework to explore syntactic development in academic writing produced by writers of English as an additional language (L2) (2.3). These discussions include reviews of relevant empirical research into student academic writing.

To investigate development in longitudinal data, it is crucial to establish a shared understanding of which linguistic constructs constitute typical academic writing, as this perception may vary among individuals. Additionally, agreement on what serves as a representative example of good writing is essential. Various study designs can be considered when examining linguistic use and its relation to writing development. These include comparing low- and high-scoring writing, analysing novice and professional writing, or studying individual writing over an extended period. These approaches have assumptions regarding the criteria and individuals involved in assessing writing quality, which will be further explored in this chapter, specifically to map the PELIC corpus, the main written data of this thesis, in the context of defining learner corpora types such as cross-sectional, level-stratified, and longitudinal corpora.

To link the linguistic variable of focus with developmental perspectives in research, I took theoretical and operational definitions of the linguistic construct drawing on a usage-based perspective. Regarding this, the second theme of this literature review involves a more specific interest in the focal linguistic construct of this thesis: grammatical complexity.

The second part of this literature review (Section 2.3) starts with theoretical discussions and empirical findings from the previous studies of learner writing with a particular interest in syntactic complexity. This discussion leads to discussions on grammatical complexity, which shares an interest in linguistic complexity in written discourses but is distinguished in terms of at least its consideration of the register parameter. This thesis also explores the aspect of syntactic sophistication, which is considered a sub-component of linguistic complexity. Chapter 5 explores the aspect of syntactic sophistication drawing on one of the association measures, the collostructional analysis method (Gries et al., 2005) to analyse a lexicogrammatical construction, head verbs controlling 'that' clauses. For the purpose of discussing research that makes a closer connection between each linguistic variable of focus and the methods in L2 writing developmental research, each of Chapters 4, 5 and 6 provides a separate literature review more relevant to each key analysis method and linguistic variable discussed in each chapter.

Section 2.3 reviews theoretical discussions and empirical studies on syntactic development to establish a solid foundation for the current study. The primary focus of this review is on the theoretical and operational definitions of SC and the key issues relevant to exploring this construct in English writing development research. Recently, phrasal complexity features, such as nouns, adjectives or prepositions modifying noun phrases, have garnered substantial attention within this domain. This incorporation of phrasal indices indicates increased attention to capturing register-specific SC, providing more nuanced and diverse measures. It examines the assumptions of SC in these studies by examining this shift of focus and operation of SC in SC research.

Furthermore, it delves into SC through the lens of usage-based perspectives. The form-function mapping is one of the fundamental assumptions of a usage-based perspective. It has long been recognised for its potential to deconstruct meaningful constructions and their subsequent pedagogical value. These constructions offer valuable insights into how language forms are used and learned as mental representations of communicative functions. On the one hand, the inclusion of a specific construction in the linguistic investigation is justified by its frequency, highlighting its significance for learning purposes. For example, nominal phrases with prepositional or noun modifications have been considered characteristic of academic written discourse for their relative prevalence in these registers.

Another focus of the review is on constructions associated with stance communicative functions. One of the stance devices includes Verb-Argument constructions, with some studies emphasising their usefulness to mark L2 proficiency. Although these constructions may be less associated with academic written registers overall, some stance markers with relatively higher frequency in academic written registers make them worthy of research and pedagogical attention (e.g., nouns and adjectives controlling that-clause complement). Consequently, these constructions become the focal point due to their direct relevance to learners' language development for academic purposes. The discussions include implications of empirical findings regarding these constructions in expert and learner academic writing, discussing their significance for understanding linguistic development.

The importance of, and the reasons for the limited use of, longitudinal studies will be highlighted. The conclusion will be reached that there is a need for more longitudinal research (2.4).

2.2 Theorising and Researching Student Academic Writing

This section establishes a link between linguistic distributions and academic writing development. First, it distinguishes the terms genre, registers, text type and task in the realm of academic discourse to introduce the notion of an L2 learner register as a distinguished register of focus in this study.

Next, the similarities between research interests and findings in L1 and L2 English writing development are scrutinized. The common challenges in transitioning from spoken to written discourse and the differing linguistic resources required for spoken and written language are discussed for L1 and L2 learning. The significance of formulaic expressions in both L1 and L2 student academic registers, signifying community membership and proficiency, is discussed. Then, it explores the significance of syntactic features within the framework of register, particularly in the context of academic writing. The focus is on the direct observation of linguistic features conveying essential functions for academic writing, such as nominalization. The focus extends to examining patterns of syntactic constructions in academic written registers. These studies emphasize the importance of phrasal complexity features in L2 academic writing development (Taguchi et al., 2013; Parkinson & Musgrave, 2014; Biber, Gray & Staples, 2014).

Finally, it synthesises the discussions to suggest hypotheses related to L2 writing development. The first hypothesis explores the similarity between L2 and L1 linguistic development patterns in academic writing. The second hypothesis focuses on the development of L2 syntactic complexity in academic writing, transitioning from clausal to phrasal structures. The hypothesis suggests that writing development starts from spoken register features and progresses to written registers.

2.2.1 Conceptualising L2 Writing Development

2.2.1.1 A Register perspective for researching language development

Establishing a clear relationship between linguistic distributions and academic writing development is challenging due to the complex nature of the writing process (Jarvis et al., 2003). Moreover, the concept of ‘development’ is relative to the context and purpose of making such a qualitative assessment, and therefore, it is essential that this notion is contextualised in terms of the focus language uses in reference to the target language use. This subsection discusses a theoretical ground for defining the target language use and main language data with two-fold aims. First, it establishes the status of the L2 longitudinal written data used in this thesis on the broader scope of research found in previous research in order to clarify how grammatical complexity can inform L2 development in writing. Second, it discusses different corpus data characteristics and research methods used to explore them to lay the ground for justifying the research design of this thesis, which will be detailed in Chapter 3.

The main written data of this thesis, the PELIC corpus, is a unique variety of L2 academic written *registers*. The term *register* is used in studies that describe the surface linguistic characteristics of a group of texts, such as studies of hedging devices in academic prose (e.g., Hyland, 1994; 1996). It often refers to a particular configuration of linguistic use functionally associated with particular contextual or situational factors (e.g., Lee, 2001, p.42; Biber, Conrad, Reppen, Byrd & Helt, 2002). The focal linguistic features could be based on salient patterns observed in a corpus, such as the complex noun phrase structures typical of scientific prose (Halliday & Matthiessen, 2014).

What is particularly relevant to the L2 learner corpora research similar to this study is the consideration that extra-textual variables may play a significant role in the variations of syntactic complexity in student writing. For example, topics or task types of student writing reflect “a communicative purpose” inherent in the particular writing product, which is an influential factor in language use (Ortega, 2015; Halliday & Matthiessen, 2014). Based on the theory of systemic functional linguistics (Halliday and Matthiessen 2014), Ortega argued that the complexity of L2 production is influenced by the communicative purposes of the task the learner aims to address (2003, 2015).

One conceivable assumption within the usage-based theory is that second language writing can be assessed in reference to a corpus that consists of a target register, such as academic written registers collected from university classrooms or published papers. In addition, as previously discussed, the notion of a form-function mapping underlying a usage-based approach indicates that assessing linguistic variation in L2 data can be guided by studying language forms that are mapped onto communicative functions in the target communicative context. This brings in the notion of genre, as a standardized communicative event mutually understood by participants (Swales, 1990). Genre is often referred to as social or rhetorical perspectives in these studies. Studies on academic discourse often describe the rhetorical structure of academic texts and the way the practices of researchers in particular discourse communities shape the conventions of academic *genres*, predominantly focusing on scientific or medical writing (Swales, 1990).

From this genre perspective, written texts produced in an L2 English classroom are an academically and practically important situational research context contributing the linguistic data produced in an instructed L2 learning setting. Exploring linguistic constructions through an exploratory analysis approach can reveal valuable insights into their associated communicative functions within the linguistic data. Indeed, teacher intervention can be genuinely effective when properly incorporated in process writing classrooms (e.g., Shintani, Ellis & Suzuki, 2014).

There has been another approach to categorizing texts: *text types* (Biber, 1995). While genre considers social conventions, such as writing purpose, format, and the context in which texts are typically used (as in Biber's 1988 work), text types take a different approach. This approach prioritises linguistic features. Biber (1995) first identified text types based on language characteristics to interpret the function of these text types based on the identified linguistic patterns. This allows for a more nuanced understanding of how similar texts are, even if they seem to belong to different genres on the surface.

Table 2-1. English text typology (adapted from Biber, 1989, pp.20-22)

	text type	representative genre type
Text type 1	Intimate interpersonal interaction	telephone conversations (personal friends)
Text type 2	Informational interaction	face-to-face conversations, interviews, telephone conversations (business associate), spontaneous speeches, personal letters
Text type 3	‘Scientific’ exposition	academic prose
Text type 4	Learned exposition	official documents, press reviews
Text type 5	Imaginative narrative	fictions (romance general, mystery, adventure), prepared speeches
Text type 6	General narrative exposition	press reportage, press editorials, fiction (general), biographies, humour, press

		revies, religion, nonsports broadcasts, hobbies
Text type 7	Situated reportage	sports broadcasts
Text type 8	Involved persuasion	interviews, spontaneous speeches, professional letters

The representative genres for each text type in Table 2-1 are chosen when the texts of a genre are identified in the core cluster and when more than 40% of the genre occurs in a single cluster. This is the representation of the prototypical characteristics of a text type, rather than being exhaustive of the text type characteristics as discussed in Biber (1989). For that purpose, Table 2-1 shows the genres for each text type only when satisfying both conditions, rather than showing all the identified genres associated with each text type in Biber (1989). Moreover, as the representative genres for text type have been identified through a cluster analysis, some texts fall in between clusters, therefore appearing in more than one text type (e.g., interview in text types 3 and 4). Biber (1992) used the LOB corpus, distinguishing the 23 genre classifications of this corpus from the analysed text types. It is especially helpful in understanding the register variations analysed using Biber's dimensions (1988), and as the pilot study discussed in Section 4.4., so this will be revisited there.

I have discussed three different approaches to writing developmental research, which all take into account a situational factor interacting with the textual dimension. Perhaps consideration of situational factors in the terms genre, register, and text type leads to often unclear, overlapping, or different operationalisation by different linguists (e.g., Biber & Conrad, 2009; McEnery & Brezina, 2022). For example, Biber (1989) noted that text types often cut across genre categorisations. There is a fuzzy distinction between terms referring to the situational contexts where writing data is produced and collected. Larsson, Paquot, and Biber (2021) note that the terms "register", "genre" and "task," are used to refer to similar constructs in the reviewed studies. For example, Yoon and Polio (2016) explain that genres are distinguished by their communicative goals, sociocultural practices, and roles. Similarly, an expository essay may be described as a register (Larsson et al., 2021), genre (Lu, 2011), or task (Way et al., 2000). Establishing a clear distinction between these concepts is beyond the scope of this study. Rather, I have adopted the notion of a register to clarify the focus of the study, which is to identify the linguistic characteristics that can be associated with the register of interest. This term is chosen because it is appropriate for a bottom-up approach to linguistic analysis. Section 2.2.1.2 discusses how this has been conceptualised for the analyses in this thesis.

2.2.1.2 L2 Student Written Registers

The notion of r*egisters* is effective in revealing patterns in grammatical features, which may not be as salient in language use as a whole. For example, patterns of usage are different across learner corpora, native usage in conversation and fiction, and native academic prose (Biber, Conrad & Reppen, 1998). Notably, previous research has demonstrated the significance of situational factors in student register studies, such as genre types, disciplines, and learner variables, including target language proficiency, in explaining the linguistic functions associated with co-occurring linguistic patterns in texts (Hardy & Friginal, 2016; Gardner, Nesi & Biber, 2019). Consequently, these factors contribute to the multifaceted variations observed in linguistic features (Hardy & Römer, 2013; Hardy & Friginal, 2016).

A notable example highlighting the importance of considering situational factors is the study by Gardner, Nesi, and Biber (2019). They found two dimensions associated with different stance functions for personal stance and stance for attributing or evaluating others' work. They attribute these findings regarding stance function to two methodological advances. The first one concerns the utilization of a linguistic tagset that includes more stance and evaluation features than previous versions of Biber tagger, and the second is the addition of situational factors to interpret the dimensions such as disciplines, levels of study, and genre families (Gardner, Nesi & Biber, 2019, p.671). This finding is important for how we analyse text data, especially in studies that involve examining many variables at once (multivariate analyses) like a multidimensional analysis. The key idea is that considering the situation where a text was written (situational factors) can greatly improve how we interpret it. This helps us understand the different ways language is used (linguistic variation) and get a more accurate picture. Additionally, including more aspects of the language itself (linguistic variables) that might be related to what we're studying is beneficial. This can be done in two ways: either by directly examining how these language features connect to the patterns we're interested in, or by controlling for their influence to get a clearer overall view.

In the context of English as a second language, student writing as distinguished from general academic writing is referred to as the “student register" in some register studies (Nesi & Gardner, 2012; Gardner et al., 2019). It is important to recognize that written discourses require different linguistic resources compared to spoken discourses, and the appropriateness of the written register plays a critical role in distinguishing quality writing from oral fluency. As Myhill (2008; 2009) notes, transitioning from casual speech to formal writing can be challenging, particularly for learners whose spoken English differs significantly from Standard English. Similarly, L2 students with relatively proficient speaking skills may produce fluent but not necessarily appropriate written language. Myhill (2009) notes that coordination is more prevalent in speech while written discourses are more lexically dense and integrated compared to speech through the use of constructions such as subordination.

A series of works done on spoken grammar (e.g., Carter, 2004; Hughes & McCarthy, 1998) also offer insight into how language use in speech and writing is inherently different, which provides a complementary view to the work done on the frequency parameters governing a continuum of different registers (e.g., Biber et al., 2021a). Spoken and written discourses exhibit notable differences, such as time constraints and the absence of immediate feedback in writing. Acquiring the necessary skills to produce quality writing involves adapting to the specific demands and characteristics of written registers, which require conveying ideas without relying on context-dependent cues. This transfer from speech to writing can be particularly demanding for L2 learners. Some L2 learner studies suggest that the distinction between spoken and written registers is a valid framework for understanding L2 writing quality, drawing parallels with certain observations from L1 research.

These register-specific variations of grammatical features indicate the necessity of investigating L2 grammatical development as a separate category from L1 findings. L2 Learner corpora studies based on linguistic developmental theories provide valuable information about how speakers of various linguistic backgrounds develop L2 linguistic commands (e.g., Housen, 2002; Wulff & Gries, 2011). Analysing these linguistic differences can provide evidence of linguistic development in academic writing, unique to a particular variety of student registers. Examining the occurrences of syntactic features within the notion of register allows for direct observation of the linguistic features crucial to academic writing, such as nominalization. This perspective on linguistic analysis contributes to understanding language use in contexts, overcoming the limitations of presupposed assumptions inherent in more traditional methods. Therefore, the notion of register, defined as a collection of texts sharing a similar situational context, is adopted to refer to the 'L2 student register' in this thesis.

Previous studies, such as those English university students' essays (e.g., BAWE) and L2 learners' essays (e.g., TOEFL answers), offer complementary insights in terms of student register and L2 learner register, respectively. Therefore, L2 needs to be separately established register from L1. This definition of an L2 student register implies a situational distinction between L1 and L2 registers. Therefore, the empirical findings regarding these two registers are not readily compatible in terms of drawing conclusions as to whether they follow similar paths of writing development. On the one hand, it may be due to the marked differences in research questions and focuses between L2 and L1 research, as highlighted in the comprehensive review of these areas by Durrant et al. (2021). The dissimilarity is likely to arise from the variation in the nature of the data associated with instructional and learner-specific variables as well as methodological issues such as the duration of data collection.

Phrasal indices have been identified as crucial indicators of linguistic development in both first-language (L1) and second-language (L2) academic writing (Staples et al., 2016; Gardner, Nesi & Biber, 2019; Taguchi et al., 2013; Parkinson & Musgrave, 2014; Biber, Gray & Staples, 2014). The concept of register presents inherent complexities. It encompasses situational factors, the writer's specific style, and linguistic variations that arise from the collective and conventional nature of the communicative event. These variations are distinct from the individual qualities unique to a particular writer's style. Consequently, it is essential to define the specific type of register under investigation. Furthermore, it is crucial to determine whether and how this register diverges from the target register employed as the benchmark for evaluation. To that aim, Section 2.2.2 examines the previous discussions on L2 writing as a distinguished register based on empirical findings.

It is remarkable to find a broad consensus that the communicative purposes inherent in a situational context mark the shared path of linguistic development in L1 and L2 writing. This commonality may be explained in terms of sociolinguistic perspectives. Kecskes (2018) points out that the idiom principle and the economy principle are assumed to be the default in processing across different linguistic typologies. More specifically, the quantity of formulaic items in English, German, and Korean is very similar, offering support for the idea that human languages in general benefit from retrieving easy-to-use processing expressions. A similar observation has been suggested by Biber et al. (2004, p.376), taking it from a different angle of seeing pre-fabricated multi-word units: a frequency. Noting that frequency is not the only factor determining this phenomenon, they emphasise the significance of frequency data in identifying recurring linguistic patterns and suggest that higher frequency sequences are more likely to be stored and used as prefabricated chunks in language.

It has been noted that many formulaic expressions convey social functions that may be useful. Formulaic expressions appropriate to the academic written registers are a marker of community membership (e.g., Durrant, 2018; Hyland, 2008). This leads to the prediction that some formulaic expressions may serve to mark proficiency when they are appropriate to the target register (Wray, 2018). Appropriate language use in academic writing, for example, may include sophisticated words within phrasal structures, which makes meaning more specific.

2.2.2 Empirical research on L2 writing development

2.2.2.1 Studies of University Student Registers

The theoretical discussions in Section 2.2 have provided the rationale for focusing on phrasal development, rather than the level of clauses as a more effective indicator of linguistic development, capturing the abstract and information-rich nature of academic texts and facilitating the production of concise and abstract sentences. Spoken and written discourses are well known to be different (e.g., Crystal, 1995, p.291). Writing is relatively less time-bound for reading and revising and does not have immediate feedback from or interaction with readers. On the other hand, speech is more pressurised for spontaneous production with less time for planning. If written discourses are different from spoken ones, such differences need to be acquired to produce quality writing. Academic writing often employs phrasal constructions, such as embedded prepositional phrases, adjectives, and nouns as prenominal modifiers, in a condensed and concise manner (Biber, 1992; Biber & Gray, 2011). These phrasal compressions contribute to text cohesion and coherence, enhancing the efficiency of written communication (Halliday, 1979).

Effective written communication differs from spoken communication in the way language is used. Written text lacks the context-dependent cues of speech, such as facial expressions or gestures. Therefore, writers need to employ different linguistic resources to effectively convey their ideas. Mastering written register is a crucial linguistic ability. It distinguishes between high-quality writing and simply being fluent in spoken language.

In terms of the syntactic function and structural form, academic writing prefers phrasal rather than clausal structures functioning as constituents in noun phrases (Biber & Gray, 2016, p.94). The use of prepositional phrases as post-nominal modifiers is a clear case of this type: they are prevalent in academic writing but rare in conversation. Similarly, attributive adjectives and nouns as nominal pre-modifiers behave as phrasal constituents embedded in the noun phrase. Nominalisation makes texts more concise, creates textual cohesion, and helps the reader follow the text (Staples et al., 2016).

Empirical findings of first and additional language children's English writing development inform many insights into language acquisition (e.g., Myhill, 2008, 2009; Christie, 2012; Reppen, 207; Durrant & Brenchley, 2022; Durrant, Brenchley & Clarkson, 2020). One example of such insight is Christie (2012), which notes that the expanded noun groups using embeddings are critical developmental markers in early English writing. However, young writers face significant challenges in transitioning from speech to writing. This difficulty arises because written registers, unlike spoken communication, are shaped by situational constraints. Effective written communication relies less on context-dependent cues (like facial expressions or gestures) and more on the precise use of language to convey ideas clearly. This shift in reliance presents a hurdle for young writers who are accustomed to the immediacy and context-rich nature of speech (Wells & Chang, 1986, p.123). This task can be incredibly demanding for writing in second or additional languages. Myhill (2009), who has studied children's writing development in British school contexts, noted that academic writing produced by students with relatively proficient speaking skills of English as an additional language may exhibit rather fluent but not necessarily appropriate linguistic use for written register. While studies on L1 and L2 learning in schools and university contexts should be considered separately in terms of their own contexts (which should vary by individual differences such as L1 linguistic backgrounds, age, L2 proficiency and many more), several studies have shown that L1 and L2 writing development follow a similar trajectory. I have adopted a usage-based perspective to examine the previously explored hypotheses regarding L2 writing development, taking L2 writing as a distinguished register from L1. For that aim, I have taken the English university environment as the key consideration of reference point in comparing L1 and L2 students' written essays in this thesis. This shared communicative context enables us to compare, for example, phrasal complexity features in studies of L1 academic writing development (Staples et al., 2016; Gardner, Nesi & Biber, 2019) and L2 academic writing (Taguchi et al., 2013; Parkinson & Musgrave 2014; Biber et al., 2014).

Among studies on the university student register, some studies particularly investigated the English university students' writing such as the British Academic Written English corpus (BAWE; Alsop & Nesi, 2009) and MICUSP corpora, representing L1 registers. For instance, Staples, Egbert, Biber, and Gray (2016) used the BAWE to investigate how phrasal and clausal complexities develop in student writing at the university level. They found that phrasal complexity features in writing were correlated with increasing academic levels, while clausal complexity features, particularly finite dependent clauses, appeared less frequently as academic levels increased (2016, pp.153-154).

This study is based on the idea that similarity exists between L2 and L1 linguistic development patterns in university academic writing due to the shared communicative environments. Admitting that L2 academic writing usually takes place after acquiring adequate L1 writing proficiency, this notion of shared trajectory of writing development is to be examined, rather than taken as a priori assumption. These L1 and L2 learning are usually considered separate categories for research enquires, given the potentially more diverse situational contexts of L2 learning compared to L1 students’ English use. Interestingly, Staples, Gray, Biber and Egbert (2023) found overall similarities in the use of key grammatical complexity features across university levels (from undergraduate to graduate-level writing) across L1 and L2 English writers. This observation aligns well with L2 grammatical development from a usage-based assumption that learners' grammar should reflect the general input frequencies to which learners are exposed.

The process of L2 learning shares similarities with how we learn our first language (L1). Research by Lightbown (2013) and Mitchell (2013), for example, suggests that L2 learners tend to follow similar developmental sequences in acquiring written register as they do in L1. However, a learner's native language can influence this process and cause some variations (e.g., Meisel, Clahsen, & Pienemann, 1981).

L2 learner corpora studies informed by linguistic developmental theories offer valuable insights into how learners with diverse linguistic backgrounds acquire L2 linguistic skills (Housen, 2002; Wulff & Gries, 2011). One particularly relevant factor for studying written L2 development is writer-specific variation. However, this very factor introduces complexity when interpreting how learners develop linguistic complexity.

Previous research on formality (Hardy & Römer, 2013; Larsson & Kaatari, 2019, 2020) highlights a related concern. For example, learners with different linguistic backgrounds and proficiency levels may not consistently apply "formality" across various communicative purposes and target language contexts.

To maintain research focus and utilize less heterogeneous data, this study will confine its scope to traditional English academic written discourse. While this approach provides clarity, it also poses limitations. Notably, the homogeneity limits our ability to interpret how writing development and authorship are redefined by online learning environments and technological advancements. These advancements potentially call for a reconsideration of how grammatical and lexical knowledge are prioritized in language classrooms, compared to their traditional importance.

While a full reconceptualization of writing development falls outside the scope of this present study, it underscores the importance of carefully considering which aspects of language variation matter most for data analysis, particularly if aiming for relevance to current pedagogical concerns.

The second assumption derived from the previous research on L2 syntactic complexity is that L2 writing develops from clausal to phrasal complexity. Biber and his colleagues propose that L1 English and L2 English students follow a hypothesised sequence of developmental stages along two structural parameters: grammatical type and syntactic function (Biber et al., 2011; Biber et al., 2021a). They hypothesized that student writing becomes more complex at higher proficiency levels, which phrasal-level measures can better capture based on their findings in expert-written texts (Biber et al., 2011, p.13). The studies discussed in this chapter generally confirm that there's a general pattern which is observed in both L1 and L2 data, for example, where phrasal features become more salient with more advanced writing. In line with this assumption, corpus studies have shown that academic writing has distinctive features that separate it from spoken discourses, such as nominalisation (Halliday & Webster, 2004, p.171). In addition, conceptualising writing development often encompasses the ability to use language to address complex communicative needs (e.g., Council of Europe, 2011). Consequently, the theoretical and empirical discussions on language complexity take into consideration the influence of communicative demands within a situational context. This point is elaborated on in Section 2.3, which is the basis of the discussions regarding the distinction between clausal and phrasal complexity in written registers.

A subsequent hypothesis emerging from the assumption of phrasal complexity in written registers is that writing development starts from spoken register features and progresses to written registers. As noted earlier by Durrant et al. (2021, p.3), writing development involves using appropriate language forms to fulfil communicative needs or functions. This notion of writing development entails that language use is associated with the function it intends to accomplish. The association between linguistic features and communicative functions in a text has attracted significant attention in previous studies (e.g., Ervin-Tripp, 1972; Irvine, 1979; Ochs, 1979; Brown & Fraser, 1979). When considering this form-function mapping as a crucial linguistic competence, linguistic features more suitable for academic written discourse can serve as indicators of writing development to some extent. Grammatical development in writing typically occurs later, with written grammatical structures exhibiting a significantly higher level of production complexity compared to spoken discourse (Biber et al., 2011, p.29). Academic writing is typically acquired at a later stage (Biber, Gray & Staples, 2014, p.7). Thus, it can be inferred that the developmental progression of both L1 and L2 learners is likely to involve a transition from language use associated with spoken register to written register. This transition has been observed in learner corpora studies (e.g., Kobayashi & Abe, 2016; Kim & Nam, 2019). Therefore, it is hypothesised that the elaborated and extended structure is likely to mark the relatively earlier developmental stages as novice writers may have difficulty producing elaborated structures in writing at the beginning. This will be replaced with more condensed, compressed phrasal structures more effective for packing information in written discourses.

Staples et al. (2022) argued that there is no firm ground to believe that L1 and L2 should be following the same developmental trajectories of grammatical complexity. Providing empirical evidence, they noted that L1 and L2 grammatical complexity exhibited a broadly similar pattern in that they use more phrasal features and fewer clausal features at higher levels of academic study. At the same time, they also noticed more nuanced differences between the two registers. The most pronounced difference between the two groups is that L2 English writers are using greater frequencies of phrasal features, particularly premodifying phrasal features when compared with L1 English writers in the first years of undergraduate writing. Two interpretations were posited to explain this difference: (i) L2 writers may be relying more heavily on chunks of technical language rather than creating their formulations to show sophisticated abstract connections between concepts; (ii) L2 writers may be more attuned to this register difference due to their relatively greater reliance on written English language use when compared to spoken English. These interpretations resonate with discussions made by other researchers on formulaicity in L2 writing (e.g., Wray, 2018; see Section 6.2) and multidimensional analysis of L1 and L2 writing (e.g., Pan, 2018; see Chapter 4, Section 4.2). How these discussions have informed the interpretations of the findings of this thesis is elaborated in the discussion sections of each analysis in Chapters 4, 5 and 6.

2.2.2.2 Longitudinal Studies of L2 Written Registers

Research on L2 writing has primarily focused on writing quality rather than development, with a predominant reliance on cross-sectional data. Numerous studies on L2 development rely on cross-sectional data, where texts are classified based on learners' proficiency levels or human ratings. However, such classifications are influenced by external criteria of what constitutes good writing and do not adequately capture the stages of individual learners' writing processes. Longitudinal evidence from large data, which can shed further light on the syntactic development through a personal course of study, is relatively scarce. There have been longitudinal studies drawing on L2 secondary school students’ writing (e.g., Kyle et al., 2021), L2 Test of English as a Foreign Language (TOEFL) answers (e.g., Gray et al., 2021), or English university students’ essays(e.g., Biber et al., 2021a). Still, there is a need for more longitudinal data collected in diverse situational contexts. The call for more longitudinal analysis highlights the need to enhance our understanding of the complex and dynamic phenomena of individual and non-linear language development. In addition, quantitative longitudinal analysis, along with qualitative case studies, can add a diverse range of evidence to facilitate cross-validation of our observations on development. Specifically, nonexperimental data collected in university English learning programmes contributes to research-driven pedagogy (Polat, Mahalingappa & Mancilla, 2020, p.689). Longitudinal, natural writing data collected and analysed in language classrooms enhance the ecological validity of the findings, as it closely resembles real-world writing experiences, balancing the types of linguistic data used to explore L2 writing development.

While researchers acknowledge the potential of grammatical complexity to understand writing development, it's not a simple task to directly map these variations onto interpretations of a writer's progress. Consider, for example, the significant disconnect between the focus of current research on written feedback and the actual practices teachers use in L2 classrooms. Rastgou et al. (2020) highlight this disparity, noting that research often emphasizes specific aspects of writing that may not always be reflected in teachers' feedback (p.42). Furthermore, the gap between theory and practice can lead to blurred boundaries when it comes to applying linguistic concepts. For instance, Rastgou et al. (2020, p.55) observed a decline in broad syntactic complexity measures at the sentence level, which became intertwined with changes in fluency. This makes it difficult to isolate the specific development of syntactic complexity from the overall writing flow.

One complication particularly relevant to research on longitudinal data is the duration of observation. For example, no consensus has been made about the length of the instruction to show an effect on the increase of the syntactic complexity (Polat, Mahalingappa & Mancilla, 2020, p.692). Perhaps in-depth qualitative analysis can unpack the complex nature of feedback provision and uptake process along the revision process. For example, Hyland (2000) used qualitative methods such as interviews with six L2 students from various L1 backgrounds in an intensive English programme at a US university. However, such intensive observation on a relatively small number of participants may not be optimal to explore consistent linguistic variations such as syntactic complexity development over time. While qualitative longitudinal analysis tends to have a longer observational duration but a smaller data size for an intensive analysis, quantitative longitudinal analysis can deal with larger data sizes. However, the data size and span vary among quantitative longitudinal studies (e.g., 2 years, Kyle et al., 2021; one year, Gray et al., 2019; two years, Biber et al., 2020b); six months, Crowsthwaite (2016). Considering these issues in L2 longitudinal analysis, the next section presents the overarching research aims of the longitudinal analyses discussed in Chapters 4, 5 and 6.

2.3 Theoretical Perspectives

This section focuses on different theoretical perspectives on syntactic complexity (SC). The discussion starts with a commonly accepted definition of SC as reflected in vast theoretical discussions and as operationalised in empirical studies. Then, the focus shifts to an alternative perspective drawing on a usage-based theory, which the current study is based on. The discussion involves the underlying assumptions of this alternative perspective on SC, the operationalisation of SC, and empirical studies based on this view of SC.

2.3.1 Syntactic Complexity in English Academic Writing

2.3.1.1 Theoretical Discussions of Syntactic Complexity

Establishing a theoretical definition of the SC construct is crucial for the accurate measurement of sub-constructs of linguistic complexity. However, the concept of syntactic complexity (SC) is a subject of ongoing debate among researchers in theoretical and empirical investigations. This section discusses a theoretical construct of SC which has been vastly utilised in operationalising the construct. The discussion on SC is a starting point for introducing the construct complexity construct, which is the overarching definition of the linguistic complexity variable of focus in this thesis.

Much literature on SC has attempted to define it by tapping into the formal aspect of linguistic complexity. Specifically, it has been defined as the number of structural elements and their interrelationships (Biber et al., 2020a, p.5; Pallotti, 2015, pp. 117-118). Previous research has operationalized syntactic complexity by measuring the length of a syntactic unit or the number of dependent occurrences within it. Following that, existing literature extensively examines the length of dependents within the targeted syntactic unit, often employing definitions such as the T-unit, defined as one main clause plus any subordinate clauses or non-clausal structures that are attached to or embedded within it (e.g., Hunt, 1965, 1966, 1970). Consider the two following examples that consist of nine words each:

(1) a boy and a dog went to the park

(2) he went to school :: while she stayed at home

In Example (1), the two underlined phrases have five and three words, respectively. On the other hand, Example (2) has four such underlined phrases consisting of one or two words each. If the length of phrases (the mean number of words in each phrase) is measured, Example 1 will be considered more complex, as its mean number of words per phrase (=4) is greater than Example 1 (=1.5). However, if we change the observation unit from a phrase to a higher grammatical level of units, such as a T-unit, the complexity will be measured differently. While Example (1) has one T-unit and one clause, Example (2) has one T-unit and two clauses, therefore the complexity measure for clause per T-unit would be greater for Example (2) (=2) than Example (1) (=1).

While the measurement units may differ in terms of their grammatical and syntactic level depending on the focus of SC research, these measures have the shared underlying assumption that the longer language production (having more linguistic units) is the indication of a more structurally complex language. Therefore, this measurement of SC is an attempt to learn about a purely formal property of the language itself disentangled from the dynamics of language use and learning related to language users or communicative contexts.

This illustration of measuring the SC construct entails that each unit for measurement needs to be defined at an operational level, at a lower level of word to a higher level of sentence or T-unit. One concern in SC research is the difficulty in defining and interpreting the construct of syntactic complexity. Researchers tend to have different definitions and operationalizations of syntax (Durrant et al., 2021, p.60). Moreover, the construct of complexity itself poses challenges in operationalizing syntactic complexity. A further issue regarding the operationalization of the complexity construct is defining the measurement unit consistent with the theoretical level of definition. Many complexity measures are ambiguous or hybrid measures, capturing multiple potentially independent and unrelated complexity constructs and sources of complexity simultaneously (Bulté & Housen, 2012, p.35).

For instance, Hunt's (1966) T-unit, used as a measure of a clause, has been criticized for its ineffectiveness in capturing the complexity of syntactic development in L2 writing (Ishikawa, 1995, p.56). Biber et al. (2011, p.13-14) express similar concerns about T-unit-based measures being inadequate discriminators of sentential complexity. These measures are problematic as they overlook other possible complex syntactic features or the functional aspects that inform the use of such features in texts (Biber et al., 2020, p.10). Moreover, these unitary measures reduce the descriptive information about linguistic development in different types and genres of writing, sacrificing the richness of information about syntactic complexity (Biber et al., 2020a, p.3). Researchers have emphasized the need for complexity measures to go beyond subordination-based measures and provide a more comprehensive and explicit understanding (Biber et al., 2011; Biber & Gray, 2013, p.68; Norris & Ortega, 2009; Bulté & Housen, 2012). Alternative measures, such as coordination and phrasal complexity, have been suggested to study complexity (Norris & Ortega, 2009).

In more practical terms for research design, operationalised indices of syntactic complexity can differ according to the definitions opted for research. Furthermore, the definition of syntactic complexity affects the frequency counts of the measure, which in turn affects interpretations of the linguistic phenomena of interest. Therefore, considering the importance of operational measures in empirical studies, the next section discusses the studies on syntactic complexity employing the different levels of measure granularity.

2.3.1.2 Empirical Research of Syntactic Complexity

In line with the concern about the operationalisation of SC as discussed in the previous section, the interpretation of SC study results is not straightforward due to the potential variations in how the quantified measures of SC can tap into the construct of SC (Ortega, 2003; Durrant et al., 2021). Studies conducted in tertiary education settings have yielded contradictory findings regarding the relationship between SC measures and writing quality. Some studies have reported negative relationships between writing quality and various SC measures (e.g., MacArthur et al., 2019) while others have no significant correlations (e.g., Crossley et al., 2015; McNamara et al., 2013). Despite the existence of these rather inconsistent results on SC in L2 writing development research, considerable empirical research supports the correlation between syntactic complexity and writing quality as assessed by human rating (e.g., Crossley et al., 2011; Bulté and Housen, 2014; Malvern et al., 2004).

Several earlier studies on syntactic complexity (SC) suggest that SC tends to increase at a specific level or age and exhibits a discernible correlation with the quality of L1 writing (Hudson, 2009; Loban, 1976). Furthermore, empirical findings in L2 writing also support the idea that syntactic complexity correlates with writing quality assessed by human rating (Bulté and Housen, 2014). For example, Bulté and Housen (2014) found that growth in global T-unit and mean length of finite clauses correlated reasonably well ($r = .40$ and $.48$, respectively) with ratings of writing quality, which themselves showed a substantial mean improvement (of around $d = .70$ to 1.0 depending on the rubric scales).

Recent studies in syntactic complexity development have focused more on phrasal level indices, exploring various noun phrase complexity indices in L1 and L2 writing development (e.g., Kyle & Crossley, 2018; Larsson & Kaatari, 2020; Durrant & Brenchley, 2022). Notably, findings support phrasal complexity as a crucial indicator of linguistic development in academic writing, providing more nuanced and diverse measures. This shifting focuses on the phrasal complexity in L2 writing in some studies show a similar observation to the previous studies using a sentence or clausal level indices but with a stronger correlation with L2 writing quality (e.g., Kyle and Crossley, 2018; Casal and Lee, 2019). These findings support the claim that phrasal complexity indices are better predictors of writing quality than clausal indices (e.g., Biber et al., 2014).

However, using more diverse and granular measures for SC research involves analysing language grammaticality with more specified definitions for each grammatical category. For example, Biber et al. (2021) argue that it is important to distinguish grammatical forms from their syntactic functions. In other words, a grammatical category based on its form (e.g., noun, verb, or preposition) should be identified and then followed by identifying its function in a sentence unit (e.g., preposition as phrasal modifier). Using computational tools for automatic POS tagging and syntactic parsing may help enhance the consistent identification of grammatical categories for researching syntactic development. One of the tools developed particularly for SC analysis is the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010, 2011), which includes 14 different indices incorporating the potential syntactic complexity markers discussed in Wolfe-Quintero et al. (1998) and Ortega (2003). Another computerised analytic tool is TAASSC (Kyle, 2016). TAASSC relies on a Stanford dependency parser (De Marneffe et al., 2014) to identify syntactic structures. This section discusses three studies drawing on one of the two tools. These studies have in common that their focal linguistic constructions include phrasal complexity indices (e.g., prepositions in noun phrases).

The first study to discuss is Juffs (2020), who provides detailed accounts of quantitative longitudinal analysis of syntactic complexity found in the written texts written by the eight students' 48 essays written at least throughout three consecutive semesters for grammar, reading and writing classes in the language programme at Pittsburgh University. He focused on two indices from L2SCA (Lu, 2011), complex noun phrases per clause (CNC) and mean clause length (MLC).

Notably, CNC is reported to be a stronger marker of syntactic complexity across English levels 3 to either Levels 4 or 5 (general English levels assigned by the language programme) than MLC (Juffs, 2020, p.211). However, when it comes to the correlation between these measures and the writing scores for each text, MLC was reliably correlated with the average essay writing score (Juffs, 2020, p.218), while CNC was not. Notably, MLC was not significantly correlated with the number of words. Juffs interprets this finding as an indication that the raters paid considerable attention to both the length of clauses and the number of words, but the length of clauses was considered even in shorter texts.

Overall, the study findings regarding the two syntactic complexity measures showed significant correlations with writing quality. Notably, MLC, together with text length, was reported to explain some of the variance associated with the human rating of the texts (approximately 25%). Furthermore, Juffs notes that the partial explanatory power of these measures indicates that human raters considered other factors than these elements, such as a semantic and organisational aspect of writing (Juffs, 2020, p.218). Taken together, these findings indicate that some of the SC measures based on elaborated clausal constructions can explain the syntactic variations related to the qualitative assessment. However, it also informs us that human raters consider complex elements of quality writing appropriate to the writing assignment in the target levels, and therefore, the patterns of the syntactic variations not explained by the underlying assumptions of these measures might also require to be complemented by qualitative analysis for a fuller understanding of the syntactic complexity in the L2 writing.

The second example study utilising an automatic analysis tool is Larsson and Kaatari (2020), who used L2SCA to explore formality in L1 and L2 student corpora as an essential characteristic in an academic written register. They found that noun phrases are a particular indicator of the difference between a learner and an expert-written register and used TAASSC (Kyle, 2016) to examine complex nominals further. Their finding showed that learners underused adjectives and prepositions inside noun phrases in comparison to the expert writers. The adjectival and prepositional modification by expert writers function to add more detail to technical descriptions of methods or results (Larsson & Kaatari, 2020, p.11). The adjectives and prepositions in noun phrases and that-complement clauses were among the key features studied in Chapter 6. While many of these features did not show significant variance in the analyses in Chapter 4, these features were deemed to hold research value and therefore separately investigated.

Kyle & Crossley (2018) explored the relationship between syntactic complexity and writing quality evaluated by human raters in 240 TOEFL argumentative essays using L2SCA (Lu, 2010) and TAASSC (Kyle, 2016). They found that fine-grained indices of phrasal complexity (e.g., number of dependents per prepositional object) showed stronger correlations with the writing score than any type of clausal complexity indices. They conclude from their finding that phrasal indices of syntactic complexity prove to be a better predictor of writing quality than clausal indices. This remark should be noted with caution, as they also note that clausal indices also showed a statistically significant correlation with writing quality, showing a complementary relationship with phrasal indices. The combined syntactic complexity indices were reported to explain 20.3% of the variance in holistic essay scores. In other words, these syntactic complexity measures should be interpreted as only a partial indicator of writing quality.

The studies discussed so far have found that the prepositional constructions prove to be a marker of academic prose and predict the quality of learner writing. This finding is consistent with the studies not using one of the linguistic analysis tools, such as the study by Durrant and Brenchley (2022). These authors, exploring British children’s school writing, have used manually tagged written data to find that prepositional constructions are one of the significant markers of language variation across Key Stages. While these studies are not directly comparable in terms of methods and linguistic data, it is notable that prepositional construction was identified as the key marker of writing development. These studies operationalized syntactic complexity by measuring the length of a syntactic unit or the number of dependent occurrences within it. This operationalization is based on the notion of elaboration, which defines syntactic complexity as the extent to which a linguistic unit is extended. Empirical research findings of studies of SC suggest the importance of considering syntactic complexity in research on L2 writing development. Syntactic complexity correlates with English writing quality as assessed by human ratings, study level, or L2 English proficiency (e.g., Bulté & Housen, 2014; Kyle & Crossley, 2018; Lu, 2011; Juffs, 2020).

The empirical studies of SC discussed in this section were the basis of some of the pilot analyses, which are reported in Chapter 4 (Section 4.5) and Chapter 6 (Section 6.3). As noted earlier, a consensus on the theoretical and operational definition of syntactic complexity has yet to be established. Particularly, the inclusion of phrasal indices in syntactic complexity research indicates the importance of a situational parameter, *register*. In other words, phrasal complexity is particularly relevant in academic written discourses, which has led to increasing attention to this aspect of linguistic complexity in academic writing developmental research. However, the approaches to researching syntactic complexity in writing share the notion of *elaboration* as an underlying parameter of development, overlooking another distinct quality observed in written registers by researchers: structural compression. This aspect of linguistic complexity will be introduced under the notion of grammatical complexity in subsequent sections. The final analyses in Chapters 4, 5, and 6 incorporate the methods taking this notion of grammatical complexity reflecting a register perspective and a usage-based theory as an overarching frame to map the grammatical variation on developmental interpretation. The following section (Section 2.3.2) discusses this usage-based perspective on grammatical complexity.

2.3.2 A Usage-Based Perspective for Grammatical Development in L2 Academic Writing

The growing interest in phrasal complexity in academic writing developmental research, as discussed earlier, reflects the influence of communicative purposes inherent in the situational context on language use, in addition to the structural complexity of linguistic constructions, which has most often been operationalised in SC studies. One specific critique made by Biber and colleagues is directed towards approaches that measure "*elaboration*" by counting words or dependents within the targeted syntactic unit. Biber collaborated with several researchers to take further steps to propose a different conceptualisation of ‘grammatical complexity’ in linguistic analysis in several publications (e.g., Biber et al., 2023).

With this, the first part of this section reviews the rationales and application of their theoretical views of grammatical complexity from a usage-based perspective. The discussions form a partial review of the theoretical definition of grammatical complexity from a register perspective. Investigating a particular written corpus produced in a shared situational context, such as a language programme in a university, from a register perspective, may contribute to understanding linguistic complexity development in L2 corpora as a collective entity in general, and as a corpus collected in unique situations in specific. The underlying assumptions of usage-based theories (e.g., Ellis, 2002) and a register perspective to researching linguistic development (Biber et al., 2021a) are discussed in the following subsection.

The second subsection, subsequently, discusses some empirical study findings from this perspective on grammatical complexity. The implications of the theoretical discussions and empirical findings provide the foundation of the research aims for the current study, which will be introduced at the end of this chapter. A conclusion will be reached as to why a register perspective can aid in understanding grammatical development in writing, which will lead to the research aims of this thesis in Section 2.4.

2.3.2.1 Challenges in Operationalizing Syntactic Complexity (SC)

Researchers acknowledge the multifaceted nature of SC, highlighting its complexity as a concept (Bulté & Housen, 2012, 2014; Biber et al., 2011; Biber et al., 2021a). This complexity creates challenges when translating theoretical definitions into practical methods for measuring SC (operationalization). For example, Ortega (2015, p.82) defines SC as "the range and sophistication of grammatical resources exhibited in language production." This definition encompasses related concepts like diversity, variety, and the degree of syntactic structures used.

A key challenge in SC research lies in the distinction between the theoretical concept and how we measure it in practice (Bulté & Housen, 2012). Many studies have focused on measures at the clause and sentence levels (a more detailed discussion on this is found in Section 2.3.1). However, Bulté and Housen (2012) argue that including measures at the phrase level can provide more fine-grained insights into learners' grammatical development. Their framework suggests subcomponents of SC, such as grammatical diversity and sophistication, observed through "depth, embeddedness, and compositionality of grammatical L2 structures" (2012, pp.27-28).

Discussions surrounding SC often separate it from other interacting constructs, such as the difficulty of a grammatical feature (Bulté and Housen, 2012, p.36). While acknowledging the importance of the cognitive aspect of complexity,many researchers argue for SC as a purely linguistic property, distinct from factors like how difficult it is to learn or use (e.g., Bulté and Housen, 2012; Pallotti, 2009, 2015).

However, the multidimensional nature of SC means it interacts with other relevant constructs, including cognitive difficulty (as noted earlier). One such construct is "syntactic sophistication," which refers to a distinct aspect of development not directly captured by traditional SC measures (Kyle & Crossley, 2017, p.514). Sophistication focuses on the level of development or control exhibited in complex language production. This intricacy makes it difficult to separate SC from what Pallotti terms cognitive complexity (difficulty; 2015, p. 118).

While sophistication is a separate construct, some studies have explored the relationship between SC and sophistication, suggesting they might be complementary. For instance, Kyle et al. (2021) found that "VAC sophistication" (a measure of verb-argument construction complexity) was a stronger predictor of human-rated writing quality than traditional SC measures.

Arguably, syntactic sophistication and complexity are two distinguished but closely interconnected, therefore complementary, constructs. However, many studies investigating these constructs focus on infrequent elements, considering language use that is "infrequent but appropriate to the target register" as sophisticated or complex (e.g., Romer & Berger, 2019; Kyle et al., 2021; Biber et al., 2021).

This highlights the multidimensionality of SC and the existence of related constructs like cognitive complexity. While a direct exploration of these relationships is beyond the scope of this thesis, this study acknowledges the complexity of SC. This study adopts grammatical complexity as an overarching construct. To analyze the development of this construct, I primarily rely on frequency measures to interpret the linguistic variation within the learner corpus

(Chapters 4 and 5). However, it's important to distinguish between two types of frequency:

> Learner corpus frequency: This refers to how often a particular grammatical form appears within the learner corpus itself. This type of frequency measure is traditionally associated with linguistic complexity.
>
> Reference corpus frequency: This approach compares the frequency of grammatical forms in the learner corpus to their frequency in a reference corpus representing target-like language.

My interest is in how complex the learners' writing is (grammatical complexity). I analyse how often they use specific grammatical features (learner corpus frequency). However, to understand sophistication, I also compare this to how often those features are used in the target register (reference corpus frequency). This comparison allows us to identify forms that learners use more or less frequently than the target register, potentially reflecting their developing control over sophisticated grammatical structures.

Additionally, I investigate grammatical sophistication separately using methods that explore the association strength between lexico-grammatical elements in construction (Chapter 6). Chapter 3 (methodology chapter) will introduce the specific methods used for this aspect of the analysis, situated within the broader framework of grammatical complexity explored in Chapters 4 and 5. Chapter 6 will provide a more detailed account of the theoretical foundation for the analyses using association measures.

Another relevant construct is the notion of *developmental complexity* (e.g., Biber et al., 2021; Pallotti, 2015). Pallotti (2015) notes that what is especially relevant to L2 learning, distinct from the formal linguistic complexity or cognitive difficulty is that of developmental complexity where the order of the emergence of linguistic structures is considered in terms of the target communicative context. While Pallotti considers this construct should be excluded from discussions of defining SC, usage-based complexity taps into this aspect at least to some degree (e.g., Biber et al., 2022). Admitting that these related constructs may interact with linguistic complexity, elucidating the relationship between these constructs is beyond the scope of this thesis. Instead, I attempt to discuss the common ground for interpreting development from linguistic variation that is observed in the theoretical discussions on the notion of developmental complexity, a usage-based perspective, and the approaches to grammatical complexity development in written registers. For that aim, I will introduce an underlying rationale for the approach to grammatical complexity research, by drawing on a usage-based theory in the following paragraphs. As overaraching theoretical constructs, the definitions of frequent and (semi-)formulaic constructions needs to be defined, along with how they can be measured. As Chapters 4 (and 6) and 5 adopt each of these two related but distinguished approaches to measuring language use, the literature review sections in Chapters 4 and 5 address more direct issues relevant to using these measures, respectively.

2.3.2.2 Theoretical Assumptions of a Usage-based Approach to Syntactic Development in Academic Writing

The usage-based approach posits that language use is shaped by actual usage rather than formal knowledge about language (Ellis, 2002). In other words, language evolves through interactions and communicative purposes inherent in its use. This assumption of the usage-based approach to language learning research leads to the two following assumptions.

The first assumption is the interconnectivity between grammar and lexis. These grammatical patterns joined with lexical instantiations are referred to as constructions (Goldberg, 2003, 2006). Durrant (2018) emphasizes understanding the relationship between grammatical form and its function in language for investigating L2 learner writng development. Wulff (2018) also makes a similar observation that many abstract syntactic frames can be considered constructions serving a functional purpose, simultaneously stored in multiple forms that differ in their level of complexity and abstraction. In this regard, identifying strong form-function mapping in a corpus enhances the interpretation of the corpus by providing both bottom-up and top-down information. This means that the corpus can be understood from both specific instances of language use (bottom-up) and the broader communicative function (top-down).

The second assumption is that language learners are expected to produce the target language once they have been exposed to sufficient input appropriate to the target language. The link between input exposure and output is one of the underlying assumptions that the present study draws on for interpreting quantitative measures: language input in a formal educational context is likely to facilitate the acquisition of linguistic devices associated with the academic register. Furthermore, specific constructions frequent in academic written discourses may lead to learners' acquisition and formation of constructional knowledge (Ellis, 2002; Tyler, 2010). The association between learnability and exposure to contextual language use implies an important direction for language development, laying the foundation for predicting the order of language production in developmental trajectories. A large body of psycholinguistic research demonstrates that both language processing and language acquisition are sensitive to the distributions of linguistic constructions in usage (Guo & Ellis, 2021, p.1). The emerging hypothesis from this is that more frequent and semi-formulaic items are learned earlier, which is one of the overarching assumptions for predicting developmental trajectories of grammatical features in this thesis.

This notion of form-function mapping, which posits that language consists of constructions that are learned through usage, has been examined in previous studies to demonstrate empirical evidence of using and learning language constructions as a mental representation of these constructions for L2 learners (e.g., Ellis et al., 2016; Durrant & Schmitt, 2009). For example, Ellis et al. (2016) found that the first-learned verb in each VAC is prototypical of that construction's action semantics but also generic and thus widely applicable while other verbs that are prototypical but have other meanings tended to be learned later. Based on this theory, it can be assumed that constructions that are generally infrequent but relatively more frequent in a specific register, such as academic writing in a university classroom, may be acquired later. This notion of the developmental order of grammatical features is the underlying assumption of grammatical complexity adopted in this study. By definition, this approach taps into the developmental complexity construct to some extent, at least in a theoretical term.

As Pallotti (2015) notes, however, developmental complexity is a somewhat circular notion because it assumes that the acquisition/production order determines complexity, without a theoretical ground on which the complexity can be operationalised. Indeed, it is a rather fuzzy notion without a clear definition of 'development' and a reference point as a criterion to measure development. However, a usage-based perspective provides a viable rationale for establishing a theoretical definition of developmental complexity, as discussed in this section. In addition, studies adopting a usage-based approach to syntactic development in writing use a reference corpus as a target register. Defining a reference corpus enables us to set a relative criterion against which the developmental complexity can be measured.

Biber, Larsson & Hancock (2023) conducted an assessment of the empirical adequacy of three types of studies using theory-based models of grammatical complexity. These categories revolve around distinct linguistic components being analysed: structural elaboration, system of the language, and texts. The first type aligns closely with the syntactic complexity studies outlined in Section 2.1. The second, termed 'system complexity', incorporates the notion of a language system, encompassing grammatical variations among languages or varieties (Biber et al., 2023, p.3). The third category of complexity termed usage-based or text complexity incorporates the notion of registers examining individual texts, registers, dialects, and even languages, based on analysis of a corpus of texts from the target language variety. This facet of complexity aligns with the explored definition of grammatical complexity in this thesis.

While I have argued that developmental complexity is a compatible notion with syntactic developmental research adopting a usage-based theory, I adopted the term 'grammatical complexity' for consistency to refer to the core construct of focus in this study. Grammatical complexity also adopts the theoretical and operational methods associated with developmental complexity discussed so far: the basic assumptions of a usage-based approach and the notion of a register as a target reference point.

2.3.2.3 Empirical Research of Syntactic Development in Academic Writing Based on Register Perspectives

The previous section has discussed some theoretical points of view on form-function mapping in English. These discussions offer a foundation for an alternative view of how English writers employ the language differently in their writing compared to other registers, such as spoken discourse. Moreover, these discussions indicate that specific language forms are of more research value in English academic written registers. Observations have been made from a usage-based perspective on second language writing development (L2) (e.g., Biber et al., 2021; Kyle & Crossley, 2017), which will be further discussed in the following.

The shift of research focus to phrasal complexity discussed in Section 2.3.1.2 indicates a more diversified interest and measurement in syntactic and grammatical complexity in writing. Along with this, studies adopting a usage-based approach to SC development research in L2 writing also call for attention to the syntactic function, on top of grammatical forms. Finally, within a usage-based perspective, the assumption of the interactions between language and situational factors leads to discussions on register variation. In this framework of examining linguistic variation, L2 writers would be expected to progress in their use of markers of typical academic written discourses. Much research drawn on the grammatical features marking this register distinction has shown the notion of register variation is useful in interpreting the grammatical variations in L2 writing development (e.g., Biber & Gray, 2010; Staples et al, 2016).

In light of the notion of register, syntactic complexity varies between spoken and written texts (Ortega, 2015, p.90). Researchers have identified nominalization as one of the phenomena associated with written complexity. The concept of nominalization, a key transitional marker from oral to written discourses, has been a topic of attention for some time (Myhill,2009; Biber & Conrad, 2009). Furthermore, a body of research comparing the grammatical complexity of spoken and written registers provides insights into the typical syntactic characteristics found in written academic discourse (e.g., Staples et al., 2016; Carter, 2004; Hughs & McCarthy, 1998).

Biber and Gray’s (2016, p.78) overview of the distinctive grammatical landscape of academic writing suggests that three-word classes are especially salient in written 'academic' writing: nouns, adjectives, and prepositions. These grammatical classes are more frequent in academic prose than in other registers. In addition, many related specific features are especially characteristic of academic prose (e.g., nominalisations, noun phrases with multiple modifiers, stance noun + of-phrase). In contrast, verbs, adverbs and adverbials are usually less common in academic prose, even though the specific types of these grammatical classes serve unique functions in academic prose. As for a notable construction, complex phrasal embedding is more commonly produced in the specific context of formal writing (Biber et al., 2020, p.6)

In terms of the syntactic function and structural form, academic writing prefers phrasal rather than clausal structures functioning as constituents in noun phrases (Biber & Gray, 2016, p.94). The use of prepositional phrases as post-nominal modifiers is a clear case of this type: they are prevalent in academic writing but rare in conversation. Similarly, attributive adjectives and nouns as nominal pre-modifiers behave as phrasal constituents embedded in the noun phrase. Nominalisation makes texts more concise, creates textual cohesion, and helps the reader follow the text (Staples et al., 2016).

Biber conducted a new factor analysis to explore register variations in academic registers (Biber, 2006). The findings inform functional dimensions of linguistic variation that are distinctively important in the university setting (Biber, 2006, p.211). One notable finding is the ubiquitous distribution of stance features across all four functional dimensions in academic university discourses. This general importance of stance in university registers indicates a greater reliance on stance features overall compared to the findings from using more general corpora consisting of other registers than academic registers, such as conversations. Two methodological implications for this study arise from it: first, new factor analysis may reveal the unique functional dimensions in the corpus of interest. Second, stance function and relative linguistic features may be important linguistic features to be explored for writing developmental research.

The last point to be noted in this review is the relativity of this type of register comparison. For example, Biber (2006) found stance features are important defining characteristics for all four functional dimensions found from a factor analysis in the university spoken and written registers. This ubiquitous distribution of stance features across all the dimensions contrasts with the MD analysis by Biber (1988), which has only one such stance-related dimension. Biber interprets it as an indication of the variety in the kinds of stance expressed in university registers, as well as a greater reliance on stance features overall. This is notable in that the stance function is often associated with spoken registers.

In addition, research on the syntactic sophistication of VACs has analysed the developmental patterns and writing quality of second language (L2) writers using a target register as a benchmark. These studies have identified a positive correlation between advanced writing proficiency and syntactic sophistication, characterized by a higher occurrence of uncommon words in the target constructions. Although a consistent definition of formulaicity is yet to be established, it is worthwhile to explore whether a specific set of infrequent yet academically valuable formulaic constructions gain prominence over time. The aspect of usage-based approaches to stance constructions is further discussed in Chapter 5.

This section has discussed an alternative perspective of SC underscoring the importance of considering the register-specific syntactic patterns, which link the intuitive appeal of constructions mapped onto communicative functions observed in written discourses. In academic discourse, there is a preference for concise expressions with more nouns and noun modifiers, which exemplify the concept of phrasal compression (Staples et al., 2016). This preference arises from the pursuit of efficiency and compression in written contexts, with more condensed nominal phrases assisting readers in achieving these goals. The following section concludes the implications of the empirical findings of the studies on L2 learner writing development, which leads to the research questions of this thesis.

2.4 Summary and Conclusions

In this chapter, I have argued that longitudinal data adds valuable evidence to our understanding of SC development in L2 writing. Building upon previous academic written register research, the current study aims to explore the longitudinal data collected in an intensive English language programme at an American university with the following overarching aims.

This thesis investigates the following overarching research questions:

> What are the key grammatical complexity and sophistication features exhibited in academic texts written by L2 English learners, and how

do they vary longitudinally?

Table 2-2 further details the specific research questions and the relevant chapters.

Table 2-2. Research Questions

Chapters	RQs
Chapter 4	1a. What are the grammatical complexity features that are associated with the functional dimensions typical of academic written registers in L2 English writing? 1b. What are the key communicative functions in the dimensions of L2 English writing? 1c. How do the major grammatical complexity features associated with the functions typical of academic written registers vary over time in L2 writing?
Chapter 5	2a. How do the frequencies of the two constructions (noun-that SV; adjective-that SV) vary over time?
Chapter 6	3a. What are the verbs that are highly formulaic in the *stance* constructions in the L2 writing? 3b. What stance functional dimensions are associated with *highly sophisticated* verbs (VACs that are infrequent but with high collostruction strength in academic registers)? 3c. How do the sophisticated constructions' association strengths vary over time?

The overarching theoretical frameworks for all of these broad aims and research questions are delineated in Chapter 3. Chapter 3 introduces the PELIC corpus, an L2 longitudinal learner written corpus and presents the detailed research design of this thesis to achieve the aims above by exploring the main corpus using corpus linguistics methods. Chapters 4, 5, and 6 discuss the analyses under thematic headings of method frameworks: exploratory factor analysis, collostructional analysis, and inferential regressions of SC features.

Chapter 3 Methodological Overview: Investigating Syntactic Development in Second Language Writing

3.1 Introduction

This chapter consists of six sections to establish the research design for the present study, describing academic learner writing, defining key linguistic constructs, and outlining the methodological design. The primary aim of Section 3.2 is to introduce the PELIC corpus, the main L2 written data of this thesis, defining its characteristics as a variety of L2 longitudinal learner corpora. Previous approaches and findings, and implications of analyses using longitudinal learner written corpora are discussed where relevant.

Section 3.3 presents the general research design of the analyses, including theoretical discussions of the overall rationales of the methodologies adopted in this thesis, which applies to the analyses to be detailed in Chapters 4, 5 and 6. More specifically, it discusses the methods of processing, analysing and interpreting the data, with consideration of research ethics relevant to the research aims of this thesis. In order to investigate development in longitudinal data, this study designs considered external criteria, including English levels assigned to text, target student register, and empirical findings from the previous literature. The criteria for interpreting syntactic development in writing are also detailed in this section.

Section 3.4 discusses the adequacy of the linguistic features as an operationalised construct of linguistic complexity and sophistication. The sub-sections 3.4.3 presents two types of preliminary analyses done on the longitudinal data used in Chapters 4, 5 and 6. Section 3.4.3.1 reports the tagging accuracy check of the data while Section 3.4.3.2 presents the thematic analysis of learner language identified during the tagging accuracy check. These two sections provide information about the sample data, which establishes the basis for the reliability of the analyses and contextual information for qualitative analyses.

Final conclusions are reached in Section 3.6, which previews the specific research questions and structures to be discussed in Chapters 4, 5 and 6.

3.2 The L2 Written Academic Corpus

The written texts produced by learners in university intensive English programmes are important evidence of their performance, which can provide evidence of their development in academic writing. Despite the potential disparities in essay genres typically prevalent within curricula in language programmes in contrast to those assigned in disciplinary domains, their intrinsic value persists, stemming from their function in acquainting students with the art of structuring persuasive discourse, which is pervasive across a range of academic disciplines. In that regard, the writing tasks in a language programme equip students with a skillset expected to be used in a wide range of assignment genres.

The primary data for this study is sourced from the University of Pittsburgh English Language Institute Corpus (PELIC), which is a longitudinal L2 learner corpus collected from 2005 to 2012 within an intensive English language programme at Pittsburgh University in the U.S (Juffs et al., 2020; Naismith et al., 2022). PELIC includes texts written by students with various L1 backgrounds, encompassing writing, grammar, reading, and speaking classes, collected over 18 semesters from 2006 to 2012. 1,115 texts were selected consisting of 82 students' written assignments in writing classes over three to five semesters. Sections 3.2.1 and 3.2.2 discuss the rationale of the longitudinal data selection and describe the profile of the selected corpus. Section 3.2.3 discusses the writer-specific factors associated with the data while Section 3.2.4 presents the situational factors associated with the data.

3.2.1 Rationales of Longitudinal Data Selection

This section discusses the rationales underlying the decisions made on data selection for quantitative analysis, including sample size adequacy and coding schemes for data processing.

First, the data selection considered the potential influence of the length of the analysis unit and data size on statistical analysis. Previous empirical research has discussed the guideline for determining the minimum text length in Multidimensional (MD) analyses, the primary research method used in this thesis, which employs a factor analysis. General recommendations for case observations for exploratory factor analysis solutions range from a minimum of observations ranging from 300 (Guadagnoli & Velicer, 1988) to 100 (Bryant & Yarnold, 1995). It should be noted that Guadagnoli and Velicer (1988) provide advice for deciding on an adequate sample size, mainly based on two considerations: the representativeness of one or more variables for a construct and the magnitude of component loadings. It should be desirable if the samples contain clearly defined variables which are all highly loaded to a component or a construct. However, if there are enough variables that represent the underlying construct or if even a few variables have very high loading on the construct, either possibility has a good chance of being adequate in terms of sample size.

A second consideration of data sample adequacy involves minimum text length. While short texts can be analysed using various methods, including multi-dimensional analysis (MDA), they pose significant challenges for quantitative corpus data analysis. As Biber & Gray (2013) note, short texts often result in unreliable normalized rates, particularly for rare features, and a high occurrence of zero values, which can negatively impact inferential statistics.

While MDA can accommodate a large number of linguistic variables, reducing the minimum sample size threshold (Friginal & Weigle, 2014, p.83), the literature generally recommends more substantial text lengths for robust analysis. For instance, Biber (1988) excluded texts under 500 words, and subsequent studies (Biber, 1990; Biber, 1993; Biber & Gray, 2013) have supported a minimum text length of 100 words to ensure reliable normalised rates. L2 learner corpora research, considering the shorter nature of learner texts and the relatively low frequency of target linguistic features, has also adopted the 100-word minimum analysis unit (Crosthwaite, 2016; Friginal & Weigle, 2014; Biber, Gray & Staples, 2016; Yan & Staples, 2020; Gray et al., 2020).

However, the suggestion that 500-1000 words is a more suitable minimum for MDA and that 2000-3000 words is preferable for complex and infrequent features, raises concerns about the adequacy of the current dataset. As Clarke, McEnery, & Brookes (2021, p.148) highlight, standard MDA relies on relative frequencies of linguistic features, which are typically only reliable in text samples exceeding 1000 words. Given that the majority of texts in our corpus are under 1000 words, this limitation is particularly relevant to the present study.

Biber (1993) extensively discusses the relationship between text length and the reliability of grammatical feature frequencies. His work on "Representativeness in Corpus Design" emphasises the importance of adequate sample size for accurate linguistic analysis. While the current study employs a 100-word minimum, the potential implications of this decision on the reliability and generalizability of the findings warrant careful consideration (see, for relevant discussions, pages 157 (Section 4.5.3), 212 (Section 5.4.4) and 261-262 (Chapter 7)).

Table 3-1 presents the number of texts and words in the written PELIC corpus, including those meeting the 100-word minimum. The 1,115 selected represent a longitudinal subset of the overall corpus. While these texts were chosen based on writer consistency, the potential impact of their relatively short length on the MDA results requires careful consideration and discussion.

Table 3-1. Texts of 100 words or more that were produced from writing classes

duration[1]	Course 1	Course 2	Course 3	Course 4	Course 5	Sum (average per text)	Minimum/ Maximum text length
student	601	330	71	10	1	1013	
text	2615	3052	1068	210	33	6978	
words	894,420	1,055,237	398,098	72,966	17,235	2,437,956 (349.3)	100 / 2432

Another important consideration to be made in sample selection is the *balance* of a corpus, which may be a potential factor influencing linguistic variations (Nelson, 2010, p.60). Egbert et al. (2022), who distinguished between the target domain and the operational domain in corpus design, provide a more nuanced understanding of a corpus. The target domain represents the ideal collection of texts that perfectly captures the specific language use we're interested in studying. Imagine a corpus aiming to analyze scientific articles – the target

[1] duration: the number of semesters in which the students submitted at least one text (the duration of data contribution by each student)

domain would encompass all scientific articles ever published. However, practical limitations like time and resources necessitate an operational domain, a feasible subset of the target domain. In this scientific article example, the operational domain might only include articles published in the last five years from specific journals due to time and resource constraints. When selecting texts within this operational domain, researchers consider relevant domain characteristics to ensure the corpus best reflects the language use found in the target domain. This distinction helps us understand the ideal scenario (target domain) while acknowledging the practical limitations of corpus creation (operational domain).

The authors argue that these domain considerations have an impact on the accurate measurement of quantitative linguistic analysis, which cannot be addressed by solely increasing sample size (Egbert et al., 2022, pp.159-160). However, the PELIC corpus composition is uneven regarding learner-specific variables such as age, first language, and proficiency. Their influence on analysis results may be subject to the research focus, as to how much variability across these variables is allowed within the research design. In that regard, some studies focus on the usage patterns similar across the learner corpora consisting of academic prose by different L1 backgrounds (e.g., Biber, Conrad & Reppen, 1998, p.145) while others emphasise variations of L2 writing related to learners' L1 backgrounds (e.g., Kobayashi & Abe, 2016). The present study intends to keep the sample unbalanced as is to maximise the sample size. However, considering the potential influence of this factor on the analysis result, a mixed-effects model is employed to make inferential estimations of the linguistic patterns controlling this potential moderating effect on linguistic variation.

The last consideration on sample selection was the uneven text length per each text across the texts in the corpus. The text lengths are considerably uneven, which may cause an inaccurate representation of any feature frequency data in the PELIC corpus. This wide difference between text length and relatively short average text length as a whole also raised concern about using much longer texts as a reference, such as the MICUSP corpus. Therefore, some linguistic variables especially susceptible to text length, such as Type-token ratio, were removed from the research design, and reference data of similar text length produced within similar situational contexts was chosen (full details are in Chapters 4-6). These points are revisited in the methods sections of each chapter where these points have more direct relevance.

Finally, a coding scheme was established to categorise relevant variables systematically for quantitative analysis. Establishing a systematic and transparent coding scheme of the data is essential for consistently retrieving relevant contextual background information such as learner background. This study uses the coding scheme provided in the metadata of the PELIC corpus. In addition to these established codes, another code is necessary to trace the temporal order of the written production by individual writers. Since students usually contributed three or four semesters during the data collection points throughout 18 semesters, students' starting and ending points of data contribution vary. The new codes assigned to each text show the relative temporal order of the texts produced by the same student. Table 3-2 shows the complete coding scheme of the corpus metadata used in this thesis.

Table 3-2. The corpus data code scheme with the examples from the samples

text	student	L1	English Level	Question	Text length	Course
22521	dq9	Chinese	4	2968	254	1
23684	dq9	Chinese	4	3081	199	1
24594	dq9	Chinese	5	3226	464	2

It should be noted that data collection occurred throughout the semesters. Therefore, the production of the text was continuous, and the coding scheme simplifies the temporal points of writing production for ease of quantitative analysis.

3.2.2 The Chosen Sample Data

Based on these considerations regarding sample selection discussed so far, the present study has selected 1,115 texts, written by 82 students over a period spanning three to five consecutive academic semesters. The texts are selected under the conditions as follows:

1. Texts of a minimum of 100 words

2-1. (For the main longitudinal analyses) The first version of texts, excluding the drafts with any revisions based on teachers' feedback

2-2. (For the two pilot analyses on the revision process) The texts with at least one revised version, regardless of their duration of contribution across semesters

3. Texts produced by a student who contributed at least one text throughout three, four, or five consecutive semesters

The focus of this analysis is linguistic changes across students' temporal order of learning (the semesters that each learner took). Since the data was collected throughout 18 semesters, but students usually contributed for from three to five semesters, the texts were coded according to the semester in which it was produced concerning each writer's course of study, such as in their first, second, third, or next semester. The semesters in which each student produced the selected corpus were coded as 'Course' in order, providing a pivotal reference point to track linguistic changes over time. Table 3-3 provides an overview of the composition and size of this sample, including information on the number of texts in each Course within the column 'sub-corpora'.

Table 3-3. Composition of the Pittsburgh English Language Institute Corpus used in the study (1st version of drafts only)

Sub-corpora	Number of texts (students)	Total word count	Average word count[2]
Course 1	328 (82)	80,322	246
Course 2	408 (82)	156,454	383
Course 3	335 (82)	157,763	471
Course 4	37 (11)	17,108	462
Course 5	7 (1)	3,792	542
Total	1,115	415,439	420.8

[2] The minimum text length is 100 words.

The common duration of text contribution was three consecutive semesters, with 908 out of 1,115 texts collected throughout three consecutive semesters. The inclusion of texts produced by a minimum of three academic terms enables the longitudinal tracking of writing development in a natural classroom environment, thereby resembling more genuine academic writing activities.

Figure 3-1 presents the corpus composition by both production time points (Course) and L2 proficiency level (L2 level). The L2 level of students increases as their study proceeds, with most of the students given level 5 in course 4, the highest proficiency assessment frame accredited to the students in the language programme. On the other hand, the majority of students are given level 4 in the whole PELIC corpus. Notably, this gradual improvement of their proficiency level may exert an influence on interpreting students' longitudinal development.

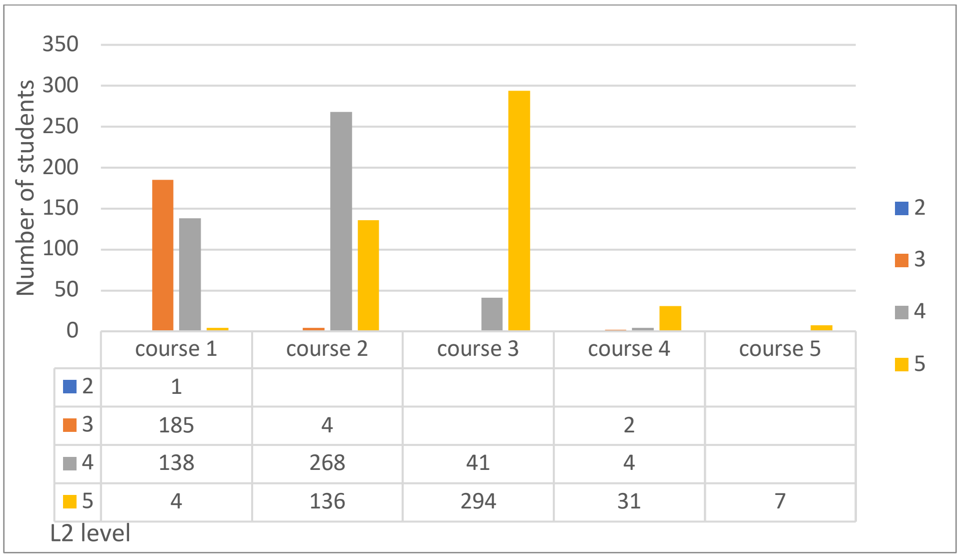

Figure 3-1. Number of students by L2 levels and course of study

The information about learners' L2 English levels in Figure 3-1 comes from the PELIC corpus metadata, which provides information on learners' English proficiency levels based on the Common European Framework of Reference (CEFR), ranging from A2 to B2/C1. The proficiency levels in the dataset correspond to CEFR A2 to CEFR B2/C1, as indicated in Table 3-4.

In Figure 3-1, the average values for Courses and Levels correspond with each other. This means that Course 1 has an average level of 3, Course 2 has an average level of 4, and Course 3 has an average level of 5. For a quantitative analysis, this may cause multi-collinearity deriving from high correlations between these two variables. Despite the potential for a statistical issue, as the level assignment was conducted for every semester, the Course and the Level were considered as two separate variables in this study.

3.2.3 L1 backgrounds and L2 English levels

This section describes two learner-specific variables identified within the PELIC corpus, including first language, and L2 English proficiency. The PELIC corpus consists of learners from 30 L1 backgrounds, but for this study, the selected texts are associated with nine linguistic backgrounds including Arabic, Chinese, Japanese, Korean, Spanish, and Turkish, as shown in Figure 3-2, which depicts the distribution of texts according to the L1 background.

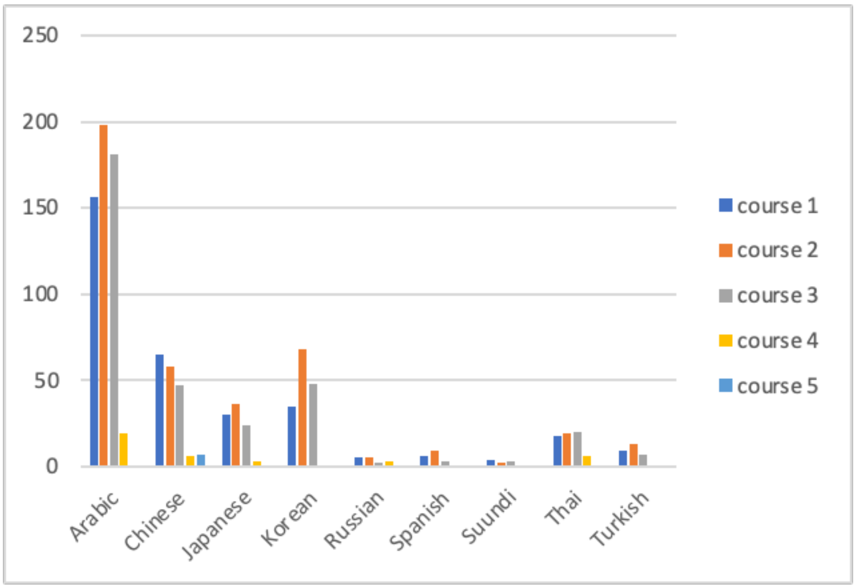

Figure 3-2. Number of texts by L1 background (total 1,115 texts)

Figure 3-2 shows that the variations of the L1s across Courses were not patterned. As the number of texts written by the different L1s is not balanced, I entered this as a random variable in mixed effects models during analysis (see Sections 4.5, 5.3, and 5.4), not imposing any theory-informed hypothesis regarding its influence on grammatical variations.

Table 3-4 shows a coding scheme for English proficiency levels assigned to students at every semester in the intensive English programme. The corpus metadata also includes the CEFR levels corresponding to the assigned level, which is also shown in Table 3-4.

Table 3-4. The coding scheme of English proficiency level in reference to CEFR

English level assigned by the language programme	Level description	CEFR level
2	Pre-Intermediate	A2/B1
3	Intermediate	B1
4	Upper-Intermediate	B1+/B2
5	Advanced	B2+/C1

Considering the participants' age (18 and above) and the university educational context for the data collection, learners contributing to the data are assumed to possess literacy skills in their respective L1s. Literacy in the first language is a crucial consideration as research suggests that the skills acquired in L1 can have an impact on L2 writing, whereas a lack of literacy in L1 may hinder the L2 writer's abilities (Carson & Keuhn, 1994). Although resemblances in usage patterns might be discerned among learner corpora encompassing a range of native language (L1) foundations, it is imperative to acknowledge the potential existence of divergences stemming from the distinct L1 backgrounds characteristic of individual learners.

The L2 English levels 3, 4 and 5 are the general proficiency levels assigned by the language programme, using both the Michigan English Placement Test (Michigan EPT) for listening, grammar, reading, and vocabulary tests and the in-house writing and listening tests (Juffs, 2020: 63-66). While level 3 is where the process writing approach is introduced to compose sentences and paragraphs, Level 5 is the final stage, where students progress to the level of writing academic papers, including the mechanics of references and appropriate rhetorical patterns (Juffs, 2020, p.205).

The writers' proficiency assigned by the language programme, as indicated in the metadata of the English level introduced in this section, has a more situational description provided by Juffs (2020), one of the investigators of the PELIC compilation project. Drawing on that, the following section describes the assessment, placement and teaching in the language programme, which contributes to the understanding of the situational context where the writers contributed their texts to the PELIC corpus.

3.2.4 Learner-centred Approach to Learning Writing

Based on the theory of systemic functional linguistics (Halliday & Matthiessen, 2014), Ortega argued that the complexity of L2 production is influenced by the communicative purposes of the task the learner aims to address (2003, 2015). Considering its potentially significant influence, this section discusses the situational background of the communicative purposes inherent in the PELIC corpus data.

Juffs (2020) notes that the language programme is run by a learner-centred approach where learners engage in discussions on topics that are personally relevant to them (Juffs, 2020, p.212). This approach forms the basis of teaching philosophy in conjunction with a process writing approach, as indicated by the holistic rubric of the writing component of the placement test (Juffs, 2020, p.65) in Table 3-5.

Table 3-5. The rubric descriptor for the English proficiency assessment

Level 3	Students are mostly focused on creating sentences and short paragraphs. The IEP seeks to introduce the idea that writing is a process and that texts may be revised. Genres at this level include definitions and narrative.
Level 4	The idea of longer pieces of writing is introduced, including thesis statements (central ideas). Genres include giving instructions and

	summarising.
Level 5	It was the highest level in the IEP at the time of data collection, where the research paper was introduced, with essential functions such as comparison and contrast, cause-effect, and processes. The goal of this highest level is to prepare students for university writing tasks (Juffs, 2020, p.205).

The writing test component in the course is specifically designed to align with the target coursework in academic writing across disciplines, focusing on the core elements of academic writing skills valued by in-house teachers (Juffs, 2020, p.65). For example, the rubric description for level 5, the highest level accredited by the programme, states that this level is aimed at university writing tasks.

Interestingly, the writing placement test results do not correlate with students' proficiency level, even though the writing score was part of the source of the placement decision (Juffs, 2020, p.65). Juffs (2020) notes that this discrepancy between proficiency and writing score level accredited to students may have to do with the complex nature of academic writing quality that may not necessarily correlate with other components of language proficiency. Along with these test rubric descriptors, the developmental writing process was likely influenced by the programme's teaching philosophies, including a process writing approach and a learner-centred approach.

The plots for L2 levels by Courses and L1s by Courses have been considered as mediating variables, but they have not been entered into the statistical analyses discussed in this thesis with a theory-informed hypothesis. Both L2 levels and L1s are not balanced in terms of sampling the data, as they were not the major variables to be explored. However, as they are likely to influence any linguistic variations, they were considered variables to be controlled in assessing the statistical significance of the major variable (Course) on the grammatical variation. The fact that these variables are not controlled nor are the data balanced is both the strength and limitation of the studies discussed in this thesis. Quantitative analyses require sophisticated and complex statistical processing to address relationships between these variables. However, as these situational factors are in many cases multidimensional and interact on a continuum rather than a dichotomous way, it was essential to complement the quantitative analyses with qualitative inspections of the variables in textual context. That is, these situational variables were controlled in statistical analyses, but their interaction with the grammatical variations is further examined in the textual context.

3.3 Research Design for Investigating L2 Learner Writing Development

This study aims to determine if existing theories and research on syntactic complexity in academic writing align with the development of L2 writing skills observed in students enrolled in a U.S. university language programme. By examining longitudinal data, I seek to contribute to the understanding of how lexico-grammatical variations indicate L2 writing progress.

To accomplish this, I employ a corpus-based approach. I formulate hypotheses about specific language features based on previous research and then test these hypotheses using corpus data. Combining quantitative and qualitative methods, I identify and analyze significant patterns in the data. Section 3.3.1 justifies the use of quantitative measures (frequency) to track linguistic development and introduces a functional perspective on language feature selection. Section 3.3.2 outlines the overall analysis framework, as an overview of Sections 3.3.5, 3.3.6, and 3.3.7.

Section 3.3.3 discusses linguistic variables of the longitudinal written corpus analysed in this study. By examining the same writers over time, this longitudinal corpus allows us to track the development of L2 writing skills. My primary goal is to determine if established theories about linguistic complexity in L2 writing match the patterns observed in the longitudinal data when viewed through the lens of how language forms convey meaning (form-function mapping).

Section 3.3.4 considers writer-related and contextual factors like first language and overall language ability as potential influences on writing development. These factors are statistically controlled to isolate the impact of language features on writing improvement. While I do not make specific predictions about how these factors change over time, their potential influence is taken into account in making interpretations an qualitatively inspected as a post hoc analysis.

3.3.1 Rationales of the Research Design

This study is motivated by the aim of investigating the extent to which previous theories and findings on syntactic complexity in academic writing are consistent with L2 longitudinal development seen in written essays within a university language programme in the United States. The objective is to contribute longitudinal evidence to the understanding of syntactic variations as indicators of L2 writing development. To achieve this, the study adopts a corpus-based approach, in which the researcher formulates hypotheses regarding specific linguistic phenomena based on previous studies' perspectives on language or knowledge, and subsequently tests the likelihood of these hypotheses based on observations made in the corpus data. The present study combines this hypothesis-driven approach with quantitative methods to capture interesting patterns within a broader scope. Once these patterns were confirmed as significant in the corpus, they were further investigated qualitatively in context. This subsection discusses theoretical and methodological justifications for quantitative variations (frequencies) as the main evidence of syntactic development.

A fundamental characteristic of corpus linguistics approaches to linguistics analysis is the utilization of numerous texts from real-world sources, as opposed to elicited sentences. This view diverges from the Chomskian perspective on language, which places greater emphasis on the researcher's intuition and analytical perspectives regarding underlying language systems than empirical evidence (Chomsky, 1988).

The corpus linguistics approach allows for quantitative analysis of a large language sample, facilitating consistent comparisons across different texts (McEnery, Xiao, & Tono, 2006; McEnery & Hardie, 2011). Additionally, the extensive collection and utilization of corpora enable objective observations of linguistic phenomena, helping to reduce researcher biases that can emerge when working with small, unbalanced data subsets that favour the researcher's preferences. The quantitative approach to text analysis employed in corpus linguistics studies is particularly valuable in mitigating confirmation bias, which researchers are prone to when selecting data that readily supports or aligns with their research hypotheses (Nickerson, 1998, p.188).

The *corpus-based approaches,* as opposed to *corpus-driven approaches*, consider concepts and assumptions formulated within the researcher's mind based on a priori reason or empirical evidence and use the corpus data to test these assumptions (McEnery & Hardie, 2011). As argued by McEnery and Hardie, corpus-based studies often use empirical linguistic data to prove a certain level of hypothetical views on language, rather than disapproving of the researchers' evaluation or interpretation of the data, as in corpus-driven approaches. The corpus-driven approaches are rather extremes of corpus linguistics philosophy represented by Neo-Firthians, who view linguistic data as the embodiment of linguistic theories (Tognini-Bonelli, 2001, p.84) and strive to describe language in a text as it is.

On the other hand, the corpus-based approaches are more compatible with the view of a hypothesis-driven approach with the underlying assumption that a priori hypothesized L2 linguistic developments can be confirmed through empirical data when they yield expected results. For example, a corpus is of use for testing falsifiable hypotheses about language (McEnery & Brezina, 2022, p.43). One of the fundamental advantages of this test-driven corpus-based approach to linguistic analysis is its compatibility with statistical analysis, which assists in assessing the plausibility of the researcher's hypotheses and the likelihood of outcomes under those hypotheses.

However, despite efforts to maintain objectivity and minimize biases or subjective/qualitative evaluations, complete neutrality may not be feasible or even desirable for educational enquiries. In other words, while descriptive exploration of linguistic variation may be strategic and appropriate in corpus linguistics studies, developmental research requires evaluative elements to a certain extent. To address this dilemma for social science enquiry, McEnery and Brezina (2022) argue for a blended approach to social science drawing on both naturalist and anti-naturalist, where the social construct is observed based on objective evidence (pp.113-114). The consistency and objectivity in data treatment enhance the transparency of research procedures and encourage constructive criticism. By accumulating and cross-checking observation data, biases in data interpretation can be routinely identified and eliminated. This iterative process of studying and testing contributes to valid attempts to seek knowledge (Popper, 2002, p.18). In this regard, corpus-based studies have the potential to not only confirm or reject hypotheses but also contribute to the refinement of the theoretical frameworks upon which they are based (McEnery & Hardie, 2011, p.6). In this way, the quantitative approaches to development research, by using the data on the presence or absence of specific linguistic features or patterns can provide more robust evidence to support claims about complex developmental trajectories (Durrant et al., 2021, p.34).

Building on the foundation of previous research (outlined in Section 2.3.2), this study posits that specific linguistic features serve particular communicative functions within academic writing. This assumption is rooted in the understanding that writing, as a later-acquired skill compared to speech, follows a developmental trajectory marked by increasing reliance on phrasal structures (Biber et al., 2011, 2021a). While quantitative analysis provides valuable insights into linguistic patterns, it is essential to acknowledge its limitations. Changes in language use over time do not necessarily equate to complex linguistic development. Furthermore, reducing rich textual data to numerical values can obscure potential alternative explanations for observed phenomena (Durrant et al., 2021). The absence of a specific feature might reflect strategic language choices rather than a lack of acquisition. To address these limitations, this study triangulates the quantitative analysis with qualitative inspection. Quantitative analysis serves as a starting point for identifying potential patterns, while qualitative analysis delves deeper into the underlying linguistic processes. This exploratory inspection aims to uncover the communicative functions governing the observed linguistic variations.

3.3.2 Analytical Framework

This section outlines the methodological approach employed in this study to investigate L2 writing development. It details the integration of various statistical methods and their application to corpus data. Within the analytical framework of this study, various quantitative and statistical techniques were combined to explore longitudinal written data. These exploratory techniques include factor analysis and principal component analysis (PCA), and the post hoc statistical analyses include linear and Bayesian mixed-effects modelling, and collostructional analysis. Table 3-1 provides an overview of the specific analyses conducted in each chapter, including the linguistic variables investigated, statistical methods applied, and corpora utilised.

Table 3-6. Overview of studies discussed in this thesis

CH	Section	Linguistic variable (source)	statistical analysis method	main corpus	reference corpus
CH4	4.4	67 MAT[3] features (Nini, 2019)	factor analysis & cluster analysis (from MAT)	1115 PELIC[4] texts consisting of five Courses	MICUSP[5]
CH4	4.5	34 TAASSC[6] features (Kyle, 2016)	PCA[7] / EFA[8]	326 PELIC texts consisting of revised texts	
CH4	4.6	71 MAT + spaCy + TAASSC features	factor analysis & linear MEM	1115 PELIC texts consisting of five Courses	
CH5	5.4	V to-inf / V that SV (MAT)	collostructional analysis & non-linear growth model	326 PELIC texts consisting of revised texts	MICUSP
CH5	5.5	V that SV (MAT)	collostructional analysis	1115 PELIC texts consisting of five Courses	PELIC texts written in LEVEL 5
CH6	6.3	adjectival/prepositional dependents per nominal (TAASSC)	linear MEM[9]	1115 PELIC texts consisting of five Courses	
CH6	6.4	prepositions (MAT)	Bayesian MEM	1115 PELIC texts consisting of five Courses	
CH6	6.5	adjective that SV / noun that SV	Bayesian MEM	1115 PELIC texts consisting	

[3] Multidimensional analysis tagger (Nini, 2019)
[4] The University of Pittsburgh English Language Institute Corpus (Juffs et al., 2020)
[5] The Michigan Corpus of Upper-Level Student Papers
[6] Tool for the automatic analysis of syntactic sophistication and complexity
[7] Principal component analysis
[8] Exploratory factor analysis
[9] Mixed-effects model

		(MAT)		of five Courses	

Factor analysis and PCA were employed to identify underlying dimensions within the MAT and TAASSC feature sets, contributing to a more parsimonious representation of the data. The collostructional analysis technique allowed for a detailed examination of the development of specific linguistic constructions, such as V to-infinitives and V that SV constructions. Mixed-effects modelling was chosen to account for the hierarchical structure of the data (students nested within courses) and to model the development of linguistic features over time. Furthermore, Bayesian mixed-effects modelling was adopted for certain analyses to incorporate prior knowledge and to obtain more precise estimates of parameters.

The integration of these methods offers a novel perspective on L2 writing development by identifying underlying dimensions within the linguistic data, revealing patterns that might not be apparent through univariate analysis. In addition, collostructional analysis provides insights into the distribution and development of stance-related linguistic constructions, contributing to a deeper understanding of linguistic complexity. Finally, linear and Bayesian mixed-effects modelling allows for the examination of how linguistic features change over time, considering individual differences and contextual factors. By combining these techniques, this study aims to follow two-step quantitative analyses: identify key dimensions of linguistic complexity and then explore the relationship between these dimensions and L2 proficiency development. This research departs from traditional approaches to L2 writing research by adopting functional perspective in cycling analaytical framework. In other words, the application of multivariate analysis to identify underlying linguistic function to interprete them in terms of linguistic complexity in longitudinal data provides insights into developmental trajectories in L2 writing from a usage-based perspective.

To comprehensively examine the evolution of linguistic complexity and sophistication in L2 writing, a usage-based perspective was adopted. Figure 3-3 shows the analytical framework of this study, where the form-function framework serves as the guiding principle for selecting key linguistic features.

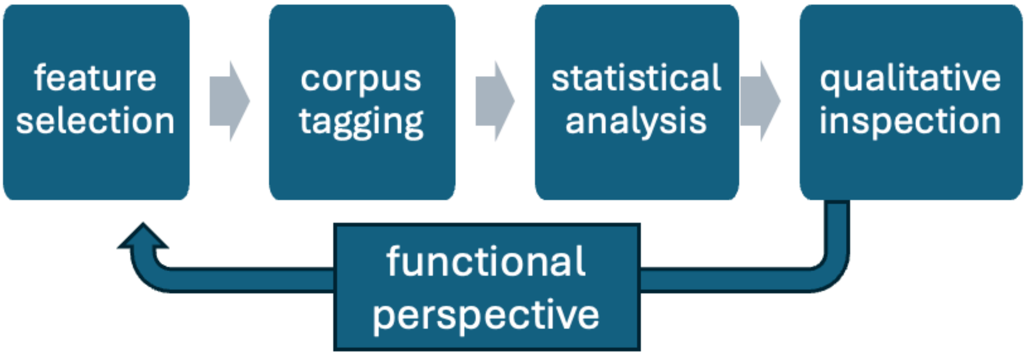

Figure 3-3. Analytical framework of the study

By examining the relationship between language form and communicative function, I decided on the key features of stance function, which are then presented in a subsequent study discussed in Chapters 5 and 6 focus. Previous research, such as Biber et al. (2021a) and Kyle (2016), informs the selection of syntactic structures for stance analysis. These structures also represent key areas of syntactic complexity in academic writing.

This approach allows for a granular examination of syntactic development, including both quantitative and qualitative perspectives. By combining quantitative and qualitative methods, this study aims to provide a more nuanced understanding of L2 writing development than traditional quantitative approaches. This approach aligns with the broader goal of analysing multidimensional variability and nonlinear developmental trajectories.

3.3.3 Definitions of Linguistic Variables

While the sample selection is of fundamental importance for ensuring valid analysis results, the selection of linguistic variables and tools and methods for extracting the information from the samples is also crucial (Biber et al., 2020, p.11; Anthony, 2013, p.147). This section sets out the rationales for establishing operational definitions of the primary linguistic variables for the research aims of this thesis.

As discussed in Chapter 2, recent studies have suggested the need for more fine-grained indices than traditionally used broad measures at the sentence or clause level (Lu, 2010; Ortega, 2003, 2009; Kyle & Crossley, 2018). As Kyle and Crossley (2018) noted, more specific and granular indices of syntactic complexity could clarify the relationship between syntactic complexity and L2 writing development. In addition, previous research has identified intriguing markers of writing development, such as nominalization. These writing-specific SC markers have increased attention given to situational and register-specific variables that influence language usage in corpora. This recognition acknowledges that the SC features in the corpus used for research may be interpreted in terms of SC features associated with written registers for writing development research.

As discussed in Chapter 2, recent empirical findings indicate that phrasal indices are a crucial indicator of academic writing quality in L2 writing development research (e.g., Taguchi et al., 2013; Parkinson & Musgrave, 2014; Biber et al., 2014; Staples et al., 2016; Kyle & Crossley 2018; Friginal and Weigle 2014; Crosthwaite 2016). The fact that phrasal indices prove to be a better predictor of written complexity in some studies (e.g., Kyle et al., 2021) indicates that the traditionally operationalised global measures of SC do not fully reflect the structural complexity in written registers. This variation of syntactic complexity in registers highlights the importance of not relying solely on global complexity measures for linguistic description (Biber et al., 2020a).

What is commonly observed in the research designs of these studies is two structural parameters, namely grammatical type and syntactic function. Biber et al. (2011; 2020a) argue that studying directly the structural/syntactic distinctions found in English writing is essential to understanding the complexities of written discourses produced at different developmental stages. The comprehensive framework of English linguistic structural form and syntactic function is shown in Table 3-7 (See Appendix 3-3 for the specific lists of grammatical categories of each syntactic function).

Table 3-7. Form-function taxonomy adopted in a study exploring syntactic complexity (adapted from Biber et al., 2011, p.19)

Grammatical type	Syntactic function	Example
Finite dependent clause	Adverbial	She won't narc on me, because she prides herself on being a gangster.
	Complement	I don't know how they do it.
	Noun modifier	That's one thing that bothers me right now about my job.
Nonfinite dependent clause	Adverbial	To verify our conclusion that the organic material is arranged as a coating around the silica shell components, thin sections of fixed cells were also examined.
	Complement	The main effect of grades has consistently been found to be the best predictor of future achievement.
	Noun modifier	The results shown in Tables IV and V add to the picture…
Dependent phrase (nonclausal)	Adverbial	Alright, we'll talk to you in the morning.
	Noun modifier	Class mean scores were computed by averaging the scores for male and female target students in the class.

Numerous studies have investigated various syntactic features in L2 writing data to assess L2 writing proficiency (Taguchi et al., 2013; Parkinson & Musgrave, 2014; Biber et al., 2014; Atak & Saricaoglu, 2021; Sarte & Gnevsheva, 2022). Overall, these studies have identified consistent patterns where the frequency of phrases increases in more advanced writing, indicating a developmental progression. As noted by Durrant et al. (2021), previous research findings suggest that L2 writing development is more reliably observed in terms of quality rather than time.

Albeit smaller in the number of studies, longitudinal studies have also contributed to understanding the usage patterns of phrasal and clausal complexity features in various contexts, such as L2 university student writing (Biber et al., 2020), L2 secondary school students' essays (Kyle et al., 2021), and L2 learners' TOEFL writing answers (Gray et al., 2019). These studies demonstrate that phrasal features become more prevalent in advanced stages of learner writing development. Gray et al.'s (2019) longitudinal study further reveals a significant task effect, indicating that the choice of language is influenced by both the task type and the writer's target language proficiency. Therefore, longitudinal evidence should be interpreted within the situational contexts of language production, taking into account factors such as task types.

In addition, the current study also incorporates the notion of form-function mapping drawing on a usage-based theory to study the multidimensional nature of syntactic complexity. As noted by researchers, syntactic complexity is a multidimensional construct that multiple measures can better capture (Biber et al., 2021a; Norris and Ortega, 2009; Bulté & Housen, 2012, 2014).

This thesis adopts an exploratory approach to the syntactic patterns emerging from a comprehensive set of syntactic features defined at both grammatical form and syntactic functions, the granularity of which reflects the register-specific communicative functions associated with them in terms of the academic written register. The 73 primary linguistic variables are based on 68 grammatical features identified by Biber (1988). The frequency data of these features in the longitudinal PELIC texts were obtained by the Multidimensional Analysis Tagger (MAT; Nini, 2019), which replicates the tagging scheme of the features identified in Biber (1988). These 74 linguistic features used in this study are found in Appendix 3-1.

It should be noted that the texts were not cleaned for grammatical, spelling and punctuation errors before annotation, which has likely affected the accuracy of the POS tagging and the syntactic parsing. In order to ensure the adequacy of the linguistic variables analysed in this study, I conducted accuracy check of the POS tagging and syntactic parsing of these linguistic variables. The result and assessment of the accuracy check for these linguistic variables are discussed in Section 3.3.1.2.

3.3.4 Definitions of Writer-specific and Situational Variables

The broad distinction between spoken and written registers as an underlying rationale for the linguistic features selection needs to be further fine-tuned to the specific register associated with the corpus for research. For instance, the influence of psychological difficulty or topic complexity interacting with structural linguistic complexity may explain L2 syntactic use beyond what can be accounted for by a usage-based definition of SC alone. Juffs (2020), in a study on semantic-syntactic complexity development, demonstrates a preference for verbs associated with phrasal dependents over verbs followed by clausal dependents in L2 written essays produced in an Intensive English program at a U.S. university. This preference for structurally simpler constructions is observed across texts written by writers at different English proficiency levels, while the frequency of both constructions increases with the writers' proficiency. These findings suggest that the formal complexity inherent in linguistic structures influences learners' choices. The preference for simpler language structure, as illustrated above, indicates the importance of considering writer-specific factors, which cannot be solely explained by exposure to target input. L2 adult learners, for instance, rely on their L1 skills as foundational knowledge for L2 learning (Koda, 2007), whereas writers from distinct L1 backgrounds may encounter challenges in transferring their L1 assets to L2 learning.

Moreover, operationalizing SC based on a usage-based perspective, as discussed earlier, may not accurately represent the linear trajectory of L2 learning, as situational and writer-specific factors are likely to influence language use. Previous findings indicating the interaction between linguistic complexity and other factors suggest the multidimensionality of syntactic complexity development in writing, which has yet to be fully explored.

Both L2 proficiency and L1 background will be considered as mediating factors and will be addressed in the current study. To address the influence of these learner variables, a mixed-effects model was fitted for inferential analysis, as will be detailed in Chapters 4 and 6. This approach aimed to minimize the influence of learner variables, specifically L1 and L2 backgrounds, allowing for more reliable generalizations regarding the effect of time.

Situational factors in the research design include writing topics and time effects. The time variable is the key dependent variable that this study aims to explore;

as the longitudinal data is collected under no experimental constraint, this construct broadly encompasses the potential situational factors, such as teacher feedback on the revision process or lectures offered in writing classes.

Writing topics are defined as an independent variable that impacts the variation of the linguistic variable in the data. It's part of the causal pathway of an effect, and it tells you how or why an effect takes place. The metadata of the PELIC corpus includes the information regarding questions to each student's answer, as shown in Table 3-8.

Table 3-8. The type of questions asked for the writing assignments in the PELIC corpus

question_type_id	Question type
1	Paragraph writing
2	Short answer
3	Multiple choice
4	Essay
5	Fill-in-the-blank
6	Sentence completion
7	Word bank
8	Chart
9	Word selection
10	Audio recording

Among the ten question types, only question types 1, 2 and 4 were relevant to the selected texts for this analysis. Some example questions of the two predominantly frequently featured question types are:

- Question type 1: Paragraph writing

  ex) Write a well-organised paragraph about how you prepare for and celebrate a holiday in your country. (Q_id: 4213)

- Question type 4: Essay

  ex) Use your outline and Introduction to complete a five-paragraph essay explaining the process of celebrating one of your holidays. Be sure to use time clauses. Try to include at least one passive voice sentence. Check your writing before you submit. (Q_id: 4235)

As shown in Question types above, the categorisation of question type does not control topics, and topic pools are significantly wide, consisting of 506 unique question prompts. Therefore, it is not likely for the approximated genre to each sub-corpus to be heavily skewed toward a specific topic only. The topic effect, as considered the influential situational variable in the current study, was considered as random effects in analysis, which will be detailed in Chapters 4 and 6. In other words, these topics of the writing was considered as influenced more by personal choices rather than a predictable and systematic parameter that influences the overall trajectories of average developmental patterns, as there is a very wide range of topics represented in the selected texts used in the analyses.

In addition, paragraph writing (question type 1) appears much more frequently in texts produced at the early stages of study (235 texts with paragraph writing type texts 91 texts with essay type texts). However, such a trend reverses as the study progresses by the third course of study (170 texts with paragraph writing type texts and 157 texts with essay type texts).

The writer-specific variables include factors related to individual differences: writers' first language (L1) background and second language (L2) proficiency level assigned by the language programme. The L1 background is defined as a variable inherent in individuals that potentially influences the presence of a relationship between the linguistic variation in a specific writer's writing and time. However, it is distinguished from individual differences that are not defined in terms of L1 backgrounds. In addition, its effect is considered a random effect; it is not assumed that its influence on the relationship between language use and time is systematic.

The L2 level variable is considered a potential factor that can interact with the time variable. The main interest of this research is to explore time effects controlling other potential factors on linguistic variables, and therefore, the L2 level is included as a mediating variable in the inferential regression analyses to be discussed in Chapters 4, 5 and 6.

As for the reference data, several sources are tried: the university student writing corpus (MICUSP), the level 5 writing, and the hypotheses drawing on the previous literature. The decisions on choosing the relevant reference data are justified in each method section of the analyses to be reported in the following chapters.

3.3.5 Overview of Methodology 1: Multidimensional Analysis

The primary method used for the research objectives of this thesis is the multidimensional (MD) analysis method, pioneered by Biber (1988) to explore register variations between 23 spoken and written genres. This method focuses on the functional aspect of language, with an assumption that core linguistic variables (e.g., pronouns and verbs) are associated with texts' communicative purposes and situational context (Biber & Conrad, 2009, p.2). This section discusses the common procedures of MD analysis and the underlying rationales of MD analysis.

The overall steps of MDA illustrated in Biber (1995) are adapted below:

Step 1: Texts are collected, or corpus is selected considering metadata regarding the situational contexts and learner linguistic backgrounds that are suitable for the research inquiry

Step 2: Grammatical research is conducted to identify the range of linguistic features for analysis, together with functional associations of individual features.

Step 3: Texts are prepared in .txt format and input into computer programs for automated grammatical analysis to 'POS-tag' all relevant linguistic features in texts.

Step 4: The entire corpus of texts is tagged automatically by computer, and all texts are post-edited interactively to ensure accurate identification of the linguistic features.

Step5: Syntactic tagging is conducted with the POS tagged texts to compute frequency counts of each linguistic feature in each text.

Step6: The co-occurrence patterns among linguistic features are analysed using factor analysis of the frequency counts.

Step 7: The 'factors' from the factor analysis are interpreted functionally as underlying dimensions of variation.

Step 8: Dimension scores for each text concerning each dimension are computed; the mean dimension scores for each register are then compared to analyse the salient linguistic similarities and differences.

As shown in Step 2 above, Biber researched linguistic variables of potential significance in accomplishing communicative functions in texts. He then targeted 67 linguistic forms for analysis. The forms occurring together with statistically significant probability are interpreted in terms of their shared factor, a dimension. Dimensions carry functional meanings, which provides a framework in which the linguistic variables are more meaningfully analysed (Biber, 1988). For example, Biber (1988) found that academic prose had very low frequencies of past tense verbs and third-person pronouns. This linguistic distribution is associated with a dimension of narrative vs non-narrative, and the low score on this dimension indicates that academic prose is largely non-narrative. For example, Biber (1988) reduced 67 linguistic features into seven dimensions using factor analysis.

The assumption is that situational contexts within which a text is produced influence the linguistic patterns observed. Therefore, the situational factors need to be carefully considered in data collection. MDA acknowledges the complexity of linguistic phenomena as dependent on situational contexts. The situational variables and learner variables (e.g., gender, age, target language proficiency) heavily influence the linguistic phenomenon in texts. Therefore, the collection of linguistic data should represent the range of communicative situations and functions available in a language for which a research question is formulated (Biber, 1988, p.65).

In step 6, a factor analysis, a statistical measure, is employed to reduce the co-occurring patterns of the linguistic features into a manageable number of factors within similar types of texts (Biber, 1988, p.79). This integration process assumes that frequently co-occurring linguistic features share at least one communicative function (Biber, 1988). This study uses 74 grammatical and semantic features. The grammatical features are further specified by their syntactic functions at both clausal and phrasal levels. In addition, the seven verb semantic categories were added.

The assumption is that certain forms co-occur in a type or genre of texts to realise the purpose inherent in the text. In other words, the co-occurring linguistic features in a type of text share a communicative function, which can be interpreted as factors or dimensions (Biber, 1988). MDA is based on the belief that if it is possible to measure linguistic performance reliably, a more abstract notion of communicative functions underlying a piece of writing can be mapped onto the linguistic forms. Therefore, it is possible to interpret what linguistic patterns are used to accomplish the communicative functions, given that other influential variables are reasonably factored out. One of the critical assumptions of MDA is that linguistic co-occurrence patterns are functional. Linguistic features occur together in texts because they serve related communicative functions (Biber, 1988). MDA involves extracting statistically co-occurring linguistic features from corpora. This procedure assumes that statistical differences in co-occurrences of lexico-grammatical features can predict different types of texts' unique characteristics. MDA presupposes form-function mapping to find out more general, higher-level functional characteristics in similar types of texts using more micro-level linguistic patterns.

The Multidimensional analysis (MDA) method (Biber, 1988), the main method for the present study has been chosen for several reasons. First, the importance of form-function mapping has been found in earlier studies (e.g., Ervin-Tripp, 1972; Hymes, 2003; Brown & Fraser, 1979). Patterns of co-occurring linguistic features can be more effective indicators of textual characteristics than any linguistic feature removed from contexts (Hymes, 2003; Brown & Fraser, 1979). Understanding linguistic patterns in broader and multi-perspective ways can yield a more comprehensive understanding of their functions in texts. The multivariate nature of the MD analysis method (employing factor analysis) can effectively capture relatively short-term variations in texts (Gray et al., 2019). In other words, MD analysis focuses on the linguistic co-occurrence of multiple features rather than any individual linguistic feature, identifying underlying incremental linguistic changes within discourse structure.

The identification of linguistic dimensions of L2 academic writing offers unique information about the linguistic choices of L2 writers that have not yet been extensively surveyed in CL. These dimensions potentially distinguish the nature of L2 student writing from other kinds of academic texts (Friginal & Weigle, 2014, p.82). In light of this, factor analyses were conducted to find new dimensions. This MDA involves conducting a new analysis to identify linguistic and functional co-occurrence patterns of the discourse domain being explored, which is what Sardina et al. (2019) refer to as the second type of MDA. Registers can then be compared concerning those new dimensions. This approach is appropriate when analysing a new discourse domain that includes many different text categories (Biber et al. 2004: 52–53). Thus, to identify the underlying dimensions of variation that can distinguish among different language backgrounds and proficiency levels, a new MDA is necessary for the present study. (Pan, 2018, p.120).

The eight established analysis steps of MDA introduced earlier can be simplified to include the following three steps: obtaining frequency data of the linguistic features of interest in the corpus, factor analysis, and computation of dimension score based on the standardised frequencies of the major linguistic features loaded in each factor (dimension). MDA uses factor analysis to identify functional dimensions from a set of syntactic features associated with the dimensions. The syntactic features used in factor analysis and computation of dimension scores for the factors are introduced in the following sections.

Detailed illustration of procedures and decisions made at each step is discussed in the following sub-sections on exploratory factor analysis, dimension score computation, and a post hoc analysis for each round of MDAs. The data processing, which is the initial step of preparing the data, will not be discussed here, as each MDA involves different linguistic variables. Instead, the selection and preparation of linguistic variables and procedures for tagging the data to obtain the frequency measures for the linguistic variables are detailed in each section of the MDAs separately.

3.3.5.1 Exploratory Factor Analysis (EFA)

This step is to identify the functional dimensions of linguistic features that co-occur in similar developmental stages of texts. A statistical measure such as factor analysis identifies linguistic features that meaningfully co-occur with others using the frequency counts to reduce the number of original linguistic variables to a smaller group of derived variables, the factors (Biber, 1988, p.79). The ‘factors’ from the factor analysis are interpreted functionally as underlying dimensions of variation. A factor analysis used in MD analysis is exploratory rather than confirmatory in that the significant linguistic features are discovered without being predetermined beforehand. Exploratory factor analysis (EFA) can be beneficial in that the derivation of dimensions for L2 academic discourse ‘offers unique information about the linguistic choices of L2 writers that have not yet been extensively surveyed in corpus linguistics’ (Crosthwaite, 2016, p.5; Friginal & Weigle, 2014, p.82). The section details this statistical procedure.

Exploratory Factor Analysis (EFA) and Principal Component Analysis (PCA) are both data analysis techniques, but serve distinct purposes. EFA excels at identifying latent variables, underlying factors that explain relationships between observed variables. This strength stems from its ability to partition the data's variance into two key components: common variance (explained by latent factors) and unique variance (specific to each variable). This distinction, as Fabrigar et al. (1999) point out, allows researchers to isolate core factors influencing the observed data, leading to a more interpretable representation. In contrast, PCA focuses solely on the covariance matrix, excluding error variance, while PCA considers total variance (Tabachnick & Fidell, 2007). This difference might make PCA better suited for data with many zeros, as noted by Gray et al. (2020, p.149). For the complementary strengths of these two methods, both PCA and EFA were tried in pilot studies. However, EFA was chosen for the final MDA due to its superior interpretability of the factors.

The general procedures are as follows. First, the frequency counts of each linguistic feature from the POS-tagged corpora are normalised per 100 words and prepared in a CSV file format, which can be read in the R environment. Factor analysis involves several decision-making steps to avoid over-/under factoring as possible. The initial factor analysis results are the basis of the decision to examine the adequacy of the linguistic variables to be used in a factor analysis, based on several statistics. The correlations between the features' normalised frequencies were assessed to ensure they do not have too high correlations. The factorability of the sample was assessed using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, and the communality of each feature was also checked. features with low factorability (under 0.5) or communality (minimum 0.18) were removed from the final analysis.

In addition, different solutions for factor analysis are compared to decide on the best number of factors in a solution, using the information regarding scree plot inspection and the interpretability of the factors extracted in different solutions. A plot of the eigenvalues provides a way to measure the amount of variance accounted for by each factor. The eigenvalues can indicate the percentage of shared variance that is accounted for by each factor. The factor solution should account for a significant percentage of the cumulative shared variance. A final step in a factor analysis before interpretation is a rotation of the factors, a solution in which each feature loads on a minimum number of factors (Gorsuch, 1983). The factor loading of a feature is the variance explained by the feature for the dimension. The initially extracted results become simpler in this rotated solution. Therefore, this simplified structure greatly facilitates the interpretation of the constructs underlying each factor. Among several alternatives, a Promax rotation has been used in previous MDA studies since it permits minor correlations among the factors (e.g., Biber, 1988; Biber & Gray, 2013).

3.3.5.2 Interpretation of Functional Dimensions

Once an exploratory factor analysis is conducted in the previous step, the results provide positive and negative features loaded on each factor, with which functional dimensions are interpreted (e.g., dimension with positively loaded features indicating informational density vs. negatively loaded features associated with personal narrative). The next step is to compare textual dimensions among text groups at different time points. A complete interpretation of a textual dimension is made possible by considering:

(1) the mean of the factor score of the texts grouped by each temporal point
(2) the linguistic features constituting the factor score
(3) the situational parameters associated with the distribution of factor scores

We will discuss each of these elements in turn. First, a factor score is the covariance score of each linguistic feature and a factor, calculated using ‘R^2’ (For simple linear regression models with only one predictor, R2 corresponds to the squared correlation coefficient; Winter, 2019, p.90) in the exploratory factor analysis stage. In a more straightforward sense, factor score is the normalised frequency of the sum of a feature per text. In terms of operationalisation, a factor score is the sum of the features of each factor (dimension) in each text. More specifically, dimension scores are computed for each factor by (1) standardizing the counts for each linguistic feature using the z-score formula and (2) subtracting the sum of the standardized counts of the negative-loading features from the sum of the standardized counts of the positive-loading features (Egbert & Staples, 2019, p.140). Following (Biber, 1988), factor (dimension) scores for each text concerning the five PELIC sub-corpora were computed to interpret textual dimensions through the following steps.

First, the normalised frequencies used in the factor analysis were standardised. The Z score is computed for the finally retained number of linguistic features to compute the factor score. The loading of a feature on a factor reflects the extent to which the variation in the frequency of that feature correlates with the overall variation of the factor (Biber, 1988, p.85). It indicates the strength of the co-occurrence relationship between the feature in question and the factor as a whole. Only the linguistic features with the 'salient' loadings on each factor were finally retained features for interpretation.

The standardised values reflect the magnitude of a frequency to the range of possible variation. Therefore it is suitable to compare the relative weight of the feature in the text rather than its absolute value (Biber, 1988, pp.93-97). The standardized linguistic counts were not weighted according to their factor loading weight in these calculations. This process translates the scores for all features to scales representing standard deviation units. That is, standardised scores measure whether a feature is common or rare in a text relative to the overall average occurrence of that feature. Factor scores for each factor were computed, summing the features' frequencies with loadings greater than +/-0.35 on each dimension. Furthermore, each feature is assigned to only one dimension even if they are significantly loaded on more than one dimension. Once a factor (dimension) score is computed for each text, the mean dimension score for each sub-corpus can be computed.

The second element to be considered in the interpretation of textual dimensions is each linguistic feature with high factor loading to a factor. The complementary relationship between positive and negative loadings can also provide information for interpretation. Finally, the third element, the situational and learner variables considered in the present study includes a personal course of study, L2(English) proficiency, revision(feedback), topics, and L1 background. This step involves qualitative analysis when necessary, referring to a linguistic feature in the context to examine the communicative functions that the feature performed. Considering these elements, similarities and differences among the textual ‘relations’ can be analysed for these scores to support or refute hypothesised interpretations. Information about the most representative dimensions of texts at different temporal points and the linguistic features highly indicative of each dimension can provide the characteristics of language use along developmental stages.

3.3.6 Methodological Overview 2: Addressing Mediating Effects on Syntactic Variation

A subsequent statistical procedure, such as a mixed effect model (MEM), was conducted to test significant mean differences among Dimension scores of the sub-corpora.

The degree to which each linguistic variable and associated factors or dimensions can explain the difference among the groups of texts still needs to be checked concerning the other mediating variables such as writing topic, L1 background and L2 proficiency. As noted in elsewhere , interpretation of textual dimensions needs these situational and learner-specific variables to be considered. To that end, both statistical tests and qualitative examination of the sub-corpora can be helpful.

An inferential statistical analysis, such as linear MEMs, can help increase the precision of the regression results by factoring out extraneous effects that are not the research focus. LMM specifies fixed factors and random effects, which helps to see the more realistic effect size of the linguistic variables of interest (Winter, 2019, p.234). As discussed in Section 3.3, academic semesters and English proficiency are set as fixed effects while learner’s L1 backgrounds and topics of assignment tasks are set as random effects. The linguistic pattern across academic terms will be the dependent variable. The two steps for a recommended LMM procedure are adapted from Gries (2015b) and briefly illustrated here. First, the fixed-effect model will be selected and tested with random slopes and coefficients to find the optimal mixed-effect structure. Second, the random structure will feed into the old fixed-effect model to adjust, resulting in the optimal fixed-effects structure, which includes the only mediating effects that meaningfully contribute to the linguistic variations across text groups by temporal points.

After these statistical steps, critical linguistic features representative of the significant dimensions were identified, some of which are discussed further in Chapter 6. These subsequent analyses intend to find any possible explanations for the patterns found in the MD analysis and make meaningful interpretations of individual features rather than collective groups of co-occurring linguistic features.

In terms of data analysis for generalization, there is a shift in focus towards probability estimation rather than binary testing. The importance of uncertainty estimation with this dataset supports the use of inferential statistics that can incorporate previously known parameters into the models and handle more flexible data distribution to approximate the probability of the data being generalised to a greater population. In light of this, the MD analysis results were further analysed using mixed-effects models drawing on Bayesian theory for interpreting the significance of their results. As noted by much previous research, the linguistic patterns found in learner-written texts show variations, depending on the learner's ability to make linguistic choices appropriate to the given tasks and the communicative purposes of the given task itself. Previous findings support the interaction among situational factors such as genre types, disciplines and learner variables such as the target language proficiency in explaining linguistic functions associated with the co-occurring linguistic patterns in texts (e.g., Hardy & Friginal, 2016; Gardner, Nesi & Biber, 2019). Therefore, these factors cause the variations of the linguistic feature to be multifaceted (Hardy & Römer, 2013, p.185; Hardy & Friginal, 2016, p.121). Another consideration of importance in studying linguistic development in writing involves the writer-specific variables. Therefore, considering these factors together, the interpretation is based on mixed-effects models, a regression analysis controlling random effects on the linguistic variables.

3.3.7 Methodological Overview 3: Collostructional Analysis

The analysis aims to compare the ways and extents to which different stance functions are employed over time: overt indications of either epistemic certainty/doubt or other attitudes over time. The linguistic focus is on the VAC constructions used to express stance. *That* complement clauses are arguably the most important grammatical stance device in English because the full set of words that can control a that-clause all express stance meanings (Biber, in Biber et al., 2021, p.124). As discussed earlier, constructions for epistemic stance function have been noted as a marker of academic writing while attitudinal stance function is more frequently used in spoken registers. Therefore, it was expected that epistemic stance would be more frequently used in later Courses. To examine the preferred stance functions across Courses, I listed the sets of verbs that occur before that-clauses according to two broad categories of stance functions: epistemic and attitudinal stances. Then, the collostruction analysis method (Gries et al., 2005), an association strength measure, was applied to examine the attraction between the verb and the constructions for each of these two stance functions.

3.3.7.1 Linguistic Variables

The focus feature is verb-argument constructions, defined as a main verb and *that*-complement clauses, including *that*-deleted constructions (as in *I think this is good*). I consulted the following framework of stance form and function: verb semantic categories indicating stance function (Biber et al., 2021). The frequency of verbs in the constructions has a predicting power of acquisition as the frequency input plays an important role in the production of generalized constructions (Tomasello, 2003). Biber et al. (2021b) categorise verbs into seven semantic categories and report that most mental verbs and some communicative verbs followed by that-complement clauses are markers of epistemic or, less often, attitudinal stance. Table 3-9 shows a framework of stance markers identified and analysed in the present study. See Appendix 6-2 for a full list of verbs in that-complement construction identified in the longitudinal PELIC texts.

Table 3-9. Stance constructions categorised by verb semantic types and *that* dependent types

Target form	controlling verb + that-complements (THVC)

	*including '*that-deleted constructions* (THVC) controlling verb + to-infinitives (TO) *including split constructions between 'to' and 'non-finite clause'
Semantic Verb Category	*That*-complement clauses are controlled by attitudinal, likelihood, factive or non-factive verbs (adopted from Biber, 2004, pp.133-135)
Example	I *think that it is right*. / I *think that is right*.
Processing	MAT tagger is used to tag the texts to THVC and THATD -> MAT-tagged texts are lemmatised -> concordance lines are extracted using AntConc -> manually sorted by semantic category

3.3.7.2 Collostruction Analysis

Collexeme analysis (Stefanowitsch & Gries, 2003) measures the joint probability (i.e., it is not directional) that two items in a corpus will co-occur. When collexeme analysis is used to measure the strength of verb-construction combinations, it is termed collostructional analysis (Gries et al., 2005). Collostructional analysis calculates the association between a word and a given construction, the so-called collostruction strength using the Fisher-Yates exact test (Fisher, 1934; Yates, 1934). While Gries (2023) suggests using residuals from a chi-squared test as a computationally efficient alternative to established methods like the Fisher-Yates exact test or G2, he also emphasizes the need for a more sophisticated approach when the goal is to move beyond description and explore theoretical explanations.

The Fisher exact test is known for its ability to address very rare collostructions, which could enhance the reliability of the results (Stefanowitsch & Gries, 2003, pp. 217-218). However, its high computational cost can cause very slow operation when dealing with high-frequency items (Stefanowitsch & Gries, 2003, p.218). Kyle (2016) suggests an alternative way of calculating collostructional strength, which is noted to almost perfectly correlate with collostructional strength (Gries, 2015a, cited in Kyle, 2016). This is achieved by multiplying the delta P value (with construction as the cue and verb as the outcome) by the frequency of the verb. It should be noted that the approximate collostructinal analysis, using delta P, is a uni-directional measure. It has been argued that uni-directional measures have an advantage over two-directional (or mutual) measures such as the Fisher-Yates exact test (Gries & Durrant, 2020, p.144). Gries and Durrant (2020) note that uni-directional measures can overcome the shortcomings of the measures of mutual attraction, which cannot distinguish between different kinds of attracted elements. The present study adopts this measure of approximate collostruction analysis (Kyle, 2016; Kyle et al., 2021) to operationalise the construct of syntactic sophistication highly associated with VAC in a reference corpus.

The common procedures of the collostructional analysis method adopted in the two analyses in this chapter are as follows. First, a 2x2 contingency table is used to calculate the association strength. Collostruction analysis computes the *expected* frequencies *a*, *b*, *c*, and *d* that would result from *x* and *y* co-occurring together as often as would be expected from their marginal totals (*a*+*b* and *a*+*c*) as well as the corpus size *N (a+b+c+d)* (see Table 3-10).

Table 3-10. Contingency table for collostructinal analysis of word W1 and construction C (Gries et al., 2005, p.644)

	construction C	¬ construction C	Row totals
word W1	<u>a</u>	b	<u>a + b</u> (=overall lemma freq of W1)
¬ word W1	c	d	c + d
Column totals	<u>a + c</u>	b + d	<u>(a + b) + (c + d) = N</u> (total number of argument-structure construction in the corpus)

For computing the collostructional strength, the four underlined measures ((1) *a,* (2) *a + b,* (3) *a + c*, and (4) *(a + b) + (c + d) = N)* of V-that clauses and V-that deleted clauses, the following steps are taken. First, the primary frequency data came from two of the lexico-grammatical features used in the multidimensional analysis discussed in Chapter 4, two *verb-argument constructions: that-clauses controlled by verbs and to-infinitives controlled by verbs*, identified by a Multidimensional analysis tagger (MAT; Nini, 2019).

Next, the tagged texts including these two tagged features are lemmatised to extract the four underlined frequency measures ((1) *a,* (2) *a* + *c,* (3) *a* + *c*, and (4) *(a + b)* + *(c + d)* = *N)* of V-that clauses and V-that deleted clauses, respectively, using the spaCy module (reference) in the Python environment (version/platform here). Then, the remaining measures (*b, c, d* in Table 2) were obtained by means of subtraction from these four frequency measures.

The third step involves using the frequency measures from the above steps to compute collostruction analysis scores for each verb lemma in the main corpus and reference corpus, respectively, on a spreadsheet. The approximate collostructional scores are computed, following the formula in Kyle (2016) = $(\left(\frac{a}{a+b}\right) - \left(\frac{c}{c+d}\right)) * (a+b)$,, which approximates collostructional strength by multiplying the delta P value (construction as cue, verb as outcome) by the verb frequency (Kyle, 2016, p.67).

An association measure for 2 * 2 tables, which use the one-tailed p-value computed by the Fisher-Yates Exact test (cf. Fisher, 1934, Yates, 1934), is computed on the basis of the hypergeometric distribution (Gries et al., 2005, p.647):

P observed distribution = $\frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{N}{a+b}} + \sum P\ all\ more\ extreme\ distributions$

The two studies discussed in Section 5.4 and 5.5 use the association strength variation of VACs associated with stance function in L2 writing: the pilot study (discussed in Section 5.4) explores *Verb-That clauses* and *Verb-To infinitives while the final study (in Section 5.5) focuses on Verb-That clauses* only. The frequency of the target constructions is considered additional information only, and they were not considered as a variable itself in the subsequent analysis to explore the aspect of sophistication and regression analyses.

3.3.7.3 Measuring Grammatical Sophistication

The overarching construct of syntactic sophistication is operationalised drawing on the high-frequency controlling verbs in the VACs in reference to the reported frequency of verbs in general registers by Biber et al. (2021). The frequency distribution of common verbs in seven semantic domains was used to classify verbs: activity, mental, communication, existence, occurrence, causative and aspectual, by order of proportional importance in the corpus, referring to Biber et al. (2021, p.363). The semantic categories of verbs are useful to understand the communicative functions in the that-complement constructions.

This study pre-defined sophisticated constructions to examine based on three criteria. I used three types of reference data: MICUSP, PELIC Level 5, and the verbs listed in Biber et al. (2021, p.369) as a general reference point when analysing both MICUSP and PELIC, two academic student registers.

In addition, the two reference data are used to represent the target registers: the corpus consisting of Level 5 texts from the PELIC corpus (for the final study in Section 5.5) and the MICUSP corpus (for the pilot study in Section 5.4). Level 5 is part of the data collected in the same situational context - English writing classes. However, they include texts written by writers with different entry levels. In other words, the corpus of Level 5 is representative of writing where the writers are closest to the target language use assessed by the programme. In a sense, MICUSP is more approximation of the target language used for university disciplinary purposes (assignments). However, PELIC texts consisting of Level 5 are more controlled in a way that they share the situational context with the longitudinal PELIC data, which reduces the variations that cannot be explained by the specified variables in analyses. In sum, these two reference data are representative of target languages chosen by different criteria, bringing different types of reference points to compare the longitudinal data with. The PELIC Level 5 texts will provide the criteria to assess whether the longitudinal variation from an analysis is headed toward the direction of what is valued within the programme by the teachers, while the MICUSP corpus will be the empirical reference point against which the previous findings in other studies on development can be compared.

The stance constructions with the high association strength scores in the reference data are expected to be more saliently observable in later stages of the writing process as evidence of development. The results can provide evidence to either prove or reject the assumption that writers expand their stance marker repertoire and productivity over time, and stance word-dependent associations move closer to the target-like usage.

This stage of analysis involved selecting sophisticated constructions by 1) semantic categories of controlling words (following Biber et al., 2021, p366), 2) type of that-dependents (that-clauses and that-deleted clauses) and 3) functions of the constructions. Finally, the data is sorted by five Courses of study and compared with reference data (Level 5 corpus and the verb frequency reports by Biber et al., 2021b).

The present study uses the collostruction analysis data to assess the communicative functions of controlling verbs in the target construction. First, verbs' semantic categories and their usefulness for linguistic analysis were assessed based on their frequencies in VACs, semantic categories, and associated communicative functions examined in the concordance lines. Communicative and mental verbs were expected to occur more frequently in the argument structure than in other structures. Second, some verbs with high collostruction scores were considered for further qualitative inspection. The potential formulaic constructions in the L2 data were compared to their expected general frequencies in academic registers, consulting Biber et al. (2021b). I did it to check the first hypothesis (refer to Section 6.2.3).

The analysis focused on ranking the scores for each criterion instead of using the raw scores themselves. This is done to account for potential variations in corpus size and sub-corpus sample sizes across the different criteria. As previously mentioned in Gries (2014), collostruction analysis based on rankings can effectively account for variations in verb frequencies and corpus sizes associated with different constructions. This allows us to focus on the relative ordering of scores across constructions, rather than the absolute values, ensuring a more systematic comparison even with potential discrepancies in sample sizes.

3.3.8 Ethical Considerations and Practices in data collection and analysis

While general research ethics apply to corpus linguistics as in any other academic discipline, there are specific ethical issues related to handling extensive linguistic data. One such issue pertains to the responsibility of corpus distributors in ensuring that data contributors are informed about and approve of the use and distribution of their data.

In the case of the PELIC corpus, which is a secondary documentary dataset containing participants' information, the sensitivity of the data and its intended uses in the research necessitate careful ethical considerations. If the data is publicly stated as available for academic research purposes, it can be assumed that data contributors have granted their consent for its utilization in this proposed study (British Educational Research Association [BERA], 2018, pp.11-12). The ownership of the PELIC datasets lies with Juffs, Han, and Naismith, and they have designated the dataset repository under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://github.com/ELI-Data-Mining-Group/PELIC-dataset#11-License).

Respecting the privacy of informants is a general responsibility for researchers (The British Association for Applied Linguistics; BAAL, 2021, p.4). Accordingly, researchers are obligated to obtain informed consent from individuals providing data for the research prior to data collection (McEnery & Hardie, 2011, p.61). Furthermore, researchers must ensure the confidentiality and anonymity of data during the data preparation process. Personal data pertaining to informants should be coded in a way that conceals their identities while preserving the data useful for research analysis and interpretation. The coding system should be systematic to ensure that data users understand and utilize the metadata. Corpus compilers provide a detailed account of the procedures and coding system developed to anonymize the data for the purpose of systematic retrieval. In the case of the PELIC corpus, the anonymization process involved coding systems for corpus data (including places, personal names, and multiple different personal names) and metadata (student information and linguistic proficiency), as shown in Appendix 3-5.

The study employs a multidimensional analysis to address the first and second research questions. This analysis involves identifying linguistic features and conducting factor analysis to explore emerging functional dimensions in L2 writing data. Additionally, a collostruction analysis is conducted to analyse the linguistic features associated with stance function (Chapter 5). The third and fourth research questions are addressed through inferential analyses, measuring the longitudinal change of key syntactic constructions while considering mediating effects such as L1 backgrounds, L2 proficiency, and writing topics. The attempt to generalize the analysis results to a broader population of L2 longitudinal writing and its contribution to previous research in L2 writing development is discussed. The results of a subsequent mixed-effects model are presented in Chapter 6. Detailed information on the data processing and analysis procedures for each research question can be found in the respective chapters.

Regarding ethical considerations, the replicability of results is emphasized as a crucial concern for corpus data users. To ensure replicability, researchers are advised to make all data preparation and analysis procedures transparent. Transparency is vital in scientific practice as it allows for potential improvements or corrections to be made. This applies to analyses relying on algorithms or automatic statistical analysis embedded in computer programs. It is considered unsafe to report the output from a program without providing the underlying mechanism of the automatic procedures. Transparency should be observed in all analyses, including manual analyses of corpus data. Researchers are encouraged to maintain detailed records of the operating procedures that produced the results and reference the software version used, considering that corpus linguistics software is continuously updated and improved. These ethical codes of conduct should be rigorously followed throughout the research process and reporting stage.

3.4 Tagging Accuracy Report

3.4.1 Introduction

The purpose of this section is to present and report the results of assessing the accuracy of grammatical tagging and syntactic parsing of the data. The research instrument for primary data collection for the present study is discussed, focusing on the pre-processing of the data (data cleaning/ accuracy check for tagging linguistic features) for automated analysis. Established data preparation schemes which consider learner corpora tagging accuracy can improve the accuracy of the quantitative results. Many researchers recommend checking taggers' output as thoroughly as possible before using it, especially when this is more of a quantitative analysis, which can enable us to make more informed decisions about using the tagged features for analysis (e.g., Durrant, 2023; Gray, 2019). In particular, learner corpora are likely to contain misspellings and ungrammatical structures, which can influence tagging accuracy.

The first aim of this section is to discuss the process and result of the manual checking of the whole corpora before being automatically tagged by the software MAT for the pilot study and to compare it with the automatically processed data; the check and comparison aimed to examine the adequacy of the tagged features, and if necessary, to reprocess the tagged features or text data to obtain a reasonable level of accuracy of the tagged features for the Pilot analysis. The second aim is to discuss the result of checking tagging precision and recall rate of two features: *that*-clauses controlled by nouns and adjectives. There are three types of checks reported in this draft: tagging precision, recall, and overall accuracy. Each of them will be further illustrated in subsections 3.1 and 3.2 in turn. The following subsection discusses issues regarding the reliability of data analysis when using learner corpora. The operational definitions of three reliability measures for tagging accuracy and example studies of checking and reporting these measures are also discussed.

3.4.2 Literature review on tagging accuracy

Corpora are useful data for linguistic analysis, but they also require careful examination before using it for analysis. This literature review discusses definitions and implications of reliability issues in analysing corpus data. It first addresses theoretical and practical issues regarding *reliability* in using and analysing corpus data, and then moves on to discuss ways of measuring reliability and their implications on the results of corpus analysis.

Reliability, referring to the extent to which a test or any measuring procedure is consistent on repeated trials, is important in generalising the research findings in quantitative research (Creswell & Creswell, 2017). When using automatised research tools, however, the results of the observations will be affected by the definitions provided by the tool as well as the designs of the tool (Anthony, 2013, pp.143-144). This observation leads to the importance of understanding analysis tools. While many automated tools for corpus annotation provide valuable linguistic data for a large number of corpora, the researcher needs to ensure that the theoretical definitions of the linguistic variable of interest are compatible with the data provided by automatic parsers and taggers.

Syntactic parsers such as the Stanford Parser (Chen & Manning, 2014) and spaCy (Explosion AI, 2018) automatically annotate texts for syntactic constituency or dependency relationships (Kyle, 2021). Texts are first tagged for part of speech, then the tags are used, in conjunction with the phrase structure rules, to generate several competing sentence-level parse trees (Kyle, 2021, p.7). TAASSC relies on a Stanford dependency parser (De Marneffe et al., 2014) to identify syntactic structures, with a reported tagging accuracy of around 90% (Chen & Manning, 2014). On the other hand, MAT (Nini, 2019), which relies on a Stanford tagger (Toutanova et al., 2003), makes modifications to part of the tagging algorithm to replicate that of Biber (1988). As these two taggers, both used in at least one of the analyses reported in this thesis, rely on different operational definitions of linguistic constructions, a separate tagging accuracy assessment is required.

While high levels of annotation accuracy have been reported for some L2 English texts (Kyle, 2021, p.6), the tagging accuracy can be more likely to be problematic for learner corpora tagging (Meurers & Dickinson, 2017). Most automatic annotation tools are trained on well-edited L1 corpora that may be quite different in the language used in learner corpora (Kyle, 2021, p.2). As with tokenization and lemmatization, typos and spelling errors, language-specific issues may cause POS annotation errors. Berzak et al. (2016), for example, found that POS annotation accuracy for tokens used ungrammatically was 88.61%, compared with 95.37% for tokens used grammatically. Some study that used SpaCy (Explosion AI, 2018), a Python module to process and extract dependency structures from texts, reports high tagging accuracy of some structures (e.g, verb-direct object, 96.0%) in L2 learner corpora (Kyle, Eguchi & Granger, 2021). However, lower accuracies were reported for more complex linguistic features. Kyle et al. (2021), for example, reported 80% annotation accuracy for verb argument constructions (which were defined as a main verb and all of its direct, non-auxiliary dependents) in lower proficiency L2 English essays using spaCy.

The increased importance of checking the adequacy of tagging accuracy of different taggers leads to checking the following features' tagging accuracy. As different POS taggers may reflect a different grammatical theory in analysing the data, using data tagged by different taggers may require separate checks for tagging accuracy. Therefore, I checked the tagging accuracy for 34 features parsed by the Stanford parser via TAASSC (for the analysis in Section 4.5.), two MAT-tagged features and two TAASSC-tagged features (for the analysis in Sections 4.6, 6.4 and 6.5). The following section discusses the procedures, results and implications of these checks.

3.4.3 Analysis & Results of Tagging Accuracy Checks

The purpose of evaluating POS tag accuracy in this report is to diagnose the adequacy of features based on accuracy measures such as *precision, recall* or other measures for using these features for quantitative analysis. As Gray (2019) noted, a full evaluation of POS tagging accuracy may be desirable when a new register is being tagged, or when updates are made to the tagger itself.

3.4.3.1 Accuracy check for TAASSC-parsed features

To evaluate the accuracy of dependency features parsed by the Stanford parser via TAASSC, I compared them to manual annotations in 20 randomly selected PELIC texts (used for pilot analyses in Chapter 4, Section 4.5). This analysis focused on feature frequency (e.g., number of adjectives). I assessed clausal/phrasal dependencies and verb semantic categories (parsed by spaCy) as follows.

Step 1. Data Extraction: TAASSC-extracted data for the chosen texts were copied into a spreadsheet.

Step 2. Manual Annotation: Manual counts for each feature were added to the same spreadsheet.

Step 3. Accuracy Calculation: For each feature in each text, the TAASSC value was divided by the corresponding manual count and expressed as a percentage. This revealed over/undercounting by TAASSC.

Step 4. Average Difference: The mean percentage difference was calculated for each feature across all texts.

Step 5. Additional Analysis: An extra 20 texts were used to assess features with a matching rate below 80%.

The accuracy rates for these features ranged from 75% to 100 % (Refer to Appendix 3-4). The overall accuracy rate of the dependency features passed 90%. The recheck with additional smaples resulted in a significant increase in accuracy, with most features exceeding 90% accuracy. However, exceptions included ‘rcmod’ (relative clause modifier), ‘preposition of’, and ‘parataxis’ (juxtaposed clauses). These features exhibited lower accuracy due to their general infrequency in the learner corpus. Learner language itself also contributed to tagging errors due to the use of unidentifiable grammatical structures (discussed in Section 3.5). The full report of the accuracy check for this set of features is available in Appendix 3-4.

3.4.3.2 Precision and recall check for MAT-tagged features

The 73 MAT-tagged frequency features were not assessed for accuracy in this study. MAT's tagging scheme is modified to replicate Biber's features (Biber, 1988). MAT further modifies the Stanford tagger's scheme for Biber's features. The Stanford tagger, used by MAT, relies on grammatical forms, not syntactic functions.

While the accuracy check in Section 3.4.3.1 provided information about the accuracy of most of the features tagged by Stanford tagger, some concern remained regarding structurally complex features, particularly those involving ‘*that*-complement’ constructions (including the manually tagged ones). Biber's analysis (1988) did not differentiate between these features based on the relatives and complements within noun-controlled 'that' clauses.

During the feature reduction process of factor analysis (reported in Chapter 4), some structurally complex and theoretically significant features were lost, including ‘that-complement clause’ constructions (discussed in Chapter 4, Section 4.5). These constructions are crucial for L2 research, so a separate accuracy check was conducted to determine their suitability for the measures reported in Chapter 4 (Section 4.5) and Chapter 5 (Section 5.4).

One potential source of tagging inaccuracy is the MAT tagger's limited operational definitions, which did not incorporate the syntactic function dimension, and only considered definitions of grammatical form. In that regard, the definitions of MAT tags are not able to identify words with complex forms with different syntactic functions. For example, it does not identify that-complements on objective position, as in 'I told you that S V' while it does tag that-object complements that directly follow verbs, as in 'I said that S V.' Moreover, the extraposed *that*-clauses are not distinguished from other that complements in object positions. The issue was that the extraposed *that*-clause following nouns was tagged as object relative complement, which should be identified as a subject complement.

Therefore, some more structurally complex constructions among the MAT-tagged features, including the manually tagged 'noun-that complement clauses', were separately checked for their tagging accuracy.

This report focuses on the accuracy of two manually tagged constructions: *that*-noun-complement clauses and *that*-adjective-complement clauses. The verb-that-complement clause accuracy is not reported here because spaCy (AI Explosion, 2018) was used for tagging, and its accuracy has been reported in previous studies (Kyle & Eguchi, 2021; Nini, 2019; refer to Section 3.4.2) and also the verb tagging accruacy has been conducted (refer to Section 3.4.3.1).

In this section, I report the result of tagging accuracy analysis, following the definitions and procedures of this in Durrant (2023, pp.116-117). In technical terms, instances where a parser accurately recognises a feature are known as true positives. Alongside these correctly identified items, there are two main types of errors that a parser might make. The first type of mistake, false positives, is to incorrectly tag words as one type of grammatical category even though they belong to that category. The measure of false positives is known as precision. This tells us what percentage of tags in the annotated texts are correct. The second type of mistake, false negatives, is for the parser not to tag a word as a grammatical category when it does belong to that grammatical category. The measure of false negatives is known as recall. This refers to the percentage of actual cases of a feature in the corpus that the annotator has found. Using these definitions, precision and recall are calculated as:

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}$$ (Durrant, 2023, p.117)

This section describes the assessment of two stance features: "That noun complements" (THVC) and "That adjective complements" (THAC). The goal was to determine their suitability for statistical analysis in PELIC texts, which might contain interlanguage errors.

Following Durrant (2023), precision and recall were calculated for these features using manually annotated data. This involved creating a coding template in CSV format (based on Durrant, 2023) and R scripts. A random sample of 80 texts from Courses 1-3 and the L2 level 5 corpus (excluding Courses 4-5 due to their small size) was chosen for manual annotation. MAT-tagged texts were converted to CSV for compatibility with the coding template. Finally, R scripts were used to calculate precision and recall based on the manual annotations.

While the Stanford parser via TAASSC offers tagging accuracy advantages due to its form-function approach, the linguistic variables it processed were not informative for identifying communicative functions in the factor analysis (refer to pilot multidimensional analysis in Section 4.5).

Since linguistic variables are governed by language rules or usage, a comprehensive set of syntactic features proved inadequate for this thesis's goal of finding functional dimensions. Therefore, the focus shifted to including variables that might contribute to identifying communicative functional dimensions marking academic written registers.

Informed by prior research and established hypotheses, I primarily used linguistic variables from Biber (1988). However, these features were refined based on three criteria:

1. Functional interpretability: Ensuring features could be linked to communicative functions.
2. Tagging feasibility: Focusing on features that could be accurately tagged.

3. Structural/functional distinction: Balancing capturing syntactic structure and functional meaning.

Due to their relevance to Study 3, only a subset of features used in the final multidimensional analysis (MD) were assessed for accuracy. These features included tags for stance function-related forms, which were relatively rare in the corpus. Manual re-tagging of the that-relative feature was conducted based on grammatical and syntactic definitions from "Grammar of Spoken and Written English" (Biber et al., 2021b).

While precision for the checked features was acceptable (87% - 98%), recall rates varied considerably (average 70%). The low recall rate (43%) for "that clauses controlled by nouns" led to its exclusion from the final factor analysis.

Accuracy rates for 'that-noun clauses' and 'that-adjective clauses' tagged by TAASSC and MAT are presented in Table 3-11.

Table 3-11. Accuracy rates (precision and recall) for that-complement clauses

Tagger: Feature	TAASSC 2.1.4: *that* noun clause	TAASSC 2.1.4: *that* adjective clause	MAT 1.3.8: *that* noun clause (THNC)	MAT 1.3.8: *that* adjective clause (THAC)
precision	0.4	0.88	0.98	0.87
recall	0.33	0.82	0.43	0.98

As noted by Nini (2019), the MAT tagging scheme does not distinguish between 'that-relatives' and 'that-complements.' This distinction was crucial for the analysis, so these features were manually retagged based on grammatical definitions (Biber et al., 2021). Consequently, precision for this feature was high, but recall was low.

Both TAASSC and MAT tagging showed low accuracy for "that-noun-complement" clauses. This highlights the need for further development to improve the tagging accuracy of this specific feature.

The tagging precision check revealed slightly higher accuracy for MAT-tagged features compared to TAASSC-tagged features. However, the focal features, particularly "noun-controlled that-complement constructions," were relatively complex and exhibited lower accuracy than other constructions.

As previously mentioned, one selection criterion for features was the ability to distinguish syntactic form and functional meaning. However, tagging accuracy was also considered. This is why MAT-tagged frequency data was ultimately used for the analyses in Chapter 6.

This section focused on evaluating two specific features for stance analysis. While this approach offers valuable insights, it necessarily excludes other features. Discussions regarding the adequacy of features analyzed in Chapters 4-6 are detailed in the methodology sections of those chapters.

3.5 Learner Language Analysis

Regarding the issues regarding tagging inaccuracy, another consequence of such operational definitions of tagger is that it may miss complex linguistic constructions.
As discussed above, the tag '*that*-relative clauses on object position (TOBJ)' was manually checked and the tag was re-coded as '*noun complement'* when *that* was in noun complement clauses position. The manual retagging process revealed that there are a variety of inaccurate forms that violate grammatical rules and therefore cause tagging inaccuracy. These instances required decision-making on categorising them, which was done based on a set of rules. These rules are established by drawing on grammar use patterns described by Biber et al. (2021) and examining typical anomalies found in examples in the data. This section discusses the established rules and underlying rationales for manually categorising the TOBJ feature into *that*-relative clauses and *that*-verb-complement clauses.

There is a wide range of POS annotation schemes (hereafter 'tagsets') that have been used in corpus linguistic research (e.g., Santorini, 1990) and annotation schemes for English syntactic dependencies (e.g., Universal Stanford Dependencies, de Marneffe & Manning, 2014). I use the Stanford Universal Dependencies scheme for dependency annotation as a reference point to the manual coding.

My initial approach to categorizing features as complement or relative clauses relied heavily on the operational definition provided by MAT for TOBJ (that-clauses as object complements of verbs). Since MAT doesn't differentiate between that-relatives and that-verb-complement clauses, it only tagged them all as relatives. In simpler terms, I used MAT's existing tags as a starting point and manually reclassified them further into the two specific categories. This method, however, had limitations as my focus was solely on refining the existing TOBJ tag, potentially missing constructions that weren't initially flagged by MAT.

This approach likely resulted in overlooking features that could have been either complement or relative clauses simply because they weren't included in MAT's initial tagging. It's important to note, however, that these missed features might not necessarily represent valid constructions according to grammatical rules. Instead, they were more likely grammatical errors that MAT didn't capture. Given the significant effort required for comprehensive identification and the fact that these errors weren't the primary focus, further investigation wasn't conducted to definitively categorize every missed instance.

3.5.1 Case 1: Decision made based on a Surface-level definition

(1) *the fact that these thing will affect in the environment by the years*.

-<u>*Text id: 45150*</u>

The surface structure of the sentence taken from Sentence (1) is closer to a relative clause where 'the fact' is controlling '*that* clause'. However, it seems that the writer intends to use the construction to complement the meaning of 'the fact' in the following *that* clause. While the intended function of the construction seems to be 't*hat* complement', the case in text 45150 was considered as a relative clause, considering its surface structure, following the definition of the MAT.

While it's true that incomplete learner language can lead to inaccurate frequency distributions and potentially mislead interpretations, there's a trade-off to consider. In this case, the tagging process prioritized consistency with the MAT even if it meant potentially miscategorizing some that-clauses between complement and relative forms.

This consistency ensures the results at least provide a reliable picture of the learner language, albeit with some limitations. The key is to be aware of this potential ambiguity and interpret the findings cautiously. While a more nuanced approach to classifying that-clauses might be ideal, the significant effort required for such precision may not be practical for the current purposes.

3.5.2 Case1-1. Surface structure regarding adverbial vs relative

(2) Happiness is very important thing in own life *that we spend all things to get some of happiness*. -*Text id: 36661*

The function of the italicised form is not clearly distinguished between relative or complement, as *that* clause is structurally complete with a subject, a verb and an object but the preceding noun 'life' seems to be better connected with adverbial relatives such as *because*. Nonetheless, it was categorised as a complement considering the surface structure (being complete in itself with the necessary arguments). The vagueness regarding the categorisation, in this case, arises from the mismatch between the surface structure and their function in the deep structure; sometimes confusion may occur even with controlling nouns, such as 'fact' or 'reason', typically controlling *that* complement. For example, consider the following example sentences consisting of the two nouns, fact and reason, respectively.

(3) the fact that it came true

(4) the reason that it came true

The types of structures as exemplified in (3) are usually categorised as complements while those of (4) are as adverbials. Their functions are different, and sometimes the differences are visible in variations of the forms of the adverbials. For example, sentence (4) can have alternative forms such as '*the reason why it came true*', '*the reason for which it came true*' or '*the reason it came true*' while sentence (3) does not have such variations.

Considering the meaning of the sentence in Sentence (2), *that* seems to be used as an adverbial relative connector, and therefore this form may be associated with the adverbal use, not belonging to either of the two categories of relative or complement; however, in the manual tagging, these instances were treated as if they were all complements as the purpose of the tagging did not require further specification between adverbial relatives and complements and their surface structure is not distinguished if the meaning and function of the form are not considered. A similar source of vagueness in categorising the form is found in the following example.

(5) their life *that no one can take care of the family*. - Text id: 3994

The italicised part in Sentence (5) was also categorised as a complement. This case seems to be another example of a learner's confusion with the use of the concrete noun 'life' in the adverbial sense. These instances indicate that the writers are using the construction with nouns having concrete meanings when abstract nouns are required. Due to its inappropriate use of controlling nouns, the intended meaning of the construction is not clear to the reader, as it opens a possibility for the construction to convey different functions. Despite such vagueness, the use of these constructions indicates the writer's attempt to use 'noun that clauses', requiring the writer's ability to elaborate on the intended meaning in an extended structure.

3.5.3 Case2. Decisions made based on deep structures

When a construction seems to have a surface form that can be identified with either of the two structures, they are considered in terms of their meaning and intended function conveyed by the structure in a sentence. Consider the following example sentence.

(6) things *that the like doing to spend their free time on*. -Text id: 713

The *that*-clause in Sentence (6) was marked as relative. Since this construction is not easily categorised into either a relative or a complement at the surface level, it was looked at holistically; the function that performs in the sentence indicates conjunction, not a demonstrative, which leads to confirmation that it should be either complement or relative. Then, the meaning of that complement leads to guessing that the writer intended to use they instead of the, which seems to be a typo considering the rest of the clause. Therefore, the intended meaning of that clause led to the decision that the preceding noun *things* was the object of the verb like in *that* clause, which led to the decision that it was a relative clause, not a complement.

3.6 Conclusions

While the accuracy checks do not fully cover all the features analysed and are done on a limited number of samples, some implications emerge from these accuracy analyses. First, the accuracy rate of the features seems to conform to the previous reports in studies discussed in the literature review. While many of the features showed a reasonably acceptable accuracy, some of the more structurally complex features, such as the two *that*-complement clauses, showed a relatively lower rate of accuracy. Second, the notably poor accuracy rate of one of the *that*-complement clauses, *that*-noun-complement clauses, seems to require further investigation into why its accuracy is so poor. MAT performs better with the data, so it has been adopted for the analysis in Chapter 6.

As for the learner language analysis reported in Section 3.5, concrete vs abstract noun distinction may be worth further investigation. From my observation, one cause of systematic tagging inaccuracy seemed to derive from the writers' using 'the' instead of 'they', among some other similar writers' ungrammatical language use being the source of the tagging error. This type of error identification is limited as it was not systematic, only done on one mutually re-tagged feature.

Chapters 4, 5 and 6 present the specific research questions, methods, and findings drawing on the general hypotheses and methodological approaches discussed in this chapter. Chapter 4 presents MD analyses, addressing the aim of exploring the syntactic complexity features associated with functional dimensions inherent in the L2 writing developmental courses. Chapter 5 discusses a collostructional analysis of verb-argument construction, as part of the complementary analysis of the MD analysis. Chapter 6 presents a mixed-effect model of stance device distribution across time points, as a subsequent analysis of the MD analysis.

Chapter 4 Multidimensional Analysis of Syntactic Complexity Development in L2 Writing

4.1 Introduction

This chapter explores the development of English syntactic complexity by focusing on co-occurring patterns in linguistic measures in written texts produced by writers of English as another language (L2) over time. This chapter aims to answer the question of what communicative functions are associated with the syntactic features as markers of academic writing development in the L2 longitudinal data, using the multidimensional (MD) analysis method (Biber, 1988).

Many studies discussed in Chapter 2 have shown that academic writing has distinctive features that separate it from other discourses. Spoken and written texts involve different types of syntactic complexity (Ortega, 2015, p.90). For example, phrasal structures functioning syntactically as modifiers of noun phrases are more frequent in written academic registers, while spoken texts include more finite clauses and verb + to constructions (Biber et al., 2014; Biber et al., 2020a).

Previous MD studies also contribute to this spoken vs. written register variation.

Moreover, Learner-written corpora studies using the MDA method provided evidence of phrasal complexity as a marker of writing development. The following section of the literature review presents a selection of the previous MD study findings and their implications for studying L2 academic registers.

This chapter discusses three MD studies (Sections 4.4, 4.5 and 4.6) aiming to capture the functional dimensions of syntactic complexity features in the L2 longitudinal writing corpus. As MD studies require complex analysis procedures, it was essential to revise the linguistic variables to produce meaningful and statistically reliable interpretations. For that aim, several pilot analyses were conducted, and two among them are introduced in the subsequent sections. Even though the linguistic variables and some of the analysis procedures for these pilot analyses were not finalised, the analysis results provided information on the grammatical variations of the longitudinal PELIC corpus, which helps understand the final analysis result. In addition, these pilot analyses' results provided the rationales for revising the linguistic features that are suitable for the data being analysed in this thesis.

Among the three studies to be discussed in this chapter, the two pilot studies (4.4 and 4.5) tried different sets of linguistic variables to examine the interpretability of the results, leading to the final set of 73 lexicogrammatical features used in the final study (4.6). These 73 features incorporate syntactic features that reflect the phrasal and clausal functional dimensions in order to capture the complexity specific to academic written registers. Subsequent mixed-effects models were run to obtain information on the significance of the results while ruling out situational and writer-specific factors from the linguistic variation over time. The analysis enables us to find the genre/register-specific characteristics of the L2 data from the form-function mapping unique to the L2 writing. Much of the literature review in Section 4.2 and the analysis discussed in Section 4.5 is also detailed in Kim (forthcoming).

4.2 MDA Literature Review

MDA is based on the notion of form-function mapping to interpret co-occurrences of linguistic features in terms of their shared functions. The ample volume of the previous MDA studies in academic registers in general and L2 learner corpora in specific provide important reference points to compare and contrast. The MD analysis method helps find similarities and differences among different texts, and such differences provide data upon which to interpret the functional aspect of the texts. The previous MD studies have contributed to understanding different types of register variations such as cross-linguistic variations (e.g., Kim, 1990; Biber & Hared, 1992; Biber, 1995), genre variations (e.g., Biber, 1988), disciplinary variations (Hardy & Römer, 2013). Many studies also used MDA with a focus on academic registers (Biber, 2006), academic level (Biber, Conrad, Reppen, Byrd & Helt, 2002), academic genre (Nesi & Gardner, 2012; Hardy & Friginal, 2016), or academic disciplinarity (Biber et al., 2002; Biber, 2006; Hardy & Römer, 2013). This section aims to review the substantive applications of MDA in previous research by cross-checking the study designs and results regarding similar interests or corpora.

The MD analysis method has also been used to elucidate the multidimensionality of syntactic complexity features that significantly co-occur in academic registers comprising expert, L1 or L2 writing (e.g., Biber, 1992; Gardner et al., 2019; Gray et al., 2019). The spoken and written register analysis shows that they are fundamentally different (Biber, 1988; Biber, 1995; Biber et al., 2021b). Some of the L2 studies show that novice L2 learners tend to use linguistic features similar to spoken register (e.g., Kobayashi & Abe, 2016; Kim & Nam, 2019). Based on these studies, the proposed study expects to find evidence that L2 writing progresses to show more linguistic use associated with written registers.

This literature review consists of three subsections. Section 4.2.1 overviews the MDA studies exploring academic register variations. Sections 4.2.2 and 4.2.3 review MDA studies by registers: general written registers, university student registers, and L2 adult learner writing. The focus of the reviews in the last two sub-sections is on learner corpora studies. The review of the MDA studies delineates different functional dimensions in different academic written registers and discusses the implications and potential areas for further investigations arising from it. The frequently discussed linguistic features and associated functions in student writing can indicate typical characteristics of a specific register, which has in common in terms of situational contexts where the writing took place.

4.2.1 Overview of Multidimensional Analyses of English Academic Writing

In terms of its methodological approach, multi-dimensional analysis (MD; Biber, 1988) is inductive in that it uses statistical methods to find patterns of a range of grammatical features and then identify the underlying discourse functions. The texts analysed may consist of different registers (e.g., spoken and written registers) or more specialised subregisters (e.g., academic textbooks used in different levels of education) (Biber & Gray, 2016, p.74).

As for the grammatical features used in this method, Biber and Gray (2016, p.78) suggest the three distinctive grammatical classes frequent in academic writing: nouns, adjectives, and prepositions. These grammatical classes are more frequent in academic prose than in other registers. In addition, many related specific features are especially characteristic of academic prose (e.g., nominalisations, noun phrases with multiple modifiers, stance noun + of-phrase). In contrast, verbs, adverbs and adverbials are usually less common in academic prose, even though the specific types of these grammatical classes serve unique functions in academic prose.

Some linguistic features of research value (e.g., hedging) are not necessarily more frequent in an academic register. Instead, they are salient because they perform special discourse functions in academic writing (Biber & Gray, 2016, pp.76-77). In that respect, linguistic forms not particularly frequent but co-occur with other features to perform specific communicative functions may go unnoticed in quantitative approaches. In other words, looking for linguistic features that share the same communicative functions in a register may capture the specific characteristics of academic discourses.

Another way to define the methodological characteristics of MD analyses is whether the research interest is diachronic or synchronic variations of linguistic features in corpora. Most MD analyses reviewed in this thesis focus on synchronic variations among academic subregisters or comparing academic registers with different registers. MDA studies have been used to elucidate linguistic variations that differentiate groups of texts associated with different situational contexts.

Biber (2019) notes that “oral" versus "literate" discourse is found to be the most critical dimension that emerged in previous MDA studies regardless of the research focus or the corpora. This distinction has been attested in many of the studies of syntactic complexity discussed in Chapter 2. However, this distinction is relative depending on the specificity of the sub-registers to be analysed (e.g., Biber, 2006; Biber & Gray, 2016). For example, academic prose more frequently uses nouns, adjectives, and nominalisations than fiction and newspapers (Biber & Gray, 2016, p.105). However, popular science employs more nominalisations than any other sub-register, while specialist science research articles make minor use of these devices (Biber & Gray, 2016, p.113).

Another example of this relativity of prominence is found regarding stance functions. Biber (2006) notes that stance features are predominant in university registers in comparison to the more general corpora studied by Biber (1988). Stance features are ubiquitously prominent on most functional dimensions emerging from more specialised corpora of university spoken and written registers (Biber, 2006, p.212). Overall, stance features were crucial to identifying how academic written narratives are represented in these very specific sub-registers.

Similarly, academic written texts produced by students are likely to provide evidence of SC variations specific to student registers. For example, the salience of adjectival features found in university students' written texts is worth noting, as they are not typical in published academic writing (Hardy & Römer, 2013, p.193; Hardy & Friginal, 2016, p.124). The results of these MD analyses are an indication of heavy influence by the specificity of the registers represented in a corpus.

Multi-dimensional analysis (MDA, Biber, 1988) has been used to explore multi-faceted communicative dimensions in register variations, primarily the spoken and written register distinction in 23 different spoken and written genres. The following subsection 4.2.2.1 discusses the findings and significance of the MDA study of English spoken and written registers by Biber (1988). As will be discussed in Section 4.2.3, many studies have shown that the spoken-written distinction is a marker of writing quality in student registers. Therefore, the MDA findings of English spoken and written registers provide foundational discussions for establishing assumptions regarding L2 writing development.

Section 4.2.2.2 focuses on MD analyses that explored two English student writing corpora collected in the U.K. and U.S., respectively. The British Academic Written English (BAWE, Alsop & Nesi, 2009) and the Michigan Corpus of Upper-Level Student Paper (MICUSP, O'Donnell & Römer, 2012; Römer & O'Donnell, 2011) consist of university student writing.

4.2.2 A multidimensional study of spoken vs. written registers (Biber, 1988)

Biber (1988) described the similarities and differences of the spoken and written texts based on the linguistic patterns that are supposed to be associated with their communicative functions. He found six functional dimensions based on the patterns of the linguistic features co-occurring within each spoken and written corpus. MDA selects linguistic variables of potential significance before analysing data based on the previous literature regarding linguistic forms that are potentially important in accomplishing communicative functions in texts. Biber provides detailed accounts of his review of potentially important 67 linguistic variables for register analysis (1988, p.64). He organised these linguistic features into 16 major grammatical categories, on which he based his interpretations of the six functional dimensions in 23 English spoken and written genres. In MDA, interpretation of the dimensions extracted from factor analysis requires examining the shared functions of these co-occurring features (Biber, 1988, pp.104-105).

Among the seven dimensions found in Biber (1988), this section reports the interpretations of Dimensions 1, 2 and 5, focusing on written and spoken distinction characteristics. This limited scope of discussions is because these dimensions are first found in Biber (1986) and then confirmed in the subsequent study (Biber, 1988), gaining more confidence in their interpretation.

Dimension 1: involved vs informational production

Factor 1 represents a dimension marking high informational density and exact informational content versus affective, interactional, and generalised content, thereby labelled as 'Informational versus Involved Production' for the dimension underlying this factor (Biber, 1988, p.107).

The primary linguistic features loaded negatively on Dimension 1 include nouns, word length, prepositional phrases, type/token ratio, and attributive adjectives. Together, these five features contribute to integrating dense information into a text and communicating concise and precise ideas. First of all, a high type/token ratio indicates more lexical use in a text, likely reflecting more extensive use of words carrying particular meanings (Biber, 1988, p.104). Appropriate lexical use is a complicated production task and therefore is more likely to be observed in written discourse than spoken discourse. Likewise, longer words convey more specific, specialised meanings than shorter words used more frequently and correspondingly more general in meaning (Zipf, 1949). Another linguistic feature to be noted is attributive adjectives, which are used to further elaborate on nominal information. In other words, they pack information into relatively few words and structures in comparison to predicative adjectives or relative clauses.

Among the linguistic features with positive loading on Dimension 1, four types of subordinations are associated with loose information packaging, more common in spoken discourses (Biber, 1988, pp.106-107). Notably, subordination-based measures have received extensive attention as a marker of syntactic complexity in written discourse (Biber et al., 2011; Biber & Gray, 2013, p.68; Norris & Ortega, 2009). Further research may be beneficial in elucidating the relationship between syntactic complexity and quality academic written discourse with an expanded frameset of linguistic features to compare and contrast.

Overall, Dimension 1 represents a fundamental variation parameter among texts that mark spoken vs written register in terms of their production characteristics and primary communicative purposes.

Dimension 2: Narrative vs non-narrative concerns

Biber labelled Dimension 2 as ‘Narrative versus non-narrative Concerns.’ The narrative dimension involves considerable reference to past time, third person animate referents, reported speech, and depictive details. On the other hand, non-narrative concerns are marked by immediate time and attributive nominal elaboration (1988, p.109).

Only two features have large negative weights on Factor 2: present tense and attributive adjectives (Biber, 1988, p.109). Notably, the present tense and the past tense show a complementary distribution on D2. More specifically, present tense typically reports current events and does not mix with past tense in dealing with more immediate matters, representing non-narrative types of discourse. Thus, the co-occurrence of attributive adjectives and present tense verbs characterises a more frequent use of elaborated nominal referents in non-narrative types of discourse.

Dimension 5: Abstract vs non-abstract information

Dimension 5 is labelled as ‘Abstract versus non-abstract Information’, representing informational discourse that is abstract, technical, and formal.

In Dimension 5, passives are loaded as a significant linguistic feature, typically carrying abstract and technical meaning and associated with a more formal style (Biber, 1988, p.112). Conjuncts and adverbial subordinators frequently co-occur with passive forms to mark the complex logical relations among clauses that characterise this type of discourse.

The type/token ratio with negative weight (TTR; — .31) is notable. Biber interprets the negative loading of TTR in Dimension 5 as an indicator of low lexical variety than non-technical informational discourse with even more negative TTR loading in Dimension 1 (1988, p.112). In other words, the informational discourses with general high TTR (negative TTR factor loading) in contrast to interactive types of discourse further divide into non-technical informational discourse with more lexical variety (with far more negative factor loading on Dimension 1) and technical, informational discourse with less lexical variety (with relatively moderate negative factor loading on Dimension 5).

Regarding the text types identified by Biber, the most relevant text groups are clusters 3 and 4, called *‘scientific’ exposition* and *learned exposition*. Biber identified cluster 3 (scientific exposition) as a text type associated with primarily academic prose texts from natural science, engineering/technology, and medicine. On the other hand, cluster 4 (learned exposition) represents a relatively broader range of texts, including academic prose from the humanities, social sciences, education, and law (Biber, 1989, pp.27-28).

As has been implicitly indicated in the discussions so far, the functional dimensions emerging from MDA studies help identify the syntactic constructions mapped onto writing development. More direct support comes from Biber

(1992), who used syntactic complexity features to conduct confirmatory factor analysis to determine the relative strengths of a unidimensional model and several multidimensional models hypothesized on theoretical grounds. The finding indicates the multidimensional nature of discourse complexity (Biber, 1992, p.137).

4.2.3 English Student Registers: New Dimensions in BAWE and MICUSP

The British Academic Written English corpus (BAWE; Alsop & Nesi, 2009) consist of university student writing with good marks in British university contexts. This type of corpora provides information unique to the student register. The new dimensions found in these corpora can adequately describe the underlying linguistic characteristics of 'quality' university-level writing.

Nesi and Gardner (2012) explored genre variation in the BAWE corpus, using Biber's (1988) dimensions, to find that personal stance features were one of distinguishing dimensions from more expert academic writing, a consistent finding with previous studies in the student registers (e.g., Gardner, Nesi & Biber, 2019; Hardy & Römer, 2013). Building on the study, subsequent studies found that the lexical features associated with a personal stance are potentially specific to student writing and differentiated from a different stance typically found in professional writing (Nesi & Gardner, 2017). Utilising the frame of genres established by Nesi and Gardner (2012), Gardner, Nesi and Biber (2019) integrated the academic levels, genre families and disciplinary variation to find the new four dimensions of the BAWE corpus. Their findings suggest that various situational considerations such as genre, discipline, and level of study all had considerable influence on interpretations of the new dimensions within the corpus.

The Michigan Corpus of Upper-Level Student Paper (MICUSP, O'Donnell & Römer, 2012; Römer & O'Donnell, 2011) corpus was collected from final year undergraduates and first through third-year graduate students in 16 different disciplines. Hardy and Römer (2013) explored cross-disciplinary variations in MICUSP, finding new dimensions. Interestingly, disciplinarity exhibited interaction with student text types. For example, dimension 3 represents situation-dependent writing with nouns and passives commonly associated, highly loaded with a report from Philosophy and an argumentative essay from English (Hardy & Römer, 2013, p.197).

Hardy and Friginal (2016) used Hardy and Römer’s (2013) dimensions to explore genre variation in MICUSP. The salience of adjectival features found in the corpus (Hardy & Römer, 2013, p.193; Hardy & Friginal, 2016, p.124) is worth noting, as they are not typically found in published academic writing (Gray, 2011). Hardy and Friginal interpret these students’ predominant use of adjectives as a process of acculturation into the discipline. The use of adjectives becomes replaced with more nominal options as students increase their academic literacy.

Another MDA study by Friginal and Mustafa (2016) attempted to explore cross-linguistic features between U.S.-based and Iraqi research article abstracts, utilising three of Hardy and Römer’s (2013) four dimensions as primary points of comparison. Despite the limitation inherent in their use of dimensions extracted from learner corpora for their corpora composed by expert writers, their findings suggest potential distinctions in these dimensions regarding directness and argumentation in these two different groups of writers.

As discussed earlier, there have been observations on the variety in the kinds of stance and a greater reliance on stance features in academic written registers, such as adjectival and verbal predicates with extraposed that-clauses with the pronoun 'it' (e.g., it is likely that, it seems that) (e.g., Omidian et al., 2021). These structures allow writers to foreground their evaluative stance on the information presented in the following clause (Biber, 2006).

Stance verbs controlling to- and that-clauses occur alongside mental verbs, that-deletions, first-person pronouns and past tense verbs. Together, these indicate a 'Personal stance' (Gardner, Nesi and Biber, 2019). Personal stance features were consistently typical in the student register, differentiated from a different stance typically found in professional writing (Nesi & Gardner, 2017; Gardner, Nesi & Biber, 2019; Hardy & Römer, 2013).

4.2.4 MDAs of L2 written registeres

Studies indicate that the trajectory of linguistic development mirrors a speech-writing distinction, reflecting primarily spoken style in the early stages and then moving to a more written style in their writing (Biber et al., 2011). The MDA method has been used in L2 learner corpora research to learn about student language, distinct from the written register in general and the expert academic written register in particular. The first dimension emerging from MD analysis usually has the most substantial explanatory power among the dimensions extracted from factor analysis, as the first factor represents the most significant portion of the variance. Table 4-1 summarises the significant linguistic features of the first dimensions of some of the MD studies that explored L2 learner registers. A full summary of the MD study review including a description of key functional dimensions and linguistic and situational variables is found in Appendix 4-1.

Table 4-1. The functional dimension (Factor 1) in selected MDA studies of L2 learners' academic register

Corpus type	Corpus description	Dimension 1 (major linguistic features loaded on D1)	MD studies
Course work	university students writing (Brazilian)	Expression of personal opinion vs compressed procedural information (third person pronoun, stance adverb, stance noun + to/that clause vs premodifying noun, passive voice) (Goulart, 2021)	Goulart (2021)
	Master's theses, PhD dissertations, and research articles (L1 & L2 speakers)	Attitudinal vs descriptive (stance verb vs attributive adjective) (Pan, 2018)	Pan (2018)
	L2 student essays produced in a semester-long English programme at university (longitudinal)	involved focus vs informational focus (noun (concrete & place), preposition, attributive adjective) (Friginal & Weigle (2014)	Friginal & Weigle (2014) Crosthwaite (2016) Issitt (2017)

Answers with controlled topic/time	TOEFL iBT speaking & writing test answers	literate vs oral responses (Nouns, Prepositional phrases, noun + of phrase, adjectives, word length, Passives (Biber & Gray, 2013)	Biber & Gray (2013) Biber, Gray & Staples (2016)
	TOEFL iBT responses spoken and written (longitudinal)	literate vs oral responses (nominalisation, a noun with multiple pre-modifiers, PPs as abstract postmodifier / attributive adjective as a single modifier in a noun phrase, adjectival diversity) (Biber & Gray, 2013)	Gray, Geluso & Nguyen (2019)
	International Corpus Network of Asian Learners of English (argumentative essays)	Cluster 1: academic register (nominalisations, predictive modals, conjuncts) Kobayashi & Abe, 2016)	Kobayashi & Abe (2014) Kobayashi & Abe (2016) Kim & Nam (2019)
	L2 English test-takers written answers	Literate vs oral discourse (nominalisations and phrasal features such as attributive adjectives and passive voice, fewer stance bundles & more referential bundles) Yan & Staples, 2019) Compressed procedural information vs stance towards the work of others (Staples, Biber & Reppen, 2018) Involved, academic narrative production vs descriptive, informational discourse (Llosa et al., 2019)	Yan & Staples (2019) Staples, Biber & Reppen (2018) Weigle & Friginal (2015) Llosa, Grapin, Gfiginal, Cushing & Malone (2019)
	L2 Learner English corpus (Pakistani, part of the International Corpus of Learner English (Granger et al. 2002)	Advanced literacy (attributive adjectives, nouns, phrasal coordination, prepositional phrases, word length, nominalisation) (Rooy & Terblanche, 2009)	Rooy & Terblanche (2006) Rooy & Terblanche (2009)

[1)] studies drawing on functional dimensions established in Biber (1988)

This selection of MD studies on L2 student writing in Table 4-1 is categorised by the situational context of corpora compilation or production, such as EAP programme/university and postgraduate / language test. It is also marked when it is longitudinal data. Studies listed in Table 4-1 generally confirm the findings from MD studies on academic registers, the most fundamental spoken and written distinction. Longitudinal MD studies are relatively rare, with the period of writing production relatively short. For example, the longitudinal studies exploring L2 corpora collected from EAP contexts (e.g., Issitt, 2017; Crosthwaite, 2016) were usually interested in a semester-long instruction effect on linguistic variations, which became similar to the written register over time. Notably, one longitudinal study using TOEFL answers found more literate language variation, utilising noun phrase complexity features to a greater extent over time (Gray et al., 2019, p.20).

The following two sections (4.2.3.1 and 4.2.3.2) discuss some of the MDA findings in L2 student registers listed in Table 4-1. The MDA studies in these sections are sorted by methodological approaches: 1) studies using established dimensions in previous MDA studies (e.g., Biber's (1988) spoken vs. written register) and 2) studies exploring new dimensions in the corpus of interest by conducting exploratory factor analysis.

4.2.4.1 L2 learner writing studies based on Biber (1988)

This section introduces some of the studies of L2 learner writing using the previously established functional dimensions identified in 23 genres of spoken and written texts (Biber, 1988). These dimensions were initially used to elucidate the situational difference between typical speaking and writing discourses. Academic written texts, for example, are found to have more nouns and adjectives than spoken texts, which was interpreted regarding the absence of interlocuter and dense information packing typical in written context (Biber, 1988).

As introduced in Step 6 of the general MDA procedures in Section 3.4.6, factor analysis is part of MDA procedures. However, the dimensions established in Biber (1988) have been used as a reference point for register studies (Berber Sardinha et al., 2019, p.166); Goulart & Wood, 2021, p.128). Subsequent studies have used the dimensions to analyse specific academic registers in greater detail (e.g., Biber, 2006). This study has also provided a baseline for the study of writing development, comparing the characteristics of essays written by students at various levels to the characteristics of general written registers studied in Biber (1988).

Research findings have shown that academic written texts produced by lower linguistic proficiency tend to show spoken discourses more frequently (e.g., Nam, 2017; Kim & Nam, 2019; Staple et al., 2016). I present two MDA studies that explored the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2019), using the 67 linguistic features in Biber (1988) for academic register analysis in L2 learner corpora (Kobayashi & Abe, 2016; Kim & Nam, 2019). It should be noted that Kobayashi and Abe (2016) did not adopt the dimensions found in Biber (1988) nor conduct factor analysis as usually done in MDA studies. Instead, Kobayashi and Abe used correspondence analysis, a statistical technique that identifies frequency-based associations between corpora and variables, to identify functional dimensions in the corpus. They argue that these methods improve the replicability of their findings compared to factor analysis. Kim and Nam (2019) also made methodological advancements, adopting mixed-effect models in statistical analysis to control other mediating effects such as L1 background or academic major in exploring linguistic variations in written texts produced by L2 learners of different English proficiency.

One notable longitudinal MDA study adopting Biber's (1988) framework is Crosthwaite (2016), who used Biber’s (1988) dimensions to explore written EAP essays and reports collected in the Hong Kong university context through three semesters. He used the longitudinal writing data to explore the effects of instruction on linguistic features associated with literate academic writing. Notably, the data used in the study were produced under timed-test conditions, which is likely to derive a test effect. The study also confirms that learner writing became more associated with academic writing over time. I am not aware of any MD analysis exploring PELIC corpus, the main data used in the present study. The study adopting a similar approach to these reported studies is discussed in Section 4.4.1.

The findings from these studies regarding linguistic variations in the ICNALE corpus generally confirmed the previous argument that L2 learners tend to use features more characteristic of spoken registers than generally expected in academic written texts. With interest in writing quality by different writers’ linguistic proficiencies, research findings have shown that academic written texts produced by lower linguistic proficiency tend to show spoken discourses more frequently (e.g., Nam, 2017; Kim & Nam, 2019; Staples et al., 2016). Notably, Kobayashi and Abe found considerable variations associated with the L1 background of the learners (2016, p.10). More specifically, the register awareness of Hong Kong learners was higher than that of other learners in East Asia, which indicates the influence of linguistic backgrounds. The findings in the studies discussed so far inform unique linguistic patterns in academic written texts produced by writers of a homogenous linguistic background and collected from such linguistic contexts.

Another point to note is the underlying primary interest of their analysis relying upon the established framework (Biber, 1988), built around 23 spoken and written genres of English texts. Relying on the established dimensions renders the new research interest in line with the potentially meaningful linguistic feature used in the previous research. It follows that these studies adopting Biber's (1988) dimensions reflect their research interest in the spoken vs. written distinction found in this previous work. Basing the interpretations of linguistic patterns on previous findings can provide comparable points, which can be primarily an appropriate choice if the corpora consist of similar situational or genre characteristics that are a vital concern of the study (Biber, 2006). For example, Aguado-Jimenez, Perez-Paredes and Sanchez (2012) explored spoken academic register, using the dimensions adopted from Biber (1988). Their decision to adopt the dimensions (Biber, 1988) can be considered appropriate in that their interest was in spoken academic register appropriateness in learner corpora. Moreover, they collected spoken texts from both learners and native speakers. Using a fully comparable corpus, they could have more data regarding spoken register variation in their learner corpus.

In summary, the findings in the studies discussed above consistently show that L2 learner writing uses more of a spoken register. The linguistic features associated with the spoken register can provide the foundation on which L2 development in academic writing can be explored further. In such a sense, considering the influence of situational factors inherent in a specific corpus, dimensions found in learner corpora targeting academic contexts can provide a valid basis to accumulate or add the knowledge of the academic register. The following section discusses some MDA studies that explored new functional dimensions in L1 and L2 learner corpora.

4.2.4.2 New Dimensions in L2 Registers: Longitudinal and Cross-sectional Corpora

This section presents some of the MDA studies of L2 learner corpora, such as English test answers, using the MDA method to identify the emerging functional dimensions in L2 writing. As Biber (2006) notes, it can be appropriate to conduct factor analysis or adopt established dimensions from previous MDA studies depending on the nature of research inquiries and corpora of interest.

One of the notable among the MDAs of L2 writing development is that spoken vs written discourse distinction was often used as an index of development in writing. Studies exploring the L2 learner texts collected from TOEFL iBT data found a fundamental spoken and written distinction in both written and spoken answers (e.g., Biber & Gray, 2013) and written answers only (e.g., Weigle & Friginal, 2015). For example, Biber and Gray (2013) examined the linguistic features and their correlations with the scoring, writing and speaking mode, and task types of the TOEFL iBT test answers using MDA analysis. Notably, this dimension significantly differentiated TOEFL iBT score levels, showing that spoken and written register awareness is a crucial quality of academic writing (Biber & Gray, 2013, p.57). The general findings regarding functional dimensions show a clear distinction between written and spoken modes. Furthermore, the co-occurring linguistic patterns were a more effective predictor of the scoring than any individual linguistic variables (Biber & Gray, 2013, pp.50-57). These results show that the co-occurrence of linguistic features can effectively explain the quality of texts valued in academic texts in different modes (spoken vs written). The ICNALE corpus provides learner writing data created under strictly controlled situations for the writing prompt and time, similar to language testing situations. The findings from two studies (Kobayashi & Abe, 2016; Kim & Nam, 2019) are consistent with the claim that academic written texts produced by lower linguistic proficiency tend to show spoken discourses more frequently.

In another theme of stance functional dimensions, many studies have effectively shown that stance helps show L2 student-specific variations in academic writing (e.g., Staples, Biber, & Reppen, 2018; Weigle & Friginal, 2015; Pan, 2018). The linguistic features in this dimension generally include *verb, noun, adjective* + *that*-complement clauses and mental verbs (e.g., I think that. . .). Studies using TOEFL iBT answers showed that higher-proficiency test-takers tended to make less use of the stance dimension (Weigle & Friginal, 2015; Staples et al., 2018). These features were often associated with interactive narrative writing, a less commonly found type in academic writing (Weigle & Friginal, 2015, p.34; Stapes, Biber & Reppen, 2018, p.312). The assumption from this is that a stance function is likely to decrease in longitudinal observation in student academic writing.

Pan (2018), exploring undergraduate and postgraduate work, also suggests a stance dimension also effectively differentiates L1 and L2 writing in applied linguistics. L1 writing tends to be more attitudinal, employing a wider range of linguistic features to express stance, compared to L2 writing (Pan, 2018, pp.122-123). Howver, when comparing different data consisting of highly specialised and different corpora, care needs to be taken. Pan (2018) reports rather unexpected findings regarding L1 disciplinary writing compared to L2 syntactic complexity development. Specifically, L2 academic writing is consistent but slightly more phrasal than L1 academic writing, with higher scores on dimensions of less attitudinal, narrative, and academically involved types of writing. He attributes this reversed trend from previous research findings to the higher English proficiency of the L2 students in Linguistics than general L2 students (Pan, 2018, pp.127-128). As noted earlier, stance is not a more prominent function in academic writing, which is more common in conversation. This distinction indicates the stance dimension as a marker to gauge the ability to present propositions in acceptable ways by using appropriate stance expressions. In a more recent study of L1 vs L2 English writing development, Staples et al. (2023) also note that L2 writing is more phrasal when compared to L1 writing in a similar context while a broad trend of grammatical complexity development was consistent in both L1 and L2 writing. These studies make it important that the distinction between L1 and L2 studies is to be understood in a relative term being on a continuum, rather than to be distinguished on a binary term.

Moreover, previous studies indicate that the affluence of linguistic resources associated with the functional dimension often differentiates writing quality or linguistic proficiency (e.g., Rooy & Terblanche, 2009; Gregg et al., 2002; Pan, 2018). In other words, the functional use of linguistic features appears to be similar across the writing by different linguistic backgrounds or writing maturity. However, the linguistic resources to fulfil the communicative functions often distinguish different writer groups. This is the reason for adopting a different analysis method for subsequent analysis of the MD analysis method to further analyse the head verbs in the verb-that clause constructions in Chapter 5.

Finally, studies consistently report on the influence of genre, discipline and level of study or L2 proficiency on interpretations of the dimensions in both L1 and L2 student register (e.g., Biber, 2006; Gardner, Nesi and Biber, 2019; Hardy & Römer, 2013; Hardy & Friginal, 2016; Biber & Grey, 2013; Gray et al., 2019; Reppen, 2017). Sometimes genre influence is pointed out as a potential factor to overshadow study level differentiation (Nesi & Gardner, 2012, p.14). Similarly, registers reflecting the communicative purposes unique to the discourses are more influential than learner-specific variables such as first language background (Rooy & Terblanche, 2009; Larsson et al., 2021). In conclusion, it is not yet clearly known regarding the direction of influence or any causality between these factors; what seems very clear is their mutual influence on each other.

Overall, these MDA studies exploring new dimensions showed considerable compatibility with the dimensions in Biber (1988). However, situational and learner variables (Adel, 2015, p.409) influenced the interpretations of dimensions found in these student registers. For example, as discussed in Section 2.2.2.1, the BAWE and the MICUSP corpora, student academic written assignments in a university setting, exhibited linguistic features distinct from academic experts. Such student writing variations indicate learner registers in their developmental stages and the influence of communicative purposes inherent in student genre, broadly categorised into situational variables.

As discussed in previous chapters, studies using longitudinal learner corpora can contribute to the understanding of linguistic development, but longitudinal MDA is relatively scarce (Friginal & Weigle, 2014; Crosthwaite, 2016). One such longitudinal MDA study is Crosthwaite (2016), who used Biber’s (1988) dimensions to explore written EAP essays and reports collected in the Hong Kong university context through three semesters. On the other hand, Friginal and Wiegle (2014) found four new dimensions in longitudinal corpora of L2 essays, using exploratory factor analysis. The findings in these two studies generally confirm *time* as a significant variable in academic written register variation.

The discussions in Chapter 2 have shown that the spoken vs written register distinction is the key to understand the writing development in L2 research. Likewise, the previous MD analyses discussed in this section also reveal the oral vs. literate distinction as major functional dimension in L2 writing research (e.g., Biber et al., 2016; Weigle & Friginal, 2015). Furthermore, in a similar trend to the reported research in Chapter 2, the findings of many MD analyses note that phrasal features such as attributive adjectives and passive voice constitute an essential component of typical academic writing. Therefore, a good proportion of variance/covariance among linguistic features is likely associated with the informationally-driven functional dimension in L2 learner written corpora.

A methodological point to be made with these MDA studies exploring new dimensions has to do with their selection of linguistic features. These studies using learner corpora tended to significantly reduce the number of features used for factor analysis. For example, Biber and Gray reduced the 128 features they used for preliminary linguistic analysis into 28 linguistic features for MDA analysis. Selecting a smaller selection of linguistic features may provide a more precise pattern of co-occurring patterns of the linguistic features, leading to a clearer picture of emerging dimensions (refer to Appendix 4 for the number of linguistic features selected in the MDA studies discussed in this paper). Moreover, such a selection process may reflect the nature of major communicative functions in the corpus of focus. It which should be worth examining and comparing the clusters of linguistic features associated with dimensions across studies using different registers, such as L1, L2 and more general academic written registers collected from different situations and writers.

4.2.5 Conclusions

The Multidimensional analysis (MDA) method (Biber, 1988) can provide a way to integrate the multifaceted textual characteristics and various situational considerations into analysis and interpretation. Moreover, as Multidimensional analysis (MDA) is a systematic method extensively applied to register analysis, it can provide a valuable tool to analyse the linguistic variations in L2 learner academic written texts regarding L1 learner writing or expert writing. Therefore, MDA methods and the linguistic assumptions underlying the method will be the primary methods for this study in exploring linguistic development in L2 academic writing in a university setting. Some of the studies exploring academic registers using MDA will be discussed in the following section.

Employing the MDA method helps obtain information on both the frequency and co-occurrence frequency of key linguistic features. First, we can gain the individual frequency data to inform the relative saliency of the features. In addition, the co-occurrence frequency provides evidence of a correlation between the features and therefore underlying latent factor which is more interpretable in regard to the macroscopic nature of the text - the communicative purpose and writer's linguistic choice for accomplishing the purpose.

MD analysis using longitudinal learner corpora is relatively scarce (Friginal & Weigle, 2014; Crosthwaite, 2016). As discussed earlier, Crosthwaite (2016) found linguistic variation in time points towards a more academic register, with a significant effect on features such as conditional adverbial subordinators by adopting Biber's (1988) dimensions. On the other hand, Friginal and Weigle (2014) found four new dimensions in longitudinal corpora of L2 essays, using exploratory factor analysis. The findings in these two studies generally confirm *time* as a significant variable in academic written register variation regardless of the dimensions explored in L2 written texts.

Based on the understanding of L2 syntactic development in writing, this study aims to add more longitudinal evidence to our understanding of grammatical development in L2 academic writing. With that aim, this chapter explores the following research questions:

1) How do the linuistic complexity features and the functional characteristics associated with these features in L2 English writing systematically vary across different academic semesters?

2) Do the linguistic complexity features and the functional dimensions

associated with academic written registers become more salient over time in L2 writing?

One of the aims of this study, as indicated in the second research question above, is to connect the body of literature on syntactic complexity drawing on more traditional types of measures (elaboration of dependents) and literature on writing development research using measures of linguistic features indicative of register-specific characteristics. In other words, the current study adopts linguistic measures other than those of more traditional syntactic complexity measures to provide a different angle of perspectives on grammatical complexity development in L2 writing.

The following sections discuss three MDAs employing different linguistic features for factor analyses. The MD Analysis carried out on a corpus consisting of general genres by Biber (1988) is considered to be inclusive of many different spoken and written situational contexts, as a general model of the English language (Nini, 2019, p.77). The pilot study in Section 4.3 draws on the functional dimensions identified from Biber. (1988). The incorporation of these previously established dimensions is valid when addressing research concerns such as mapping a more specific register against more general reference points. An example of such applications is Crosthwaite's (2016) analysis of a longitudinal corpus of students' writings in a language program for English for Academic Purposes, in which Biber's dimensions are used to plot student texts onto Biber's model in order to assess their progress. Such applications reveal the possibilities that past MD studies offer and should encourage researchers to use this methodology to make their models available and applicable to other researchers (Nini, 2019, p.92). On the other hand, considering my key enquiry is 'longitudinal development', which is likely to be subtle in terms of linguistic variations over such a short term as one year, EFA was also considered to be appropriate to find the characteristics of the L2 writing data as a stand-alone register in the second pilot study (Section 4.5) and the final study (Section 4.6). A general conclusion drawn from these MD analyses is discussed in Section 4.6.5, leading to the research aims of stance analyses in Chapter 5.

4.3 Pilot Study 1 – Analysing the L2 written register using established spoken vs. Written functional dimensions

The general aim of this pilot MDA study is to explore functional dimensions established in Biber (1988) in the L2 longitudinal writing. More specifically, this study explores linguistic features and functional dimensions of written texts over three to five semesters. The following section describes the data sample used in the study. This study uses the MAT tagger (Nini, 2019), a replication of the Biber tagger (1988) used to tag and analyse grammatical features in texts. MAT uses each feature, the mean frequency for that feature and the standard deviation of that feature in Biber's (1988) corpus for calculating scores for the functional dimensions (Nini, 2019, p.71). The Multi-Dimensional Analysis Tagger (MAT; Nini, 2019) has been tested in terms of reliability to replicate the linguistic models in Biber (1988).

MAT tagger performs Steps 3-8 of the multidimensional analysis delineated in Section 3.3.4. MAT tagger includes a copy of the Stanford Tagger (Toutanova et al., 2003), which performs part-of-speech (POS) tagging in Step 3. MAT then conducts syntactic tagging on the POS-tagged texts by identifying the 67 linguistic features used in Biber (1988). This means that this study did not conduct Steps 2 and 7 among the eight steps. This methodical simplification is to directly compare the findings with the dimensions in Biber (1988).

Before running the analysis, MAT asks users to select the options for analysis performance and visualisation outputs. It should be noted that the number of tokens for type-token ratio (TTR) calculation is set as 400 by default, which means that the tagger considers only the first 400 tokens of the text and counts the occurrences of types in these 400 tokens (Nini, 2019b, p.25). This variable is included in the calculation of Dimension 1 only when choosing the default number for compatibility with Biber's (1988) calculations. Also, the value was appropriate for this pilot study considering that the average length of the texts used in the analysis is 386 (refer to Table N in Section 3.2 for the corpus composition).

Another point of option setting is the z-score correction. If the user has chosen to use the z-score correction, then these Dimension scores reflect the choice. The z-score correction option checks the z-scores and changes them to 5 if the absolute value of the magnitude is higher than 5. This correction intends to prevent inflation of the overall Dimension scores by a few very infrequent variables. The MAT manual advises using this option for very short texts, and this study tried both the default setting and the option for z-score correction for examining purposes. Only the results of using the z-score correction are reported considering that the minimum text length chosen for this study is relatively short (minimum 100 words per text).

4.3.1 Linguistic features identified in five PELIC sub-corpora

The MAT analysis results include the following statistical data among others: (1) the mean frequencies of each of the 67 measured linguistic features (part of the features shown in Appendix 3-1), (2) the standardised statistics for each linguistic feature including Z-scores, calculated based on the means and standard deviations presented in Biber (1988, p.77).

Table 4-2 below shows the Z-scores of selected linguistic features from the 67 tagged features by the five Courses in the PELIC corpus. It also shows the scores of the academic prose genre in the LOB corpus, as reported in Biber (1988).

Table 4-2. Z-scores of linguistic features in Dimension 1 in the PELIC sub-corpora (features are ordered by the factor loadings in Biber (1988, p.103))

Features positively loaded in Dimension 1		course1	course2	course3	course4	course5
1	private verbs	-0.32	-0.05	-0.18	-0.19	-0.55
2	THAT deletion	-0.25	-0.21	-0.20	-0.25	-0.73
3	contractions	-0.39	-0.35	-0.49	-0.65	-0.61
4	present tense verbs	-0.09	0.01	-0.31	-0.27	-0.50
5	2nd person pronouns	0.44	0.85	-0.11	-0.11	-0.70
6	DO as pro-verb	-0.29	-0.27	-0.31	-0.21	-0.70
7	demonstrative pronouns	-0.19	-0.12	-0.17	-0.44	-0.78

8	general emphatics	0.68	1.06	0.72	0.62	-0.10
9	1st person pronuns	0.56	0.28	0.23	0.10	-0.08
10	pronoun IT	0.21	0.15	-0.16	-0.04	-0.40
11	BE as main verb	-0.41	-0.55	-0.65	-0.26	-0.87
12	causative subordination	2.00	2.12	1.94	2.13	0.80
13	discourse particles	-0.39	-0.34	-0.28	-0.42	-0.52
14	indefinite pronouns	-0.64	-0.55	-0.63	-0.64	-0.70
15	amplifiers	0.44	0.06	0.04	-0.29	0.17
15	general hedges	-0.21	-0.03	-0.20	-0.26	-0.30
17	sentence relatives	0.80	1.20	2.31	2.44	3.79
18	WH questions	0.24	-0.07	0.33	0.84	-0.33
19	possibility modals	0.72	1.58	1.13	0.90	2.14
20	non-phrasal coordination[10]	0.22	0.15	0.00	-0.34	-0.32
21	WH clauses	0.11	0.33	0.53	0.92	0.33
22	final prepositions	-0.57	-0.56	-0.57	-0.65	-0.74
Features negatively loaded in Dimension 1						
1	nouns	1.30	1.27	1.25	1.04	0.00
2	word length	-0.20	0.23	0.05	0.01	-0.68
3	prepositions	-1.01	-0.78	-0.97	-1.19	-1.88
4	type/token ratio	-3.70	-2.89	-3.49	-3.24	-4.12
5	attributive adjs.	0.20	0.41	0.20	-0.15	-0.17

As seen in Table 4-2, the positively loaded features include 22 lexico-

[10] independent clause coordination in MAT

grammatical features such as private verbs and THAT deletion. These features were associated with spoken registers by Biber (1988). As seen in Table 4-2, the distribution of the z-scores is within the range of +/- 1, with some features marking greater z-scores (highlighted in Table 4-2; e.g., general emphatics Course 2, causative subordination, from Courses 1 to 5). Among some relatively salient features, emphatics and possibility modals exhibit a nonlinear increase across Courses while sentence relatives show a steady increase. On the other hand, causative subordinations tend to nonlinearly decrease. Twelve features increase when the other ten features decrease from Course 1 to Course 3. Considering that these features are associated with spoken registers, this overall increase is not expected.

As for the negative major features in Table 4-2, nouns, word length, prepositions, type-token ratio, and attributive adjectives were associated with an informational type of discourse in Dimension 1 (Biber, 1988). The rise of these features, therefore, indicates that the language use gradually becomes more informationally dense. However, the overall variations of these features' z-scores across Courses are minimal. The three features (word length, prepositions, and type/token ratio) exhibit a nonlinear increase across Courses 1 to 3 while nouns and adjectives decrease from Course 1 to Course 3. As these features are associated with written discourses in Biber (1988), this mixed trend of variations in the z-scores across Courses is questionable. It was suspected that the trajectories of these features based on the pre-established dimensions did not clearly show the communicative functions used in the PELIC longitudinal corpus. Therefore, I conducted a new factor analysis to identify the functional dimensions unique to the corpus as further methodological consideration. The analysis based on the new factor analysis is detailed in Section 4.5.

The saliency of Dimension 4 in the PELIC corpus warrants further examination. Table 4-3 shows the z-scores of the major features loaded on Dimension 4.

Table 4-3. Z-scores of linguistic features in Dimension 4 in the PELIC sub-corpora (ordered by the factor loadings in Biber (1988, p.103))

D4 features	course1	course2	course3	course4	course5

infinitives	1.24	1.58	1.40	1.86	3.20
prediction modals	-0.12	0.28	0.32	0.07	1.67
suasive verbs	0.14	0.10	0.17	-0.02	0.52
conditional subordination	0.50	0.61	0.58	1.25	3.39
necessity modals	0.64	0.74	0.38	2.22	2.46
split auxiliaries	-1.54	-1.62	-1.72	-1.55	-2.03

Dimension 4 has positively loaded major features only, lacking negative major features in Biber (1988). For the positively loaded major features, Table 4-3 shows that infinitives nonlinearly increase while split auxiliaries steadily decrease. Overall, four features increase while two features decrease. However, the trends across Courses of these features are nonlinear and the slope is small.

4.3.2 Interpretation of Dimensions

Based on the z-scores of normalised frequency of the linguistic features described earlier, MAT provides visualisation of data regarding each text's functional dimensions and text types. This subsection discusses the interpretation of Dimensions found in the PELIC sub-corpora, focusing on Dimension 1 (involved vs. informational production), as it explains the most variance in a factor analysis, it is the most informative of the dimensions identified from the corpus.

Table 4-4 describes the scores for Dimension 1 in the five PELIC sub-corpora. First, MAT calculates the Dimension scores by summing the z-scores of the variables that are loaded on each dimension in Biber (1988, p.77). MAT then classifies each text according to its closer text type as proposed by Biber (1989), as well as a representative text type for each sub-corpora, as shown in Table 4-4, among the text types introduced in Chapter 2 (refer to Table 2-1 in Section 2.2.1.1 for the text typology by Biber, 1989).

Table 4-4. Dimension 1 scores for five PELIC sub-corpora, MICUSP and academic prose (Biber, 1988)

DIMENSION 1		
	Dimension1	Closest Text Type
Course 1	3.68	Involved persuasion
Course 2	5.41	Involved persuasion
Course 3	0.83	Involved persuasion
Course 4	-0.87	Involved persuasion
Course 5	-7.27	Scientific exposition
MICUSP[1)]	-12.62	Scientific exposition
Academic prose[2)]	-14.9	Scientific exposition/ Learned exposition

1) The corpus of American university students' written essays

2) The texts categorised as the genre of academic prose in the LOB corpus (adapted from Biber, 1988, p.123)

As seen in Table 4-4, Dimension 1 shows a trend of decrease except for Course 2. This implies that the texts gradually become more informationally dense across five Courses. The score of the MICUSP corpus is lower than any of the Courses in the PELIC corpus, and the lowest score is found in academic prose. Potentially, the task prompt (question types assigned as a topic for writing) may play a role. In other words, the topic and task purpose given to each text are likely to dictate the linguistic choice. interestingly, an outlier is noted in Course 2 (5.41 in Table 4-4) in Dimension 1. As it is a pilot analysis, this nonlinearity has not been further examined. However, it has been observed in the final MDA study (Section 4.5, where a closer inspection of this phenomenon is presented.

Table 4-4 shows that among the five sub-corpora, 'involved persuasion' is the closest text type in Courses 1-3. This type of text is non-narrative and non-abstract in style and argumentative or persuasive in their primary purpose, typically associated with Dimension 4 (Biber, 1989, p.35). Dimension 4, labelled as 'Overt Expression of Persuasion' (Biber, 1988), consists of major linguistic features including infinitives, prediction modals, suasive verbs, conditional subordination, necessity modals, and possibility modals. Using these linguistic features, this dimension measures the extent to which persuasion is openly displayed, either through explicit expression of the speaker's perspective or through evaluating the persuasiveness or probability of an event presented to convince the listener (Biber, 1988, p.111). While Dimension 1 is associated with a communicative function informing the spoken vs written distinction, Dimension 4 marks a persuasive communicative function, which is one of the stance functions.

Biber (1989) describes the characteristics of the texts clustered to the 8th group identified from the Lancaster-Oslo-Bergen and the London-Lund Corpora as follows.

> *The majority of these texts are spoken: some are interactional and informational, such as the interviews and the telephone conversation between disparates; others are informational monologues, such as the spontaneous and prepared speeches. The remaining texts are written, informational texts, such as popular lore, professional letters, religion, humor, and editorials. Overall, the linguistic characterization of these texts is primarily persuasive and secondarily involved, while the texts themselves are primarily argumentative or persuasive in purpose.*
>
> Biber (1989, pp.35-36)

As summarized by Biber (1989) above, the genres suggested for text type 8 include genres such as monologues, which are particularly meant to be persuasive in their purposes. Due to the combination of the characteristics of the spoken mode (monologues) and the communicative purpose (persuasion), the text type exhibits an interactive and involved nature, which is more typical of spoken communications. This point is revisited with a qualitative inspection in Section 4.5, where the identified functional dimensions in the PELIC texts are mainly related to spoken registers.

This text type is frequent in the spoken genre, and all the PELIC sub-corpora showed positive dimension 4 score except for Course 5. The influence of task prompts may potentially impact this consistently identified text type in the PELIC corpus, but this aspect requires further qualitative inspection. This outlier of the Course 5 score resonates well with its closest text type, 'Scientific exposition', as seen in Figure 4-7. Notably, Group 5 consists of texts produced by only one learner; it is also likely that the learner variable contributed to such an outstanding result to a certain degree.

Figures 4-1 ~ 4-6 show the average scores of Dimensions 1 to 6 (e.g., Dimension 1's functional dimension is involved vs. information production) for the 1,115 whole PELIC longitudinal texts (marked as 'MAT_texts' in Figures 4-1 to 4.6) analysed in this study. The text types in the plot are identified based on the text typology introduced in Table 2-1 Section 2.2), using Euclidean distance analysis.

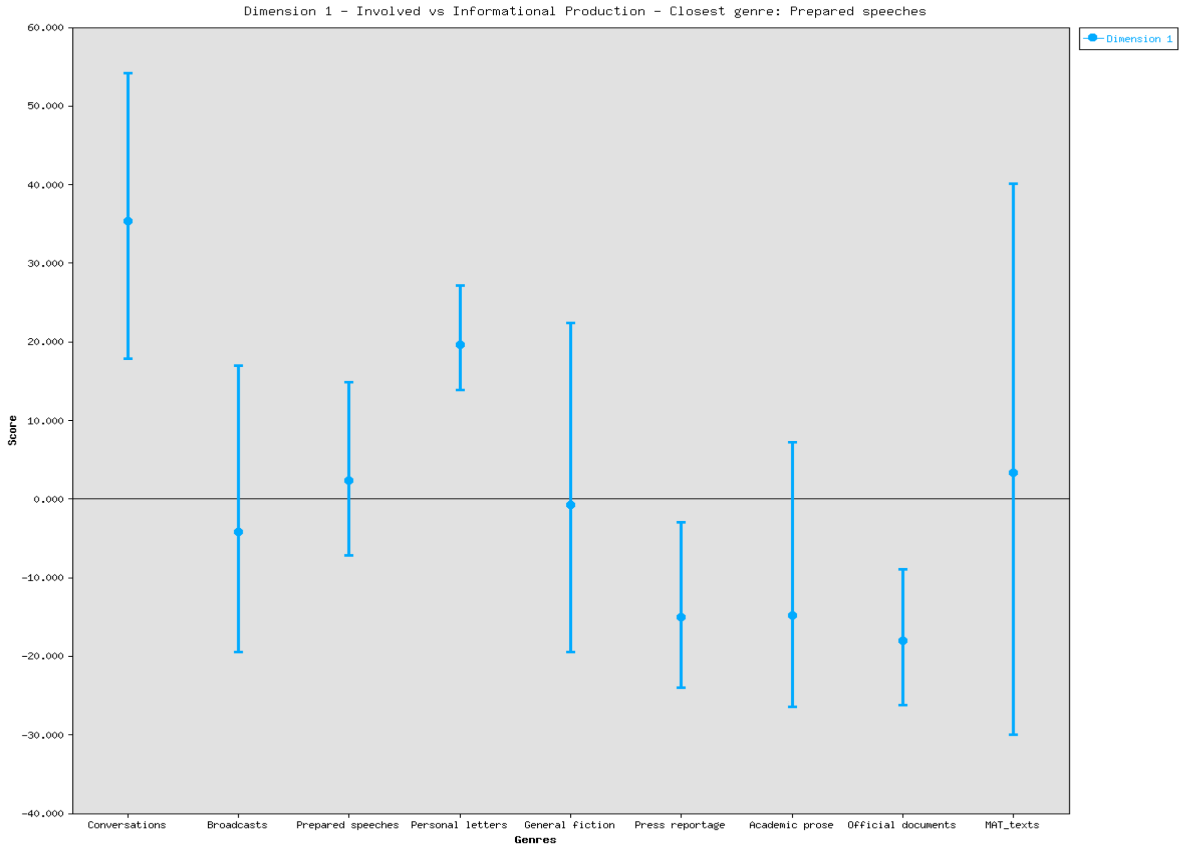

Figure 4-1. Average scores of Dimension 1 (Biber, 1988) of the PELIC longitudinal corpus

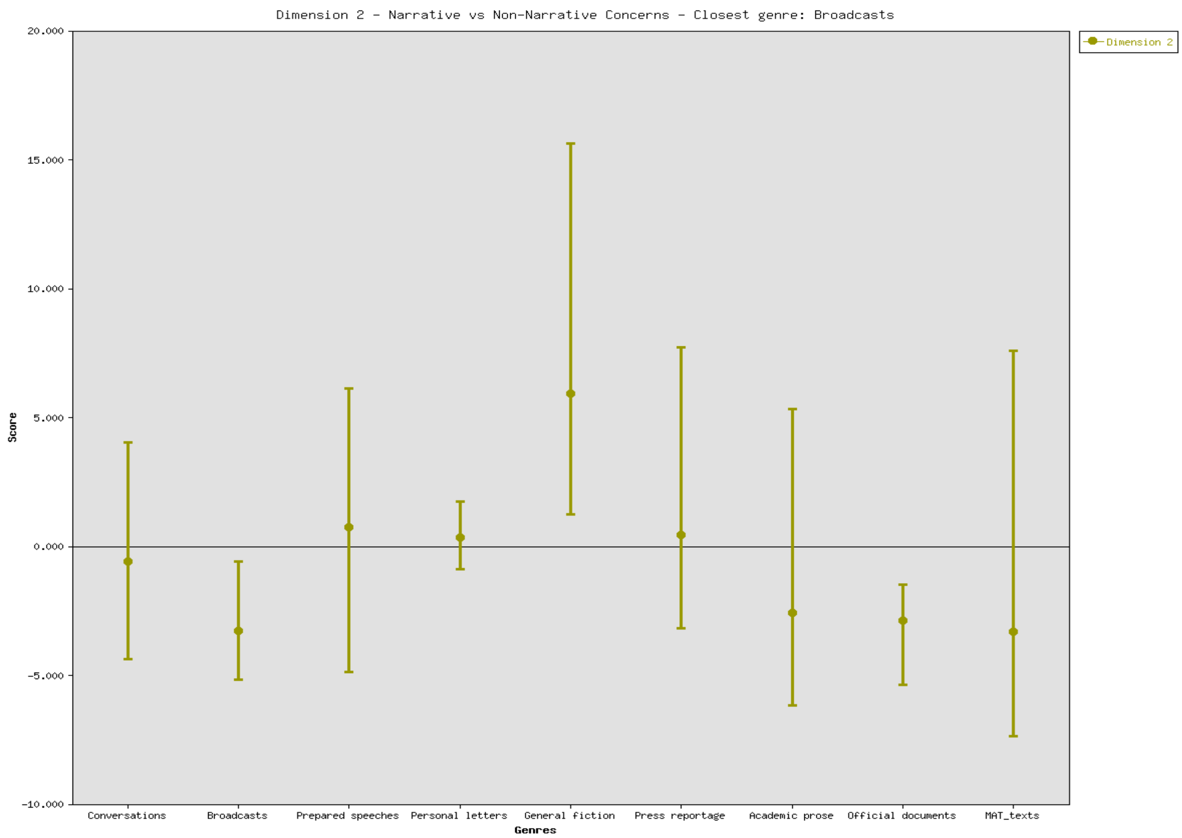

Figure 4-2. Average scores of Dimension 2 (Biber, 1988) of the PELIC longitudinal corpus

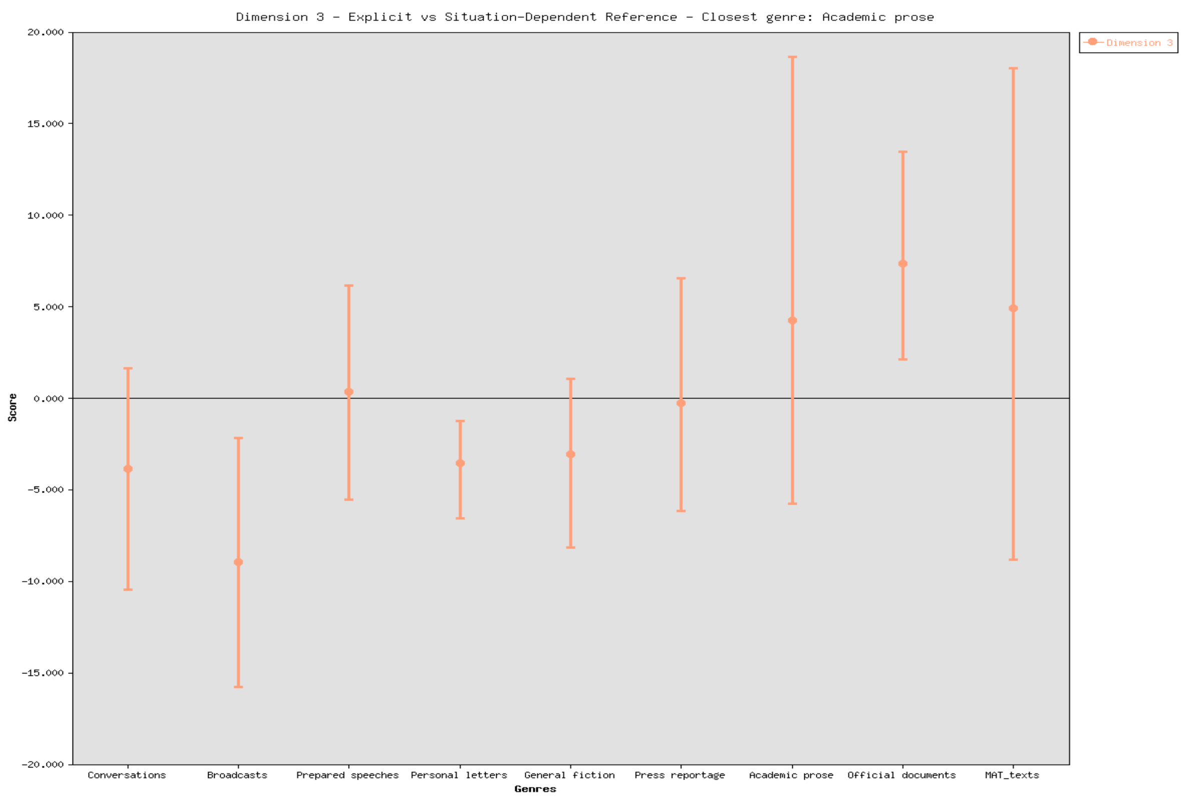

Figure 4-3. Average scores of Dimension 3 (Biber, 1988) of the PELIC longitudinal corpus

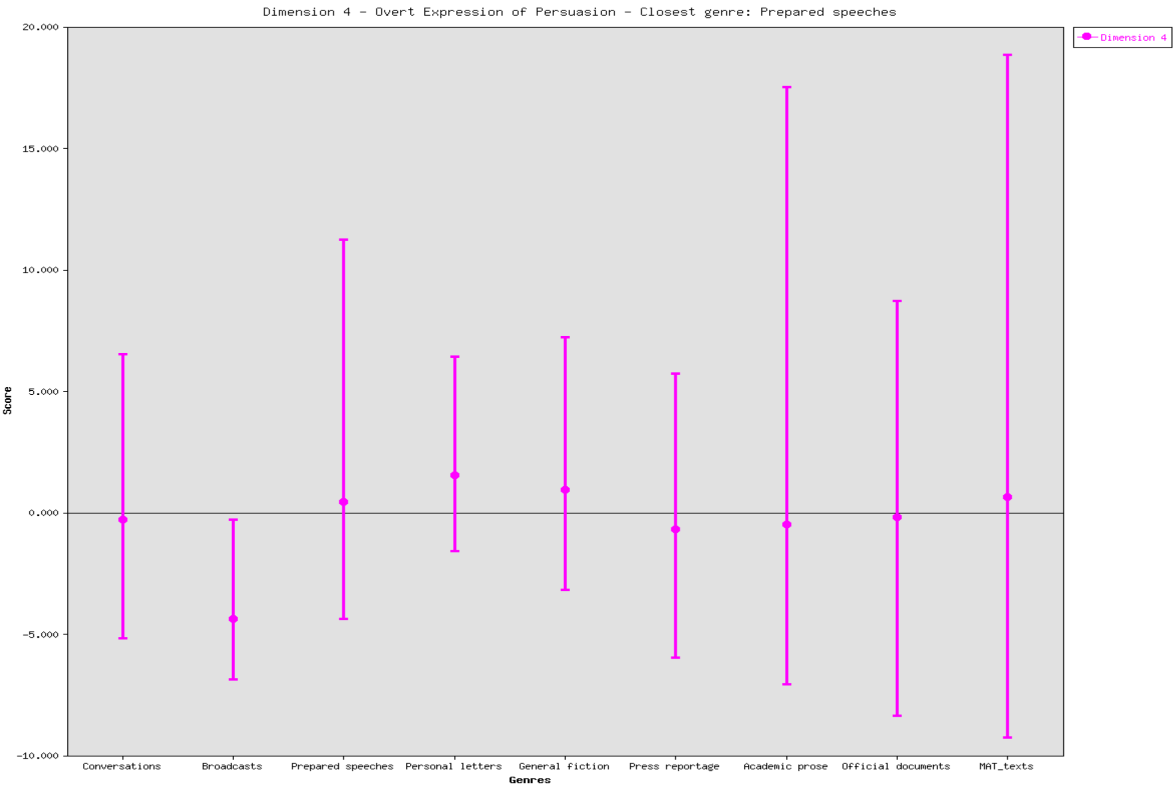

Figure 4-4. Average scores of Dimension 4 (Biber, 1988) of the PELIC longitudinal corpus

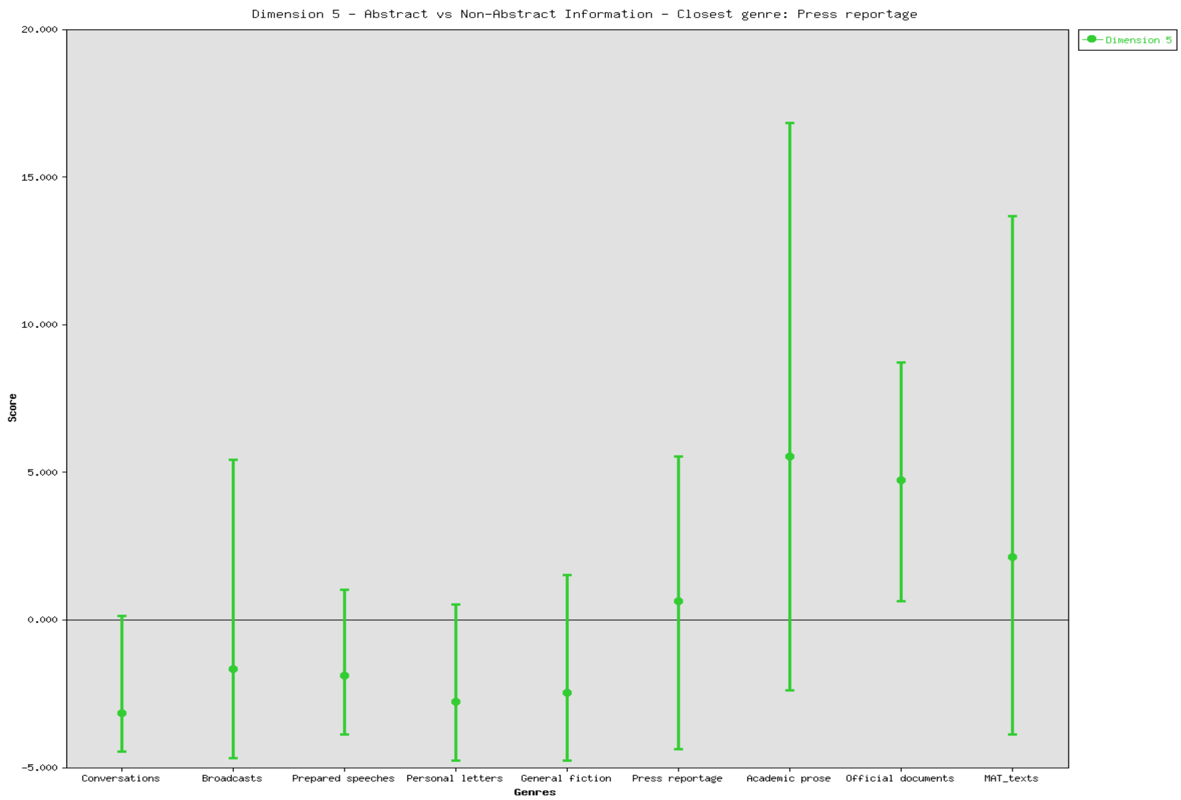

Figure 4-5. Average scores of Dimension 5 (Biber, 1988) of the PELIC longitudinal corpus

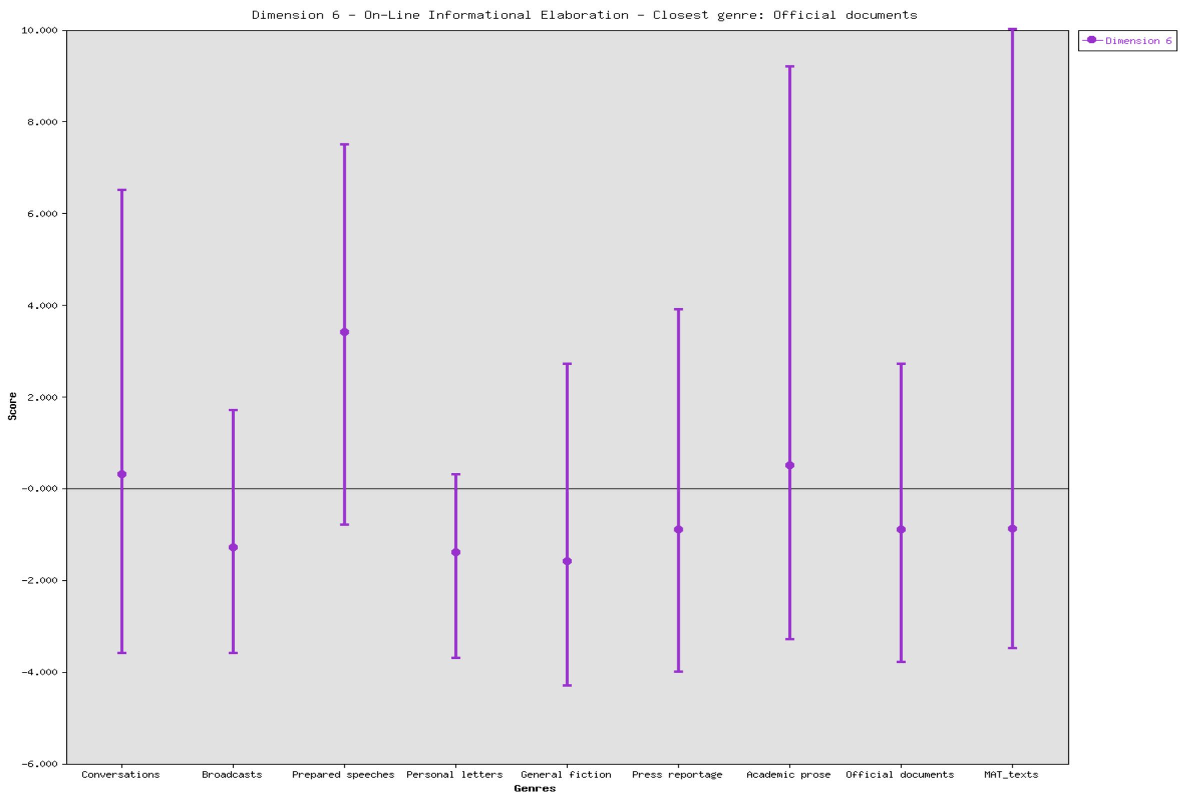

Figure 4-6. Average scores of Dimension 6 (Biber, 1988) of the PELIC longitudinal corpus

The Dimension scores for the five sub-corpora in Figures 4-1 ~4-6 are very widely dispersed, nearly reaching both poles of the Y-axis. These wide dispersions indicate that the established frame of text type does not inform much about the typical characteristics of the L2 student written corpus.

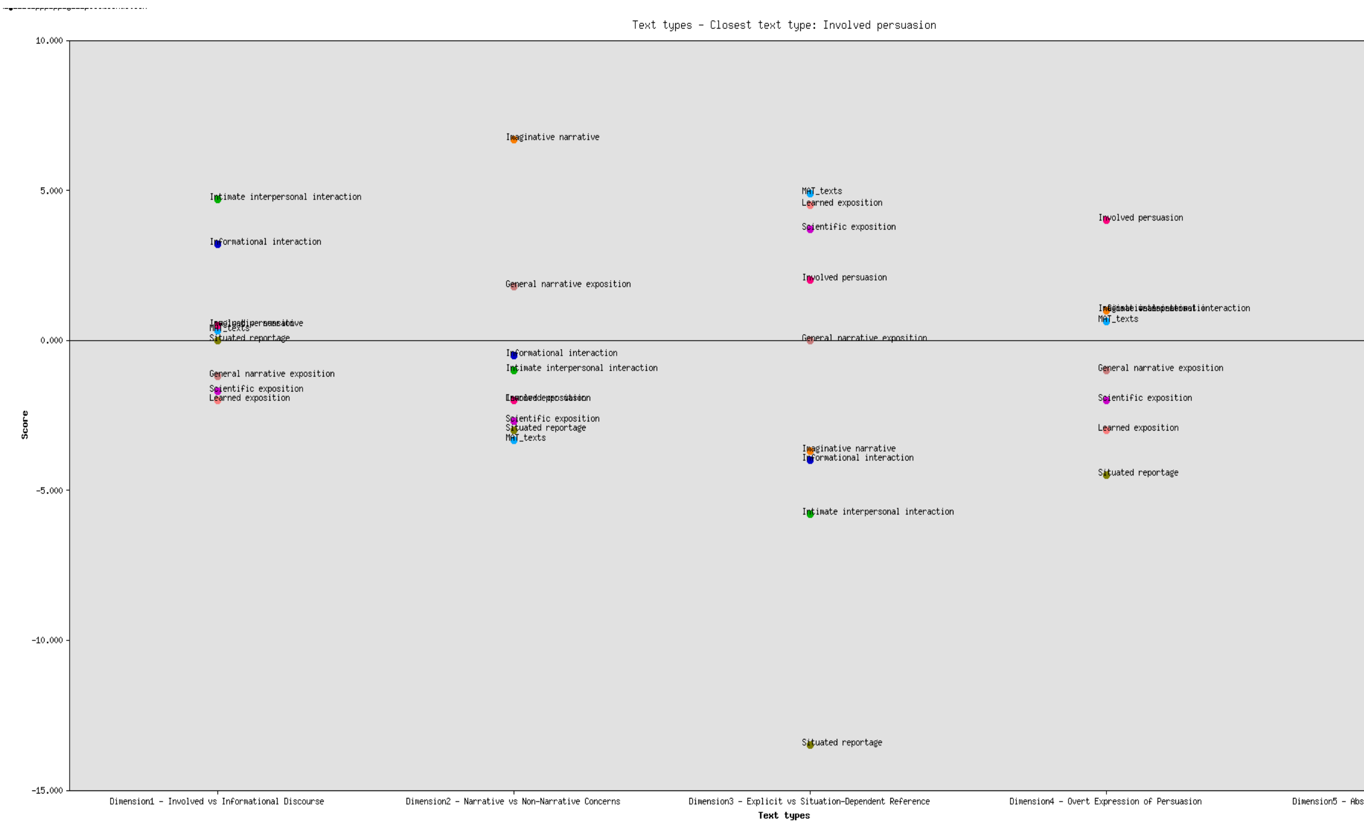

Figure 4-7. The closest text type (Biber, 1989) of the PELIC longitudinal corpus

Figure 4-7 shows the closest text type of the longitudinal PELIC texts identified based on the text typology (Biber, 1989), which is also processed by MAT. As an overall characteristic of the PELIC texts, ‘involved persuasion’ has been identified as the closest text type. However, the scores of Dimension 4 of the PELIC texts (marked as ‘MAT-texts’ in Figure 4-7) do not exhibit a clear tendency toward the involved persuasion text type. As this typology is developed based on the Brown corpus consisting 23 written and spoken genres, the interpretation based on this should be used as only a general background information for such a learner written register as the PELIC corpus.

In summary, despite such inconsistent results regarding dimensions, Dimension 1 in each of the five groups of texts shows the most interpretable results: it indicates that the texts are overall spoken register (‘prepared speeches’) except for the fifth group but become gradually more densely packed with information.

It should be noted that the interpretation of the functional dimensions is based on the co-occurrences of these linguistic features in performing a shared communicative function in Biber (1988). As I have discussed so far, however, many of the frequency data of these features in the L2 writing corpus do not co-occur the way they were in Biber (1988). For example, present tense and attributive adjectives were associated with a non-narrative type of discourse in Dimension 2 with significant negative loadings in Biber (1988). However, while attributive adjectives are relatively more frequent, present tenses are less so when compared to the standardised frequency scores in academic prose (Biber, 1988). The finding in Biber (1988) showed these two features were significantly related to performing a communicative function in academic prose. However, their patterns in the PELIC longitudinal corpus are not patterned, making it hard to interpret their functional behaviour based on the established dimensions in Biber (1988).

Similarly, conjunct is a major feature loaded on dimension 5 in Biber (1988, p.103) with passive constructions (agentless passives and by-passives). Conjunct is a consistently overused feature throughout the PELIC texts, with notably high z-scores (4.93 on average in the PELIC longitudinal corpus). However, passives show negative scores, indicating marked underuse compared to academic prose in Biber (1988). Therefore, the interpretation of co-occurring conjuncts and passives as a more abstract style (Biber, 1988) does not apply to the patterns of these features in L2 writing.

In conclusion, it appears that the linguistic features constituting the altered dimension in Biber's (1988) study are not consistently correlated in PELIC. This inconsistency limits the extent to which the linguistic features consisting of each function dimension found by Biber (1988) are interpreted for their shared communicative functions in the PELIC corpus. Therefore, further investigation based on a new factor analysis may provide information to interpret the unique variation of these linguistic features in the PELIC texts, for example, to understand why conjuncts are overused while passives are less utilized.

4.3.3 Conclusions

The study result generally confirms the previous findings regarding L2 academic writing more like spoken register, as discussed in Chapter 2 (e.g., Kim & Nam, 2019). In addition, there is evidence of the texts becoming more informationally dense, exhibiting more written register characteristics.

However, the limitation of this analysis is that many of the results are not genuinely interpretable due to the difference between the corpus used in Biber (1988) and the PELIC corpus. The dimensions found in Biber (1988) were representative of a more general corpus while the longitudinal PELIC corpus represents a more unitary L2 written register, for which the sub-corpora consists of subtle variation across relatively short time points. Arguably, the task types and writer-related variables such as proficiency and L1 backgrounds should influence the variation over time. One unexpected observation involves the non-linear patterns for some linguistic feature use and average dimension scores regarding their co-occurrences. These patterns may be better explained if a new factor analysis is conducted because that may provide the functional dimensions based on the linguistic variations, which are unique to this written data. For that reason, the analysis based on a new factor analysis was conducted, which is discussed in Section 4.5.

4.4 Pilot Study 2: Exploring Functional Dimensions using Syntactic Features Incorporating Grammatical Form-syntactic Function Distinction

The two more pilot MD analyses discussed in this section are differentiated from the first pilot analysis discussed in Section 4.3 in that these analyses included the exploratory factor analyses while the first analysis did not. These two analyses are further differentiated regarding the linguistic variables used for the factor analyses. I do not discuss the second pilot analysis (MDA) in detail here. This is because the second analysis relies on variables measured under the assumption of elaboration, which does not contribute to the overall synthesis of the studies presented in this chapter. However, the linguistic variables used and the implications of the analysis results are briefly discussed. The second pilot analysis used the frequency data of three types of linguistic variables: 1) seven semantic categories of verbs (Bibet et al., 2021) 2) 152 syntactic complexity measures and 3) 208 syntactic sophistication measures tagged by TAASSC (Kyle, 2016, ver.1.3.8). Through the adequacy check of the initial factor analyses, a total of 46 features were used for the factor analysis of this second pilot analysis. The general findings from this analysis correspond to the pre-established expectations; phrasal indices get slightly more frequent over time. However, as the co-occurrences of these linguistic variables are not easily interpretable in terms of their communicative functions, and therefore no further investigation using these linguistic variables is pursued. A subsequent pilot analysis tried different types of measures, as reported in the following section.

4.4.1 Method

This pilot analysis used a different set of SC features, the 34 syntactic features parsed and counted by the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle 2016), which relies on a Stanford dependency parser (De Marneffe et al., 2014) to identify syntactic structures. For the factor extraction method, both Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA) were tried. PCA is arguably the most commonly used approach for dimension reduction and visualization while EFA is grounded to identify theoretically interpretable latent variables. In addition, MICUSP was used as a reference corpus in interpreting developmental patterns across the L2 writing corpus grouped by time points.

After initial rounds of factor analyses for checking correlations between measures (normalised frequency of each feature) and factorability, only the 28 lexico-grammatical features were kept for the final factor analysis using the *FactoMineR* module in the R environment. There are a few features with high correlation with other features, but they are not removed in this analysis. Only features with very low correlation, very infrequent features, or features with very low communality and factorability were removed. This decision is made considering that syntactically co-occurring features, therefore having high correlations, do not theoretically indicate the same linguistic structure. Sometimes the linguistic structures attract another feature. For example, a passive construction is likely to invite ‘by agent’ to follow. However, they do not represent the same linguistic structure, even though they co-occur for a shared communicative function. However, it is still statistically multi-collinearity, which potentially causes the analytical imprecision. This aspect is improved in the final analysis (Section 4.5), where the justifications of linguistic variables selection and analysis procedures are discussed.

The scree plot in Figure 4-8 represents the proportions of variances explained by each factor and component, which provides information to decide an optimal number of factors (for principal axis factoring) and components (for principal component method).

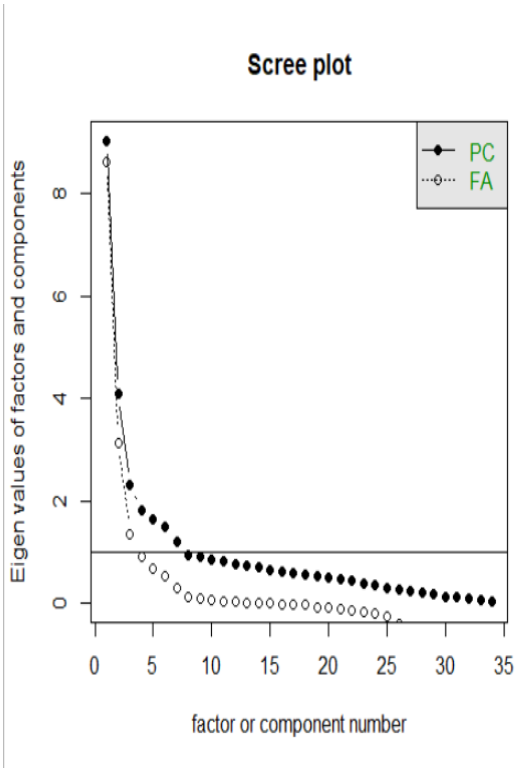

Figure 4-8. Eigenvalues of factors and components for 34 syntactic features

PCA was chosen for further exploration due to its potential for interpretability of the features loaded onto the principal components. PCA prioritises capturing the most significant variations in the data, which can make interpreting the components more straightforward (Jolliffe & Cadima, 2016). On the other hand, FA was not chosen for dimensionality reduction, in this specific case, as further investigation was needed to determine its suitability regarding its factorial adequacy, which refers to the strength with which features load onto factors. The total eigenvalue percentage of cumulative variance is 45.8 for the three components. The next section discusses the final factor analysis results.

4.4.2 Results and Discussions

Table 4-5 shows the features significantly loaded onto each dimension (±0.35). After inspecting some of the texts with high occurrences of major linguistic features in each dimension, it seemed that Dimension 1 is heavily loaded with prepositional features. Biber et al. (2021a) suggest that the variety of head prepositions other than ‘of’ is a marker of development. While it is not clear as to the reason for this result, constructions headed by the preposition 'of' are the only prepositional construction separately loaded on Dimension 2 while all the other features are loaded on Dimension 1. This will be referred back to in the next section of the final MDA.

Table 4-5. Major linguistic features (with loadings larger than ± 0.35) on each dimension (Pilot study)

<table>
<tr><td>Dimension 1</td><td>Dimension 2</td></tr>
<tr><td>positive (+) features:
Clausal prepositional complement (0.46) / discourse marker (0.65) / indirect object (0.61) / passive clausal subject (0.97) / prepositional modifiers: across (0.97), after (0.49), against (0.85), among (0.84), around (0.62), into (0.65), like (0.47), over (0.57), through (0.68), toward (0.92), under (0.91)
negative (-) features: Verbal modifiers (-0.35)</td><td rowspan="3">positive (+) features:
Adverbial clause (0.59) / auxiliary verb (0.52) / clausal complement (0.39) / direct object (0.39) / nominal subject (0.56) / open clausal complement (0.41), subordinating conjunction (0.5)

negative (-) features:
prepositional modifier: of (-0.46) / conjunction (-0.65) / determiners (-0.38) / conjunction: and (-0.68)</td></tr>
<tr><td>Dimension 3 (positive features only)</td></tr>
<tr><td>positive (+) features:
agent (0.44) / passive auxiliary verb (0.94) / passive nominal subject (0.93)</td></tr>
</table>

Many of the features with high correlations are clausal indices, indicating potential multi-collinearity issues. On the other hand, most phrasal indices were removed because they did not significantly load on any factors. This loss is rather disappointing in that the previous findings suggest that phrasal indices are significant markers of quality writing development (e.g., Biber et al., 2014; Kyle & Crossley, 2018; Larsson & Kaatari, 2020). As noted in Chapter 3, the data was not cleaned for more accurate grammatical tagging, which limits the accurate representation of the linguistic variations for analysis.

Aside from this processing limitation, a possible reason for this result is the influence of writer-specific variables. What has been considered the most plausible explanation for this is that the writers' English proficiency levels of the general population of this dataset are not advanced enough to indicate the growth of phrasal complexity. Considering that students started with relatively limited general English proficiency (average entry Level 3 as shown in Figure 3-1) it is plausible that the exposure to the tasks of conventional literary writing may predict language use of this sort even more strongly than the proficiency levels assigned during their stay at the programme. One possible explanation is that the writers are still developing the ability to produce elaborate sentences including what has been considered more spoken registers. Another possibility is tagging inaccuracy or the linguistic measures not being comprehensive enough to include interesting markers of syntactic complexity. These possibilities are discussed in Chapter 3 (Section 3.3) and the next section (4.5), respectively.

Descriptions of the major features loaded on each dimension and their shared communicative functions are discussed below to provide the basis for the functional interpretation of the textual dimensions regarding the three linguistic dimensions each. It should be noted that, unlike the conventional MD analyses that label factors functionally, this study labelled each dimension based on the formal linguistic properties - syntactic characteristics. This labelling is due to the difficulty of identifying the functional dimensions based on the co-occurrences of the major linguistic features associated with each dimension. The following paragraphs present the syntactic variation associated with each dimension.

Dimension 1: Phrasal vs Clausal complementation and modification

I have focused on dimension 1 in discussing the previous MD analyses because the first factor in factor analysis explains most variance of the corpus, and for the same reason, I pay most of the discussion to Dimension 1 for this study. Dimension 1 is heavily loaded with prepositional features. Notably, construction headed by the preposition 'of' is the only exception, loaded on Dimension 2. The only negatively loaded feature is the 'verbal modifier', a phrasal index. Therefore, it seemed that the preference to use either prepositional or verbal modification and complementation distinguished the texts in Dimension 1, which led to tentatively labelling Dimension 1 as “Phrasal vs Clausal complementation and modification". As noted in the Method section 4.5.3., the prepositional indices were collapsed and not distinguished at a phrasal or clausal level, which required them to be examined in context. This qualitative examination did not provide a meaningful difference between these two syntactic functional forms. Rather, it was mostly the case that texts with high occurrences of either of these two also used the other more often. That is, the texts tended to use both substantively without making a noticeable difference in distinguishing these two constructions.

One type of passive construction loaded significantly onto D1 is the 'passive clausal subject'. Another feature related to passive constructions, 'agent', was also significantly loaded onto D1 (=-0.39), but it was assigned to D3, where its loading was more significant (=0.44). Both these passive constructions and prepositional indices are clausal indices, while the prepositional modifiers in either clausal modification (e.g., occurring after a verb to complement the verb) or phrasal modifications (e.g., within noun phrases). Overall, the features loaded on this dimension are used at the clausal level.

Dimension 2: Phrasal versus clausal complexity

The positively loaded features are clausal indices, including the adverbial clause, and the 'open clausal complement', while negatively loaded features are phrasal indices such as conjunction '*and*' and determiners. Notably, 'of-phrase' was distinctively loaded onto D2, while the other prepositional phrases mainly were Dimension 1. Biber et al. (2011) hypothesised that *of* phrases as noun postmodifiers precede other PPs as postmodifiers at developmental stages. While this distinction between prepositional constructions has been hypothesised to be a marker of writing development, whether these separate loadings on different dimensions should be interpreted in relation to the hypothesis is not further pursued. Instead, I further analyse prepositional constructions in Chapter 6 (Sections 5.2 and 5.3) to discuss potential explanations about how different prepositional constructions cluster to perform a shared communicative function, where in-text examples of prepositions are presented.

<u>Dimension 3: Impersonal description</u>

The three significantly loaded features onto D3 are passive auxiliary verb, passive nominal subject and agent. Therefore, D3 is tentatively labelled as 'passive constructions.'

Figures 4-9 to 4-11 visually depict the distribution of the scores for each of the three Dimensions, including the interquartile range, median and presence of outliers, across five Courses. The inclusion of error bars extending from the boxplot's mean provides a 95% confidence interval, indicating the range within which we are 95% confident that the true population mean of each dimension lies. These figures include the mean values of each dimension's factor scores of the MICUSP corpus, as a reference corpus of the target student register. As the dimension scores are computed based on standardised frequencies of the features both the PELIC and MICUSP corpora were processed at the same time for the standardisation.

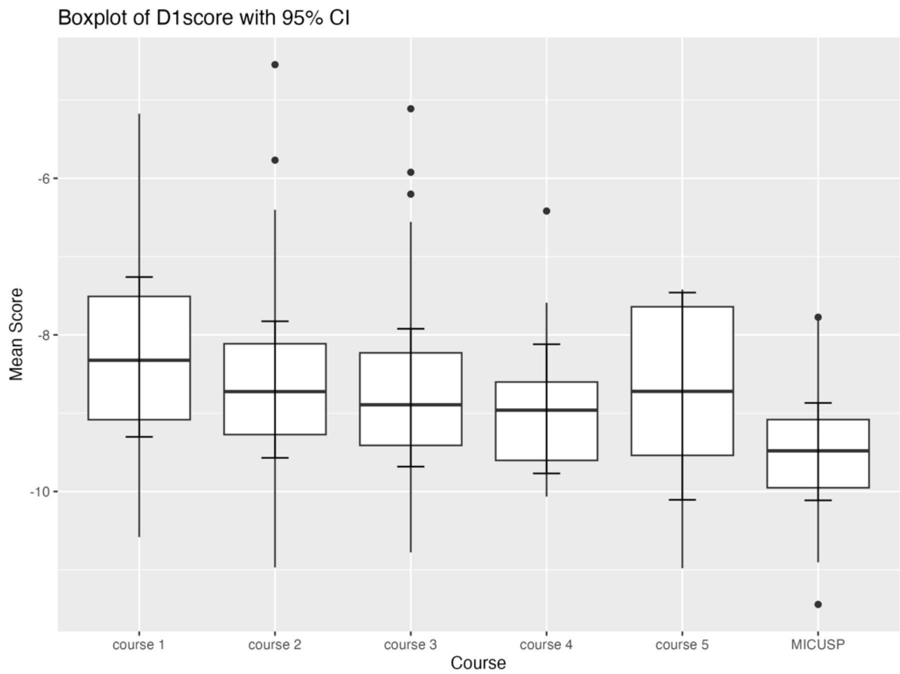

Figure 4-9. **Dimension** 1 **mean scores for five PELIC sub-corpora and MICUSP corpus**

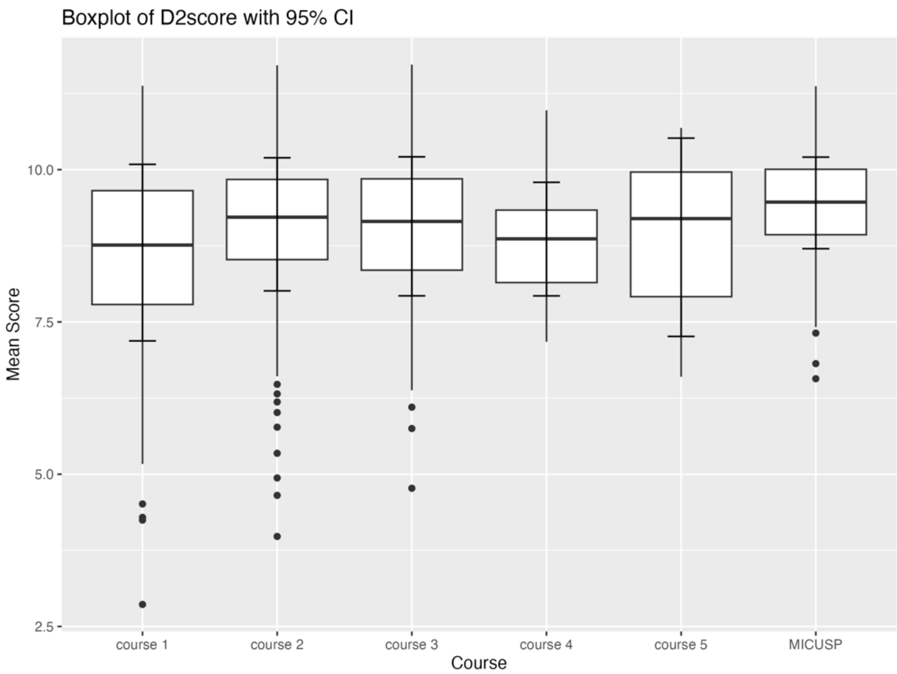

Figure 4-10. Dimension 2 mean scores for five PELIC sub-corpora and MICUSP corpus

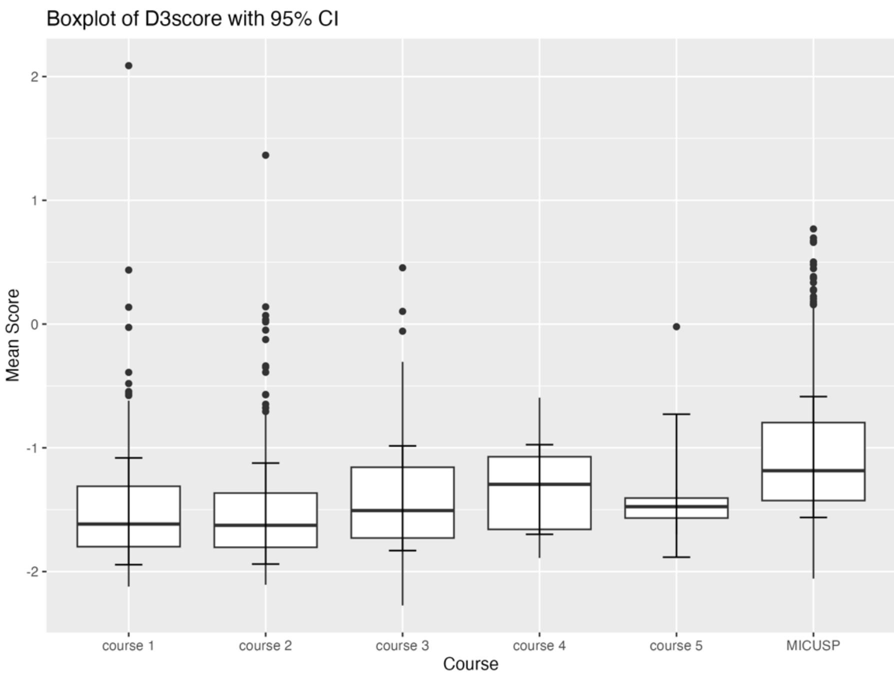

Figure 4-11. Dimension 3 mean scores for five PELIC sub-corpora and MICUSP corpus

These dimensions were identified by examining the linguistic features co-occurring in the PELIC corpus only, and the dimension scores of the MICUSP were calculated afterwards to compare to the PELIC corpus. The dimension scores in Figures 4-9 to 4-11 show nonlinear increases across the five sub-corpora, but the overall direction of change conforms to the target discourse represented by the dimension score of MICUSP. The use of passive voice is not theoretically complex in that it does not add more linguistic parts, but it is associated with more advanced developmental stages and advanced academic writing (Biber et al., 2021, p.478). Therefore it is worth further investigation in future analysis.

The negative trend of the Dimension 1 scores indicates the shift of weight from positively loading features to negatively loading features over time (Courses 1-5), which is most noticeable in the reference data (MICUSP). However, this is of relative importance and does not reflect the absolute frequency decrease.

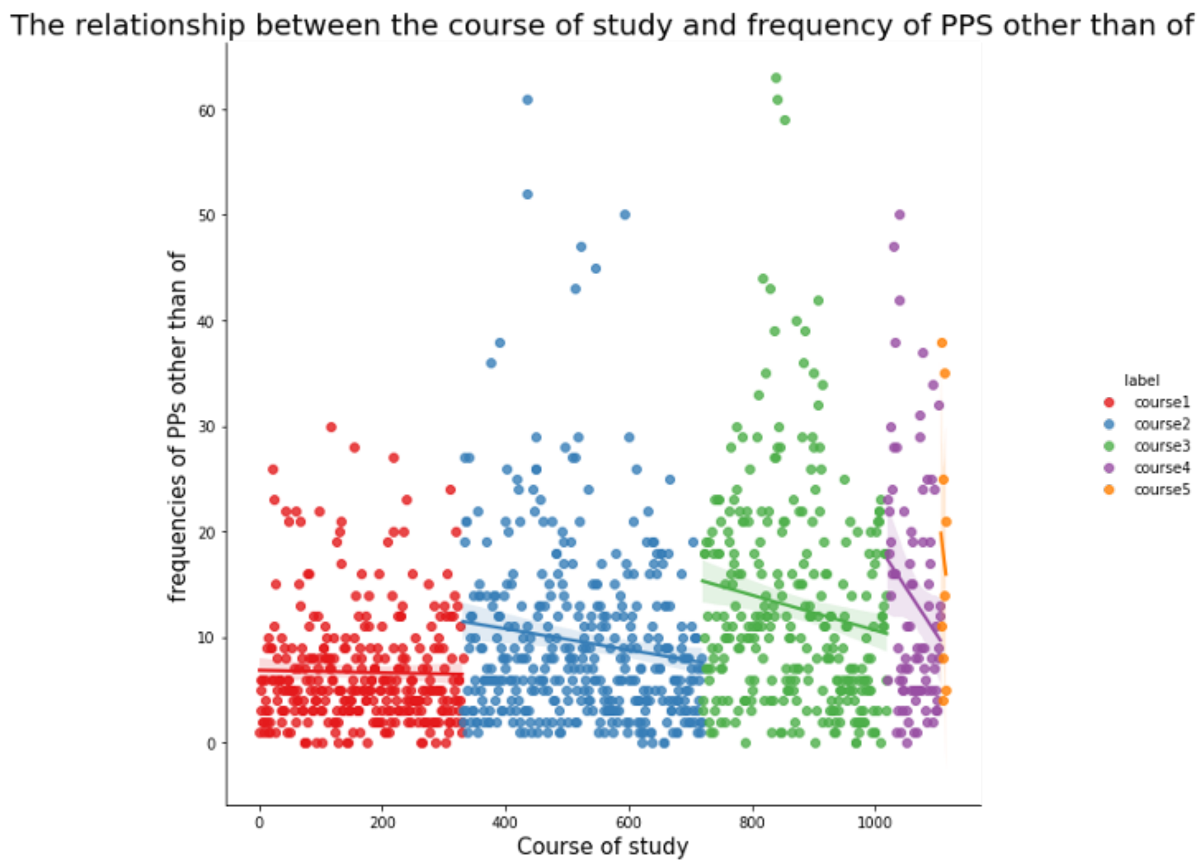

Figure 4-12. Absolute frequency counts of PPs other than *of* across five Courses

Figure 4-12 plots a scatterplot of frequency counts of prepositions loaded onto Dimension 1 in the five PELIC sub-corpora: prepositional modifiers (a clausal index) and verbal modifiers (a phrasal index). The frequencies of both indices show an increasing trend across the sub-corpora. The increase in these features may indicate an increase in fluency in using these features over time. However, as observed in the decrease in the Dimension 1 score, their relative decrease in weight among the major linguistic features invites further investigation into prepositional complementations and modifications in context, which will be presented in Pilot Study 2 in Chapter 5 (Section 5.3).

Finally, a brief examination of sample texts produced in the different courses by the same writers revealed a moderate increase in the quantity and variety of both clausal and phrasal indices, while some inaccurate uses of the syntactic forms were detected in writings produced at higher proficiency levels. Consider the following two sample texts (Text 1 and 2) with instances of the major linguistic features in Dimension 1.

Text id _ 23684, Course 1, Level 4

(The clausal features in italicised; the phrasal features are in bold)

In recent years, the problem of obese teens in the U.S. has grown. The effects of obesity are the high risk of diabetes and heart disease. The causes of obesity are complex, and include genetic, behavioral and cultural factors. Basically, obesity occurs when a child eats more calories than the body needs.

A radical solution for overweight teens was reported by CBS News medical correspondent Dr. Jon LaPook. According to the report, a weight reduction **surgery has approved,** Dr. George Fielding, who has worked at New York University Medical Center, said that 1 in 15 American children were obese enough to need this operation, and the operation was **to use a silicon band** to tie *around the upper part of the stomach*, and this narrowed the stomach and caused patients to feel full. A 16-year-old girl Jodie Babich, whose weight was 240 pounds, opted to have the operation. *After three weeks*, she has lost 21 pounds. "It's the only treatment that works for the morbidly obese," Dr. George Fielding said " not only does it treat their fat, but (it) cures with the diseases that goes with the fat." More and more obese children will get the benefit from the operation.

Text 2: id_37238, Course 5, Level 5

(The clausal features in italicised; the phrasal features are in bold)

Children begin to try to impress their opposite sex when they are *into adolescence*. During this period, boys and girls are easy to distract their mind from their study. Should boys and girls be educated separately since middle school? Will the same-sex schooling create a more relaxed social atmosphere and prevent boys from dominating classroom discussions and activities? Absolutely not, for the same-sex schools, the students may be free from distraction of the other sex; however, the same-sex schooling would harm the socializing skill of its students. *Unlike the co-educational school*, which keeps gender diversity in school, the single-sex school makes the students unable to meet the people of different gender, and thereby, decrease the ability **to communicate with** the people of opposite gender might not know about how to talk with the people of the other gender for a long time even *after graduated*. Moreover, the students from single sex school may have trouble **to get friends** of opposite sex, which not only limited their broad view of the world, but also their future such as marriage. Due to reasons above, I am bias *against the single sex schooling.*

In texts 23684 and 37238, the correlations between each text and the student's level do not seem associated with the syntactic variation. Therefore, considering that writers generally advanced to higher levels over time, achievement of communicative purposes could have been considered more important in a level assignment in the language programme. The highlighted parts in the excerpts above serve as examples for this point.

4.4.3 Conclusions

The results suggest the need to revise the linguistic features for the final MD analysis - which not only reflect the structural form - syntactic function distinction but also are informed by the previous literature discussing linguistic features characteristic of academic writing development. The final MD analysis used the syntactic feature based on 67 grammatical features via MAT, which replicates the features in Biber (1988) with some modifications made to them. The next section presents the final MD analysis, providing the rationales for choosing these 67 features provided by MAT and making modifications to them.

No close examination has been done of the key grammatical patterns, but a broad pattern emerged from this quantitative analysis reported in this section. First, the features for spoken vs written distinction proved to be a meaningful marker of development.

The preliminary findings reported in this pilot analysis show evidence of spoken register in novice academic writers and gradual replacement with that of written registers as an indication of academic English development in L2 writing (e.g., Biber & Gray 2013; Kobayashi & Abe 2016; Kim & Nam 2019). More specifically, as their studies progressed, the learner texts were closely associated with phrasal complexity. However, the pattern is not linear across the five sub-corpora, and the phrasal indices loaded onto these dimensions are limited.

Finally, the distinction between 'of' and other prepositions associated with two separate dimensions is worth further investigation with PELIC's own emerging dimensions. These patterns, nonlinearity and prepositional distributions will be further analysed in the Final analysis in Section 4.6.

Informed by the analysis presented in Section 4.5, subsequent analysis has reinforced the indices of phrasal and clausal complexities to longitudinally trace the non-linearity along with these features. Further analysis is expected to provide more explicit trajectories that individual learners follow as they progress through different semesters.

However, the limitation of the analysis is that the communicative functions were not identified from the PELIC data; it was pre-established from the spoken and written registers analysed by Biber (1988). The second point is that the pilot analysis using MAT exhibited a pattern of nonlinearity in the linguistic features associated with Dimension 1.

4.5 Final Study: Exploring Functional Dimensions using Linguistic Complexity Features Drawing on a Usage-based Approach

Discovering new functional dimensions in a corpus can provide a deeper understanding of the intrinsic makeup of the register of interest (Nini, 2019, p.70). This final MD analysis aims to shed light on the L2 written data as a unique written corpus built within a university English learning context, identifying key linguistic complexity markers that inform development in their course of writing. The trials of different specific analysis methods and linguistic variables have provided insight into how to go about this endeavour suitable for the data and research question and also have helped reveal the nature of the L2 writing on its own. The linguistic features, analysis methods and interpretation of the analysis results discussed in this section are based on the pilot analyses reported in the previous sections.

Biber et al. (2011) propose that writers progress from using features common in conversation (finite clausal features) to those characteristic of informational writing (nonfinite and phrasal features within phrases) (pp. 48-49). The Biber et al. (2011) hypothesis is based on the assumption that conversation is acquired first, with no formal instruction, while academic writing requires explicit teaching (as discussed in the passage). However, Staples et al. (2023) note that we are not fully equipped with evidence for the similarity of the established stages of grammatical complexity development, outlined by Biber et al. (2011), to L1 and L2 writing. Staples et al. (2023) emphasize the critical need for a direct comparison of L1 and L2 writers (p. 49). They argue that we currently lack a clear understanding of whether both groups follow similar developmental trajectories. This comparison is particularly relevant because L1 and L2 writers encounter academic writing in vastly different educational contexts. L1 writers typically receive extensive exposure to academic writing throughout their schooling, while L2 writers may be introduced to it later and with varying intensity depending on their learning environment.

More importantly, L2 learners may not always follow this linear path. For example, their exposure to written language in their native language (L1) and the target language (L2) can influence their development. This means analysing how learners use grammar to achieve specific communicative goals in their writing, which may not always align with the stages proposed by Biber et al. (2011). Section 4.5 incorporates the phrasal vs clausal distinction into the linguistic features for the analysis, as suggested by Biber. At the same time, it also discusses the limitations of Biber's framework (1988) in analysing the L2 writing development, based on the nonlinearity of the trajectory of the developmeantal patterns in these linguistic features.

4.5.1 Linguistic variables

The linguistic feature analysed in this section is based on 67 grammatical features via MAT, which replicates the features in Biber (1988). This section presents the process of extracting and reducing these grammatical features for factor analysis in the study. Principled variable selection for MD analysis is critical because linguistic variables ultimately determine the validity of interpretation on statistical analysis (Egbert & Staples, 2019, p.130). The present study is based on the grammatical features identified in Biber (1988). In addition, to better analyze the functional aspects of writing related to stance (a writer's attitude or position), two changes were made. First, some adjustments were made to prepositional phrases to distinguish between their grammatical form (e.g., noun phrase following a preposition) and their syntactic function (e.g., acting as an adverbial modifier within a clause). Second, the original four verb categories were replaced with the seven semantic categories, based on the list of verbs by general frequency identified in Biber et al. (2021b). This change was made because stance is one of the key linguistic functions investigated in this section, and Biber et al. (2021b)'s verb classification system is particularly useful for stance analysis. Their system allows us to identify verbs that frequently co-occur in stance constructions, such as verbs used in controlling complement clauses (e.g., *ensure that*) or with infinitive phrases (e.g., *appear to*). The complete set of 74 linguistic features used in the study is shown in Appendix 3-1.

The previous literature (e.g., Gardner et al., 2019; Pan, 2018; Gray et al., 2019) proved the importance of considering situational and learner variables into account. The Pilot analyses also indicated that the longitudinal data do not clearly exhibit the trajectories of the linguistic variable, potentially due to the interaction with the situational and learner-related factors. Even though it is still valuable to trace chronological trajectories regardless of such factors, it is hard to generalise the results as the effect of time. Therefore, the final MDA that will be detailed here made use of a different dataset, unlike the previous reports. The dataset was selected so that such learner variable (L1 and L2) was minimised. The topic effect, the situational variable, was not possible to rule out because the range was too wide (about 400 question types as categorised and recorded by the data compilation team)

The 1,115 texts from the PELIC corpus were tagged with the 67 grammatical features using the Stanford tagger in the Multidimensional Analysis Tagger (MAT). Two types of linguistic variables were considered: grammatical features and semantic categories of verbs. There were some modifications to these 67 features considering 1) syntactic form-function, 2) functional interpretation and 3) accuracy of tagging. This distinction was made, in part, to revisit previous findings (e.g., Gray et al., 2019; Parkinson & Musgrave, 2014). Specifically, I wanted to see if separating grammatical form from syntactic function within constructions would reveal any meaningful differences in the way L2 writers develop these patterns over time.

First, the spaCy (Explosion AI, 2018) module (version 3.0) in Python was employed to analyse 152 lemmas representing seven semantic verb categories (activity, communication, mental, causative, occurrence, existence, and aspectual verb; Biber et al., 2021b, pp.366-369). The frequency data for semantic verbs were normalised per 100 words, following MAT's statistics. MAT's four semantic verb categories were replaced with these seven frequency counts to align with the communicative purposes identified by Biber et al. (2021b).

Additionally, the MAT-tagged feature labelled as 'that-relative clauses on object position' was manually corrected to differentiate between relative clauses on object position and that-complement clauses, which were respectively re-tagged as relatives and complements (refer to Section 3.4.3.2 for a report of tagging accuracy check).

Finally, the prepositional feature tagged by MAT was replaced with three distinct types of prepositional features identified by the Stanford parser, as it distinguishes a preposition by its syntactic function while MAT does not. These types accounted for differences in grammatical form and syntactic function at both the clausal and phrasal levels: 1) *of* preposition phrases functioning at the clausal level, 2) other types of preposition phrases at the clausal level, and 3) any types of preposition phrases at the phrasal level. This distinction was introduced, in part, to revisit previous findings and assess whether it offers a more meaningful way to differentiate patterns of variation in L2 writing development.

The initial trials on these features for an MD analysis revealed that the TTR index did not work well with this data set, due to its uneven text lengths. As TTR is known to be sensitive to text length, the factors identified using this measure made the scores computed from the analysis result not reliable. More specifically, the initial trial with this feature along with other grammatical features for a factor analysis did not provide interpretable results. The value of TTR for each text is sensitive to its length, which did not correlate with other features loaded on the same dimension. For example, when a text is very short but discusses an informationally dense topic, TTR did not co-occur features such as nominalisation, or word length on the same dimension. More importantly, TTR is not directly relevant to the grammatical complexity construct explored in this thesis, and it has been used for analysing lexical diversity. Therefore, this feature was removed from the final MD analysis.

The frequencies were normalised to address the issue that the text lengths of the longitudinal PELIC corpus are considerably uneven, as it may cause an inaccurate representation of any feature frequency data in the PELIC corpus. In addition, this issue also raises concern about using much longer texts as a reference, such as the MICUSP corpus, whose average length is much longer. As the main purpose of this analysis is to explore the key dimensions and associated SC markers, a reference corpus was removed from this final MD analysis.

4.5.2. Analysis procedures

Each feature's frequency data extracted from MAT were normalized per 100 words to address uneven text length. Several analyses were conducted to evaluate the adequacy of linguistic variables and the factor solution. In the initial stages, diverse statistical measures were scrutinized to ensure the suitability of features for factor analysis, including assessing correlations between features.

None of the features displayed a correlation exceeding 0.80. The sample's factorability was evaluated using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, and features with a communality value below 0.18 were eliminated. Consequently, 21 features were retained for the final factor analysis, with 53 features removed. The overall KMO value was 0.67. Although Egbert and Staples (2019, p.129) suggested a cut-off value of 0.20 for communality, variables with a marginal communality value of 0.19 typically exhibited acceptable KMO scores.

The final factor analysis was conducted using the psych package (Revelle 2022) in the R environment to identify co-occurring patterns of lexico-grammatical features in the PELIC corpus.

Figure 4-13 shows the eigenvalues of the factors and components based on the 21 linguistic features finally retained for the factor analysis. The eigenvalues of factors represent the percentage of variance explained by each component and factor. The first component and factor are represented by a far-left black and white dot, taking up the predominant proportion of total variance. Several solutions with different numbers of factors were examined in terms of their interpretability. Based on the examination of eigenvalues in Figure 4-13 and the interpretability of the factors, a four-factor solution was chosen. The principal axis factoring method with a Promax rotation was employed for the factor analysis in the R environment.

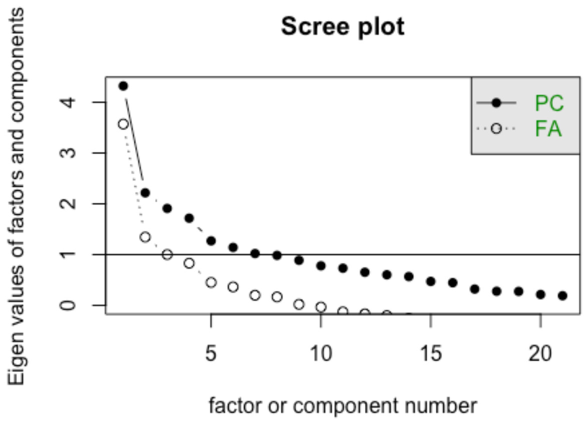

Figure 4-13. Percentage of explained variance by each component/dimension for 21 variables

Factor scores were computed to interpret textual dimensions through the following steps. The normed counts of linguistic features were standardized using the z-score formula. The sum of standardized counts of negatively-loaded features was subtracted from the sum of standardized counts of positively-loaded features (Egbert & Staples, 2019). Only features with loadings greater than +/- 0.35 on each dimension were considered for dimension score computation. Mean dimension scores were then computed for each sub-corpus consisting of five Courses, and sample texts with high dimension scores were examined to explore the functional dimensions of significant features within each dimension.

Furthermore, a mixed-effects model was employed to test significant mean differences among the dimension scores of the sub-corpora, accounting for random effects not explained by the predictors of interest. Mixed-effects models are particularly useful for addressing individual differences in developmental trajectories (Miles et al., 2012). The maximal fixed effects included personal course of study (time) and L2 level, while the maximal random effects encompassed L1 backgrounds, individual writers (nested within their L1 background and L2 level) and writing topics (operationalized based on the identification codes of the essay question prompt in metadata). The models were compared using the Akaike Information Criterion (AIC) score to determine the optimal model for each dimension score. The chosen models were evaluated through Q-Q plots to confirm the normal distribution of residuals and random effects (Zuur et al., 2009).

The following two sub-sections present and interpret the results of the analysis. Section 4.5.3 focuses on the factor analysis results obtained from the factor loadings. It explores the functional dimensions that emerge from the analysis, shedding light on the unique lexico-grammatical variation and communicative functions observed in the L2 written texts. Section 4.5.4 delves into the possible interpretations of linguistic variation over time in the learners' written texts, utilizing the mean dimension scores for the sub-corpora. Additionally, it includes the examination of mixed-effect models applied to the dimension mean scores across the sub-corpora to assess the significance of the time effect on these scores. By considering the factor score means, significant linguistic features, and the interaction between various parameters (such as time, L2 English level, L1 background, and writing topics), this analysis aims to provide insights into the textual dimensions of the L2 written texts and their relationship with the specified factors.

4.5.3 Results & Discussions 1: Linguistic features and associated functional dimensions

The final factor analysis, employing a four-factor solution, explained a cumulative shared variance of 38%. Factors 1, 3, and 4 exhibited positive correlations with each other, while Factor 2 displayed a negative correlation with the other factors. The highest correlation observed among the factors was 0.40, linking the first and last dimensions. The complete report of the factor loadings, encompassing 21 variables, as well as the correlations between factors, can be found in Appendix 4-3. Table 4-6 displays the standardized loadings of the significantly loaded features derived from the correlation matrix.

Table 4-6. Major linguistic features (with loadings larger than ± 0.35) on each dimension (Final study)

Dimension 1	Dimension 2
positive (+) features: second person pronouns (0.7) / conditional adverbial subordinators (0.65) / mental verbs (0.47) / infinitives (0.41) / contraction (0.39) / analytic negation (0.37) ***negative (-) features:*** total other nouns (-0.57), all prep modifying phrases (-0.46), OF preposition modifying clauses (-0.39), (past tense) (-0.35),	***positive (+) features:*** word length (0.58) / (present tense)[11] (0.41) / nominalisations (0.35) ***negative (-) features:*** past tense (-0.85) / first person pronouns (-0.71)
Dimension 3[12]	Dimension 4[13]
positive (+) features: causative verbs. (0.82) / causative adverbial subordinators (0.53) / other propositions modifying clauses (0.51)	***positive (+) features:*** Be verb (0.82) / predicative adj (0.74) / present tense (0.47)

[11] when a feature appears in more than one dimension, only the one with a higher loading (underlined) is included in the dimension score computation.

[12] Positive features only
[13] Positive features only

I have observed the potential risk of losing significant linguistic variability due to limitations in data processing, such as uncleaned data and limited accuracy checks on tagging. Unfortunately, upon examining the statistics for feature suitability in factor analysis, many features distinguishing between phrasal and clausal structures were lost, possibly due to their overall rarity in the corpus processing, such as finite complement clauses controlled by nouns and verbs. These infrequent features, often associated with the limitations of shorter text lengths, exhibited low tagging accuracy rates (see Section 3.4.3.3 for the report on the tagging accuracy of the MAT). Features with fewer observations have a lower chance of contributing to the dimensions (Nini, 2019, p.71). For instance, in this study's data, "that-complement clauses on object position" had a mean occurrence of 0.3 per text, and "that-complement clauses on noun position" had a mean occurrence of 0.8 per text. Conversely, all the features retained for the factor analysis exhibited higher mean occurrences, ranging from 1.4 to 85.1 per text.

Their relative scarcity is noteworthy as they represent instances of structurally complex forms. While Biber et al. (2021a) suggest that 'finite clauses functioning at the phrasal level' may emerge in earlier developmental stages, the corresponding feature analyzed in this study, namely 'that clauses controlled by nouns,' is notably rare. The convergence of general rarity and low tagging accuracy in certain complex constructions, like noun-that clauses, should be acknowledged as a limitation of this study. Enhancing the overall accuracy of tagging in learner corpora could offer insights into comprehending the rarity of elaborated linguistic structures.

Interestingly, Juffs (2020) has noted that some learner writing research suggests that learners may prefer the greater simplicity of verbs followed by noun objects to clauses, which require more extended and complex modifications. Similarly, some studies have found that L2 writers used more compressed linguistic constructions compared to L1 writers in university written assignments (e.g., Pan, 2018; Staples et al., 2023). Taken together, this indicates that writer-related variables (e.g., L1 background) should influence the linguistic variations along with the text-related variables (i.e., communicative purposes inherent in the writing task).

The relative rarity of these structurally elaborated forms is noteworthy. While

Biber et al. (2021) hypothesize that "finite clauses functioning at the phrasal level" emerge in earlier developmental stages, the corresponding feature used in this analysis, "that clauses controlled by nouns," is notably scarce in the data. The English proficiency of the students may have influenced the linguistic choice, as low to intermediate-level students may have difficulty in producing some of the elaborated structures. However, some studies also indicate that L2 learners tend to use phrasal complexity features more than L1 students in university contexts (e.g., Pan, 2018). Taken together, the use of phrasal and clausal features should be influenced by multiple factors, including the L2 writer's L2 proficiency, study level, and communicative task. While studies suggest that academic prose uses compressed linguistic structure for communicative efficiency, students' writing assignments may require different linguistic resources. Therefore, the rarity of the elaborated structure in the PELIC texts necessitates further examination of the linguistic features in context, which will be discussed in Chapters 5 and 6.

The major features loaded onto each of the four dimensions provide valuable information for interpreting the functional dimensions found in the L2 written texts. The following paragraphs focus on the interpretation of these functional dimensions, utilizing excerpts from the texts with the highest and lowest scores for each dimension. Subsequently, the grammatical variations associated with potential communicative functions are discussed, drawing upon the text excerpts with the highest and lowest scores for each dimension.

Dimension 1: personal involvement vs scientific description

text id 3387, Course 2, Level 4 – Dimension 1 score: 19.35

(the positive features in Dimension 1 are in bold)

> **If you want** to fail it, absolutely **you** can, **you** do **not** absolutely **you** can. There are many ways **to fail** a test. I'm going **to write** about it."Do **not trust yourself**" **if** you do **not trust** your ability, I **mean**, **if** you are a pessimist, you will **not** pass the test. For example, **if** you are studying and you say ohh it is too hard I cannot understand it; it'**s** a big problem. My advice for you readers is to **trust yourself** and be an optimistic.

Text 3387, an excerpt from the L2 data, gives suggestions for a good attitude for taking a test. There are occurrences of grammatical errors found in the essay above, but they do not impede understanding the writer's meaning. It uses many of the features positively loaded on Dimension 1, including second-person pronouns, conditional adverbial subordinators, infinitives, analytic negation, and mental verbs. The text employs "if" clauses to create hypothetical scenarios that the reader can imagine themselves participating in.

Many texts scoring high on Dimension 1 exhibit spoken features like second-person pronouns ‘you’, contractions, mental verbs (e.g., think, feel), and infinitive clauses (e.g., to fail). This combination, as seen in text 3387, directly addresses the reader, fostering a unique contribution to the persuasive intent of the writing. These features create an interactive style, inviting reader participation. This function in L2 writing stands in contrast to how Gardner et al. (2019) observed mental verbs with infinitives used by British university students – for establishing a personal stance, not reader engagement. For instance, consider the use of second-person pronouns (e.g., you) and negative statements (e.g., you do not absolutely you can) in Text 3387. Here, the writer directly addresses the reader, potentially trying to influence their beliefs about failing a test. This contrasts with the use of mental verbs and infinitives to express personal stance, as seen in Gardner et al. (2019). Taken together, high Dimension 1 scores in the PELIC corpus are linked to spoken features that create an interactive style for persuasion, contrasting with the use of mental verbs and infinitives for personal stance observed in other studies.

Hyland (2005) emphasizes the importance of reader pronouns in guiding readers through a text. Hyland notes that while pronouns like 'you' and 'yourself' directly address the reader, they are uncommon in most academic writing styles. This exception might be due to the implication of a less informed reader in fields like philosophy, where these pronouns might be used. In contrast, academic writing typically strives to create a sense of shared knowledge and purpose by using the inclusive 'we.' This pronoun fosters a sense of community and shared goals between the writer and the reader. This convention stands in stark contrast to the frequent use of the second-person pronoun 'you' in Text 3387, a feature more characteristic of spoken language.

Text id 26840, Course 1, Level 3 – Dimension 1 score: -18.03

(significant tense is marked in italics; the rest of the significant features are in bold)

> The line graph **of** Springfield Academy *displayed* the population **of** students. From 1992 to 1995, the number **of** foreign students remains the same, but in this period **of** time the number **of** American students increased slowly. In 1996, the number **of** foreign students *rose* to equal the number **of** American students. Between 1996 and 1998, the number **of** American students *decreased*. From 1996 to 2002, the number of foreign students *was* more than the number **of** American students.

The essay above describes a chronological shift in the student population between home students and international students in America. This essay is a good example of how communicative purposes influence linguistic choice. It uses descriptive verbs and adverbs for numeric description (e.g., remained steady, increased gradually, surged). The overall tone is neutral and indifferent, relying heavily on prepositions and nouns to convey information. Notably, the essay lacks features typically associated with a high Dimension 1 score, such as conditional adverbial clauses, analytic negation, and mental verbs. While it does use the infinitive "to equal" once, its use is still limited. These features often introduce a more complex and nuanced analysis, which is absent here.

> D1 scores reveal a spectrum of writing styles. Texts with low D1 scores tend to be more factual and impersonal, relying heavily on prepositions (both within clauses and modifying phrases) and noun phrases for description. These texts often lack features commonly associated with high D1 scores, such as mental verbs (e.g., *think, believe*), infinitives (e.g., *to suggest, to persuade*), conditional adverbial clauses (e.g., *if you consider, in some cases*), second-person pronouns (e.g., *you*)
>
> Conversely, texts with high D1 scores exhibit a more interactive and persuasive style. These texts frequently use second-person pronouns and mental verb constructions to create a sense of personal involvement. This spoken style aims to engage the reader and potentially convince them of the writer's viewpoint. In essence, Dimension 1 reflects the writer's choice between a personal and engaging style versus a more impersonal and factual approach.

The focus on reader engagement in D1 is noteworthy. While Hyland (2005) acknowledges the importance of reader positioning and stance, he highlights the relative neglect of writer-reader interaction in academic prose research. This includes how writers anticipate and address potential reader objections (Hyland, 2005, p.182). This observation by Hyland warrants further discussion, considering the unique role "engagement" plays within the PELIC corpus.

Dimension 2: informational density vs personal narratives

text id 26207, Course 2, Level 4, Dimension 2 score: 8.7 (cf. D1: -4.22)

(nominalisations are in bold)

> **Education** is one of the most important problems that affect a lot of countries. Most importantly, quality of **education** which may be qualified or unqualified has a dramatic impact on the community. Furthermore, unqualified education might affect the **community** in different features and they are: poor economy, poor social **relations** and unsafe communities. Recent researches clarify the relation between the economy and the education. Meanwhile, useful and successful financial **situation** of the **population** accomplish by qualified **education**. However, the unqualified **education** might lead to financial **corruption** and unfair spending of money by **governments** or the people themselves.

The provided excerpt 26207 represents an argumentative essay focusing on the topic of education. Throughout the essay, there is a notable utilization of features commonly associated with both literary and scientific writing styles, particularly the use of nominalization. This linguistic technique facilitates a discussion with a more objective and conceptual tone about the broader ramifications of education. Text 26207 exemplifies that Dimension 2, characterized by positive loadings for nominalizations, longer words, and the present tense, is associated with a more formal and impersonal writing style.

12158 – Course 3, Level 5, Dimension 2 score: -10.6

(significant tense is marked in italics; first-person pronouns are in bold)

> And **my** dog *barked* toward me. He *looked* hungry and wanted to go bathroom. The dog and cat *made* noisy. **I** *could* not eat breakfast because **I** should have fed **my** pet and picked dog up the bathroom. When **I** *went* outside to take my car, **I** *realized* that **I** *was* wearing a pajama. **I** *had* to change **my** clothes. Finally, **I** *left* my house. **I** *was* thought that it *was* over nightmare. However, when **I** *saw* traffic jam, **I** *could* feel frustrating.
>
> On **my** way to work, **I** *decided* to take a shortcut through on old part of town. When **I** *took* a shortcut, there ***was*** horrible car accident. **I** *could* not move anywhere. **I** *could* not wait. **I** *decided* **I** *parked* on the way

and **I** *ran* toward a main street. **I** *could* catch a taxi. However, **I** *came* into **my** boss's head.

Text id 12156, Course 3, Level 5, Dimension 2 score: -9.78

(significant tense is marked in italics; first-person pronouns are in bold)

When **I** *woke up* the next morning, **I** *was* surprised to find that **I** had overslept and *would* be late for work. As **I** *rushed* down the stairs to eat a quick breakfast**, I** *tripped* over my cat and *had* to hurry up. **I** *ran* into my car quickly, but **my** car *did* not work, so **I** *had* to walk to **my** company. **I** *had* just 30 minutes.

The provided excerpts 12158 and 12156 are identified with a similar communicative purpose, personal narrative. This type of writing purpose has been found consistently in texts with a low Dimension 2 score. The writers employ a retrospective voice to narrate a personal event. This narrative style is facilitated by the consistent use of first-person pronouns and the employment of past tense verbs throughout the text. These linguistic features effectively convey the narrative to the reader. These texts receive a low score on Dimension 2, primarily due to the increased occurrence of negatively loaded features, namely the use of first-person pronouns and past tense. These features are commonly associated with the function of narrating personal events in a retrospective manner. It is noteworthy that while the corpus primarily consists of persuasive or descriptive texts, this personal narrative, as a distinct genre, stands out within the realm of academic writing genres. The presence of this text type in the corpus can potentially be attributed to the learner-centred teaching philosophy underlying the language programme.

Considering the linguistic usage and the associated communicative functions, Dimension 2 is labelled as "informational density vs personal narrative," reflecting the identified communicative functions associated with both the positively and negatively loaded features.

Dimension 3: Persuasive explanation

31247, Course 3, Level 5, D2 score: 2.78 cf. D4: -2.09

(Nominalisations in bold; present tense in italic)

American society, as well as other societies, is used to defining marriage as a union between a man and a woman. Today, same sex marriages have become a phenomenon in the United States.

> Homosexuals are claiming their rights in all states and they want their union to be recognized by law.
>
> Consequently, same sex marriage has become one of the most heated arguments in the United States. The controversy around same sex marriages brings about heated discussion in many areas. (– the remaining part truncated)

Text 31247 employs present tense frequently, which is significantly loaded on both Dimension 2 and Dimension 4. While this feature has been categorised as a major Dimension 2 feature, it is interesting that Text 31247 scores high for Dimension 2 but it scores low for Dimension 4, indicating the complementary distribution of the features in these dimensions. This text also uses nominalisation actively, which adds to a more literary style of the texts. These features serve communicative functions in argumentative essays.

text id 20810, Course 3, Level 4, Dimension 3 (+) score 6.93

(the significant features in Dimension 3 are in bold)

> I got along well **with** my coworkers **because** we were both **from** UAE. All of us were born **in** the same city, Dubai. We also liked the same food so we would go out to lunch together. Both of my coworkers were good workers, they were not lazy. The similarities between me and my coworkers are what **caused** us to get along so well.

Text 20810 incorporates a personal experience as supportive evidence to bolster the argument presented. The writing style exhibits a personal tone and employs narrative elements, strategically employed to elucidate the significance of cooperation within the work environment. To accomplish this objective, the text makes use of causative verbs and causative adverbial subordination. In general, texts that employ the prominently loaded features in Dimension 3 tend to adopt a persuasive voice in order to argue their viewpoint or explicate the causal relationship underlying a particular incident. Consequently, Dimension 3 is appropriately designated as 'persuasive explanation,' reflecting the communicative intent associated with the features that exhibit high loading in this dimension.

As discussed in Section 4.3, the ‘persuative involvement’ text type has been identified as the characteristic of the longitudinal PELIC texts. This dimension, marking ‘persuasive explanation’ resonates well with this textual type.

Dimension 4: Evaluative description of the world

2756, Course 2, Level 4 – Dimension 4 score: 7.39

(present tense is marked in italics; 'be' verbs and predictive adjectives are in bold)

> I ***am*** agree. because if you ***are*** **lucky** the good luck will bring something that you want or need with out any thing. I mean, you might win the lottery ant take $1,000,000 dollar. then you will **be** so **happy**, and you will have a great life in spit of you didn't deed any thing. I asking my God to give me a good luck because I *think* I ***am*** not lucky person, and I *want* to **be rich** like kings. This **was** my opinion it *does*n't have to **be true**, it might **be** falls. *do*n't believe every thing that people *say*. Good Lock.

There are substantive occurrences of spelling and spacing errors found in the essay above, but they do not impede understanding the writer's meaning. The essay exhibits very low tokens but prevalent use of present tense, BE as the main verb and predicative adjectives. These devices are used for persuasion purposes. Therefore, this text is regarded as a text carrying out the purpose of persuasion as apposed to 'informational density and lexical diversity.'

text id 3631, Course 1, Level 3, Dimension 4 score: 5.11

(significant tense is marked in italics; 'be' verbs and predictive adjectives are in bold)

> Bedouins *have* a difficult life because they *live* in desert and they often *travel*, but they *have* good life because they ***are*** modest in their life and *have* special custom. I *have* many reasons for that and now I *tell* you about how do people life and customs people. Firstly, most people ***are*** *hospitable* and kindness because they *like* to meet people. For example, when anyone *visits* them, they *kill* one lamb for him or her and *give* him or her home, and it ***is*** *free*. Secondly, almost they ***are*** *honest* and *friendly*, so they often *do*n't close their houses.

Text 3631 examines the advantages of residing in a specific national territory from cultural and social perspectives, employing a personal evaluative tone. The presence of present tense "be" verbs with predicative adjectives primarily serves the purpose of describing and evaluating various aspects and entities related to the topic. This evaluative tone is frequently conveyed through the use of predicative adjectives. Hence, Dimension 4 is identified as an 'evaluative description of the world.'

An interesting observation from the aforementioned text is that the presence of these features is often accompanied by a suppression of mental verbs and to-infinitives. In other words, these two sets of features tend to exhibit complementary distributions, even though both extensively employ the present tense. Consequently, there is a positive correlation between Dimension 1 and Dimension 4 (=0.40).

Examining representative texts from each dimension reveals a close relationship between the major syntactic features, the topics addressed, and the communicative purposes of the writing (Biber & Conrad, 2009). This suggests that the choice of linguistic forms and functions is influenced by the chosen topics. Consequently, the trajectory of linguistic variation throughout the learners' study likely reflects both their individual topic preferences and their developing awareness of the more conventional forms of academic essays expected in university writing activities (Myles, 2002).

This analysis identified two key dimensions. Dimension 1 groups features associated with literate and academic writing, alongside those used in more elaborate and narrative styles. Dimension 2 captures features used for functions like explaining cause-and-effect relationships and making evaluative descriptions. These two dimensions show complementary distributions, meaning the features within each rarely overlap. However, the corpus also reveals the use of some features across both dimensions, suggesting their adaptability for various communicative purposes. These findings emphasize the importance of considering the interplay between writing topics, learner-centered teaching philosophies, and process writing methods. Understanding these relationships is crucial for interpreting the variations in sentence structure (linguistic variation) observed in the PELIC longitudinal written corpus.

4.5.4 Results & Discussions 2: Textual dimensions across sub-corpora

As outlined in the previous section on methodology (Section 3.3.4.2), the interpretation of a textual dimension involves several considerations. These include: (1) analysing the factor score means across different sub-corpora, (2) examining the situational parameters associated with the distribution of factor scores, and (3) identifying the significant linguistic features that contribute to each factor (Biber, 1988, p.97). Through a comprehensive examination of these features and their shared communicative functions, we can establish the functional interpretation of the extracted factors as three distinct linguistic dimensions. The mean values of each dimension's factor scores, along with their corresponding error terms, are presented in Figure 4-14 to 4-17. These figures illustrate the relationship between factor scores among the text groups at each time point.

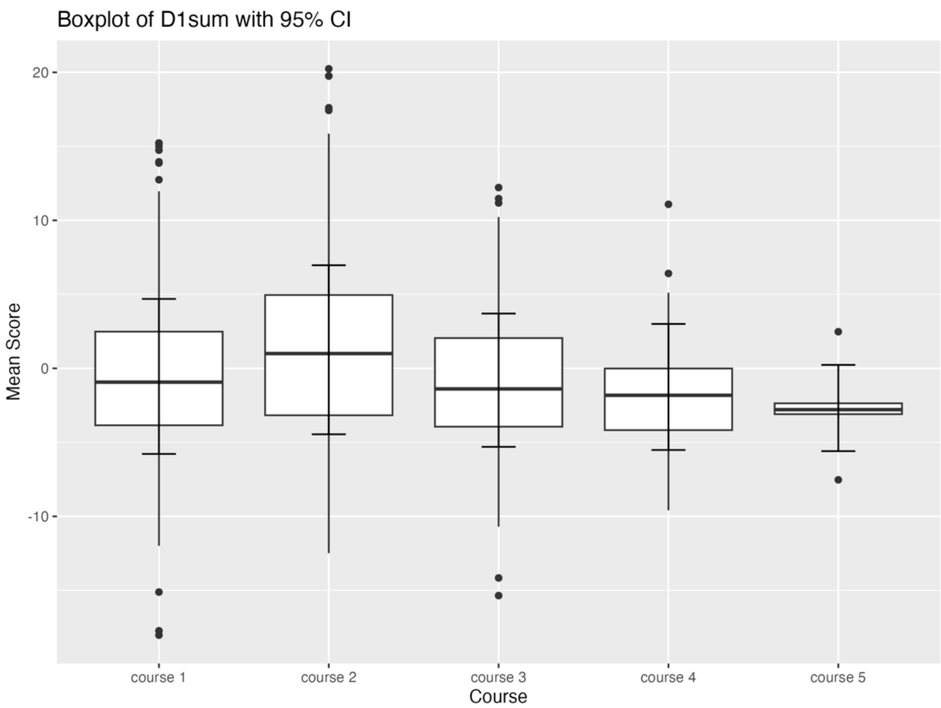

Figure 4-14. Dimension 1: Personal involvement vs Scientific description*14***.**

[14] A 'course' refers to the order of the semesters in which individual writers submitted each text while studying in the language programme. In other words, 'course' corresponds to an individual course of study.

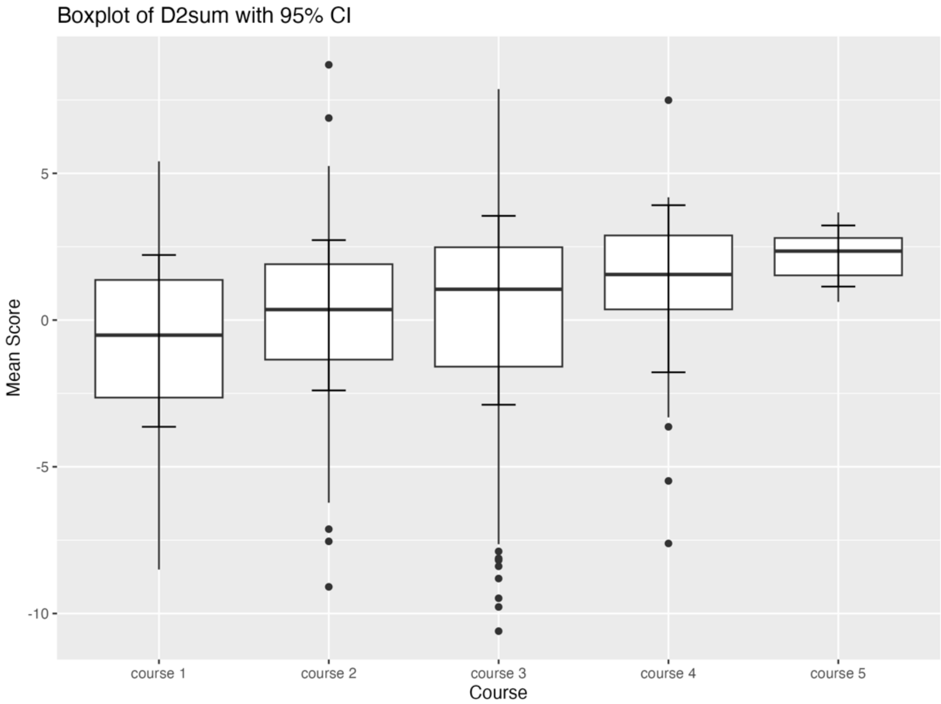

Figure 4-15. Dimension 2: Informational density vs Personal narratives

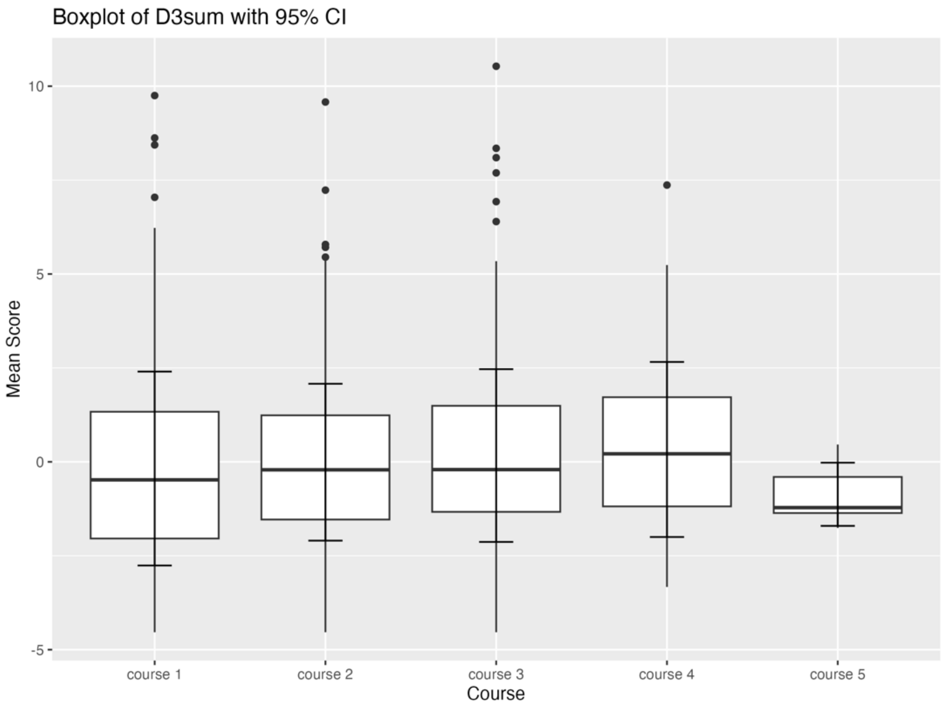

Figure 4-16. Dimension 3: Persuasive explanation

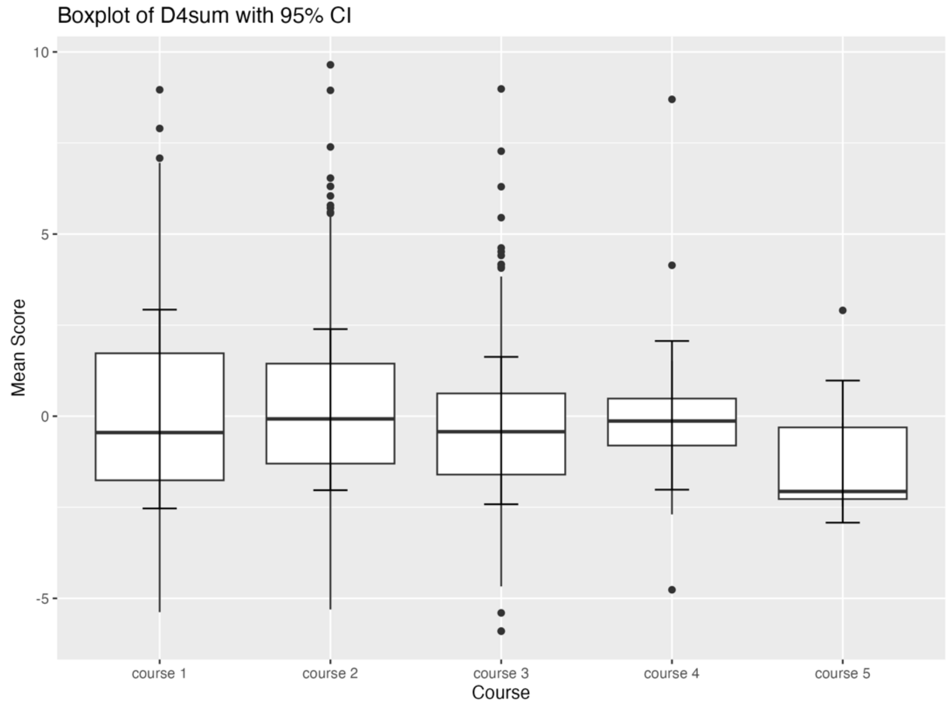

Figure 4-17. Dimension 4: Evaluative description of the word

Dimension 1 scores exhibit a trend of gradual decrease over time, although this decrease is relatively small. It's important to note that the 95% confidence intervals for these scores may overlap, making it difficult to definitively claim a statistically significant decline. Despite this uncertainty, the observed pattern does suggest a diminishing presence of features associated with Dimension 1.

Interestingly, Dimension 1 also displays a non-linear pattern in this decline. This suggests that the decrease may not be uniform across the writing samples.

While the overall trend points towards a potential increase in phrasal features, particularly prepositions, we should be cautious in interpreting this as a definitive developmental trend due to the potential for chance variation. However, this observation aligns with previous research suggesting a growing prominence of features associated with mature writing over time (e.g., Gray et al., 2019).

Biber et al. (2021a) propose that infinitive phrases controlled by verbs (e.g., *to suggest*) may emerge earlier in writing development (stage 3) compared to prepositions used as postmodifiers (stages 4 and 5). This aligns somewhat with the observed non-linear pattern in Dimension 1 scores within this dataset. However, it's important to remember that the confidence intervals for these scores may overlap, making it difficult to definitively claim that one feature consistently appears before another.

With this caveat in mind, the decline in Dimension 1 scores might tentatively suggest a pattern where features associated with spoken language, like to-infinitives and mental verbs, are initially more frequent. These features could then be gradually replaced by features associated with more mature writing, such as prepositional phrases modifying clauses and noun phrases.

Further research with larger datasets and more robust statistical methods is needed to confirm these potential trends of all Dimensions, considering their overlapping confidence intervals in Figures 14-17.

Similarly, the increasing pattern observed in the Dimension 2 scores across the sub-corpora indicates a growing prominence of positively loaded features, such as nominalization, in the later stages of the study. The mean score of Dimension 2, associated with informational density, demonstrates a relatively steady increase over time. Conversely, the sharp increase in the Dimension 1 score in Course 2 suggests that students employed features related to personal involvement more frequently compared to other time points. This non-linearity in the Dimension 1 score diverges from the usage-based perspective on linguistic development. As seen from the example texts from the PELIC corpus presented in Section 4.6.3, the linguistic use in the PELIC texts was heavily influenced by the topics and the communicative purposes. Text 26840 was a good example of this point; when written at Level 3 and Course 1, the text addressed the impersonal description of the population trend over time. The language used for that aim exhibited characteristics closer to that of academic written registers. The presence of this type of text indicates that the experience with conventional literary writing tasks may be a stronger predictor of language use than the proficiency levels assigned during their time in the programme.

To investigate the influence of both L2 level and time on the writing development of students, a mixed-effects modelling approach was used. This approach considered the standardized mean score for each dimension (1-4) while accounting for random effects like writing topic and individual writer. The models, incorporating interaction terms between L2 level and time (course) variables, are summarized in Table 4-7.

It is worth noting that the scores of Dimensions 1, 3, and 4 in Figures 4-14, 4-15 and 4-17 displayed non-linear patterns in the scores of the dimensions, which may not be adequately captured by linear regression analysis. To address this, separate predictor variables (fixed-effects) were included for Courses 2, 3, 4, and 5, with Course 1 serving as the reference time point. The aim was to determine if the slopes of these variables against Course 1 were statistically significant, thereby accounting for any potential non-linear relationships.

The time and proficiency interaction is not always present in the result of the mixed-effects models, but they always showed very strong correlations for the variables, indicating multicollinearity. However, they both were kept in the regressions for two reasons. The first reason was theoretical. Even though they showed the multi-collinearity issue, they were theoretically distinguished variables. It was suspected that the level assignment considered various components scores and qualitative decision making, where the level test closely aligned with the instructions in each level, therefore, students are likely to be better prepared for proceeding to the next levels during the semester. Therefore, the temporal order of semesters is likely to align with the level upgrade. Second, most of the models showed better fit when both variables were entered, indicating that they both contributed to explaining the variance. Moreover, several models did not show clear interaction between them, while others did. This indicates that they are appropriate to be treated separately. This is a limitation as to how to establish the model of interaction between the two variables. They should likely interact in certain situations while they do not in others. It was not possible to trace it and establish a model of such interaction, which should be worth further investigation in future research with more suitable statistical models than adopted in this study.

Table 4-7. Mixed effects models for Dimension mean scores

Models	Dimension1	Dimension2	Dimension3	Dimension4
Fixed effects				
Course 2	< 0.001	0.04	ns	ns
Course 3	ns	ns	ns	ns
Course 4	ns	ns	ns	ns
Course 5	0.009	ns	ns	ns
L2 level	ns	0.001	ns	< 0.001

Course 2 * L2 level	0.001	0.05	ns	ns
Course 3 * L2 level	ns	ns	ns	ns
Course 4 * L2 level	ns	ns	ns	ns
Course 5 * L2 level[15]	-	-	-	-
Random effects	Topic, Writer	Topic, Writer	Topic, Writer	Topic, Writer
R2				
Marginal	0.05	0.07	0.01	0.05
Conditional	0.52	0.65	0.23	0.35

Table 4-8. Mixed effects models for Dimension 1 mean scores

Fixed effects	value	Std.Error	t value
Intercept	-3.9926	1.4631	-2.729
Course 2	10.18	2.14	4.76
Course 5	-4.14	1.3	-3.18
L2 level : Course 2	-2.3	0.53	-4.3
Random effects	Variance	Std.Dev.	
topic	4.155	2.038	
student	1.574	1.255	
Residual	5.871	2.423	

Table 4-9. Mixed effects models for Dimension 2 mean scores

Fixed effects	value	Std.Error	t value
Intercept	0.05239	0.98341	0.053
Course 2	-2.12561	1.43779	-1.478
L2level	0.04266	0.27808	0.153
L2 level : Course 2	0.38753	0.35893	1.08
Random effects	Variance	Std.Dev.	
topic	1.9348	1.391	
student	0.7275	0.853	
Residual	2.5771	1.605	

[15] Course 5 consists of level 5 only, therefore the interaction term was dropped from the correlation matrix.

Tables 4-8 and 4-9 show the details of the mixed effects models of Dimensions 1 and 2, respectively (the models of Dimensions 3 and 4 were excluded due the non-significance of their results). The mixed-effects models in Table 4-7 reveal significant increases in the Z-scores of the phrasal complexity features at Course 5 in Dimension 1, providing support for the significant time effect documented in previous longitudinal studies (e.g., Gray et al., 2019). Specifically, Course 5 exhibited a significant time effect in Dimension 1, characterized by a negative slope, albeit non-linear in other courses), while Course 2 displayed a significant time effect in Dimension 2 with a positive slope. These findings indicate a growing frequency of phrasal complexity features associated with literate writing (e.g., prepositional phrases in Dimension 1) and informational density (e.g., nominalization in Dimension 2) at these specific time points.

These results partially confirm my initial prediction that phrasal features related to literate writing and informational density would become more salient over time. Analyzing the fixed effects in Dimension 1, I observed that the substantial increase in the Dimension 1 score at Course 2 was statistically significant, reflecting a preference for functions commonly found in oral communication during the early stages of language development. This significant nonlinearity observed in Course 2 may be interpreted as a period of adaptation for learners, during which they need time to activate their receptive language knowledge and transform it into productive knowledge, particularly when adapting to English writing as an additional language. This adaptation process may involve selecting topics that provide a greater sense of comfort and familiarity. However, these effects diminish from Course 3 onwards, until the negatively loaded features gain more significant saliency in Course 5, aligning with the expected trajectory of writing development.

Importantly, the random effects of writing topics and individual writers accounted for a considerably larger proportion of variance in all four models, compared to the relatively minimal variance explained by the fixed effects alone (as indicated by the marginal R2 values in Table 4-7). This underscores the significant role played by the chosen writing topics and individual writer variations in explaining linguistic variations across all dimensions, beyond what can be attributed to time and L2 level variables alone.

4.5.5 Conclusions & Next steps

The analysis results support the hypothesis regarding linguistic developmental patterns in L2 academic writing. Specifically, the L2 learner corpus exhibited an increase in the frequency of phrasal features as students progressed in their studies, indicating greater phrasal complexity. This longitudinal analysis provides further evidence for the linguistic developmental paths in L2 academic writing from a usage-based perspective. The findings align with previous research (Biber et al., 2011) suggesting that phrasal coordination is a crucial aspect of grammatical complexity in academic writing.

The dimensions capturing syntactic and other related linguistic features shed light on how these features are employed in scientific and literate writing styles. Overall, the identified functional dimensions in the longitudinal PELIC texts consist of linguistic features marking spoken registers, such as personal involvement (Dimension 1), persuasive description (Dimension 3) or evaluative description of the world (Dimension 4). The major linguistic features of these dimensions show how a writer sees the given writing topic or expresses their opinions on it, actively involving the reader. Dimension 2, on the other hand, relatively exhibits a more traditional characteristic of written academic registers, with major linguistic features such as nominalisation. In contrast to the relatively inconsistent patterns observed in the dimensions scores of other dimensions across Courses, Dimension 2 showed a steady increase over time, albeit not always statistically significant. As Díez-Bedmar and Pérez-Paredes (2020) argue, premodifying slot is a key marker of the development of syntactic complexity in noun phrases. In this regard, the steady increase of nominalisation leaves an area for further pursuit of the enquiry of nouns as modifiers in academic writing development.

The study also revealed a significant influence of writing topics on linguistic variations in the L2 writing data, echoing findings by Lan et al. (2019) regarding the impact of topic on the use of prepositional phrases. Similar to their observations, the current study found that topic significantly influenced the use of linguistic features, particularly in relation to proficiency levels. However, unlike Lan et al. (2019), who focused on noun modifiers, this study explored a shared communicative function of a broader range of linguistic features.

Notably, the nonlinearity observed in the dimension scores and the early decrease in elaborated constructions during factor analyses should be considered within the situational context. The learner-centered approach to language learning in the language program likely provided students with more freedom in selecting writing topics while aligning them with conventional academic genres. This aligns with the suggestion by Lan et al. (2019) that a balance between topic control and writing fluency is necessary in L2 writing instruction.

Subsequent mixed-effects models considered the potential mediating effects of extralinguistic variables, such as L1 backgrounds, L2 proficiency levels, and topic effects. The influence of L2 proficiency on the scores of Dimensions 2 and 4, a well-documented finding in previous literature, was not directly examined in this study. It is important to note that the data used in the current study focused only on writers who contributed for three or more semesters, resulting in an imbalance in terms of time, English proficiency, and language background. Although mixed-effects models were employed to address this data imbalance, missing information regarding students with shorter periods of study in the language program was unavoidable.

Chapter 5 Predicting Development of Clausal Complements for Nouns and Adjectives in L2 Writing

5.1 Introduction

One of the findings from the previous studies of syntactic development discussed in Chapter 2 is that phrasal complexity features are crucial markers of writing registers (e.g., Biber et al., 2014). Moreover, Chapter 4 analysis results indicate the nonlinearity of linguistic use in L2 writing. This nonlinearity in L2 writing development has been also noted in previous studies (e.g., Man & Chau, 2019). In a study of evaluative that-clauses based on a longitudinal corpus of argumentative essays by L2 university students, Man and Chau (2019) found a non-linear trajectory of change in the use of this construction over time while verb that-clauses decreased whereas the proportions of noun that-clauses and adjective that-clauses increased. This finding suggests that the nonlinear trajectory of L2 writing development still reflects the phrasal complexity as a marker of L2 writing development.

Three studies are reported in this chapter: a pilot study using two SC indices based on clausal elaboration (Section 5.2), another pilot study using one feature of prepositional construction (Section 5.3) and a final study using two clauses associated with the stance function (Section 5.4).

The analysis focuses on variations regarding the frequency measures of syntactic complexity indices processed by TAASSC at phrasal level (1) in texts among the first, second and third versions of texts and (2) in all the texts produced by the same learner throughout three consecutive academic semesters. This pilot study is motivated by the findings drawing on the SC indices with the assumptions on elaborated structures in academic writing. As has been found in previous studies, the ability to use structures associated with phrasal complexity is characteristic of academic writing. Therefore, the ability to use these structures is a marker of writing development. Empirical studies have also proven this to be significant in L2 writing (e.g., Larsson & Kaatari, 2019; 2020), and the pilot study aims to add longitudinal evidence to this. As previously noted, phrasal complexity is a meaningful marker of writing development, and the general aim of the analyses in Chapter 5 is to gain preliminary analysis results, which could provide the basis for further investigation into a meaningful phrasal complexity variation specific to the PELIC longitudinal texts. As a result of the analyses, in summary, the pilot analyses reported in Sections 5.2 and 5.3 indicate little insight into significant syntactic variation; however, the pilot analysis results still provide meaningful information about the PELIC longitudinal variations in the phrasal complexity, and therefore discussed in Section 5.2 and 5.3. These pilot analyses lead to employing a different linguistic measure for the noun and prepositional constructions (in Section 5.5). The overall results of the analyses in Chapter 5 also call for a more nuanced method, as introduced in Chapter 6.

As in the analyses in Chapters 4 and 6, the studies in this chapter also focus on the temporal order of the texts produced by the same writers, rather than the English levels assigned to the texts. The temporal order and the assigned levels tended to be highly correlated in the mixed-effects models reported in Chapter 4. However, this different focus from the previous studies reported by Juffs (2020) is based on the assumption that these are fundamentally different factors influencing the syntactic variations. In light of this, the current studies focus on the temporal order of the syntactic variations, which are supposed to complement the evidence based on the writings by human rating.

These analyses attempt to answer the following questions about the syntactic constructions of focus: 1) How do the frequencies of the key constructions vary over time? 2) Are the variations of these two constructions influenced by writer-specific (L2 level, L1 background) or situational (writing topic) variables?

To explore these questions, mixed-effects modelling (MEM) was used to make inferences regarding the linguistic variation over time, which may vary in magnitude among individual texts. As noted by the mixed-effects model results of Dimension scores in Chapter 4, the patterns of linguistic variables in the L2 data require to be modelled in consideration of its multifactorial nature. A principal advantage of using MEM is that it provides the researcher with a way to model individual differences through the inclusion of random effects (Miles et al., 2012). It should be noted that the MEMs used in this chapter draw on a Bayesian theory. While the variables analysed in a mixed-effects model in Chapter 4 were the mean scores of Dimensions, raw frequencies of the key constructions are analysed in this chapter. The Bayesian approach was chosen instead of the frequentist model (used in Chapter 4) because the count variables (frequency of linguistic variables) are handled in theory more accurately by Poisson distribution, and Bayesian models are more flexible to deal with different assumptions including different data distribution (Winter & Bürkner, 2021, p.4).

Chapter 2 has discussed that varying patterns between cross-sectional and longitudinal data indicate that while linguistic characteristics observed at specific proficiency levels might remain relatively constant, the routes individual learners follow to attain these levels can be diverse (Kyle, Crossley and Verspoor, 2021). In addition, the broad patterns of grammatical variations in L1 and L2 may be similar, but the specific differences in these two separate registers are still notable (Staples et al., 2023). These observations add weight to conducting longitudinal analyses on these research enquiries.

The motivation for conducting the second pilot analysis (Section 5.3) is to estimate the optimal prior for a mixed-effects model for the longitudinal PELIC corpus. As detailed in Section 5.1, the prior is informed by previous analyses and theory-based assumptions. As some of the grammatical variation in the longitudinal PELIC corpus has been nonlinear, it was essential to check an optimal prior with a pilot analysis. The linguistic variables for the pilot analysis are built upon the previous analysis, as will be introduced in Section 5.3.

5.2 Pilot study

5.2.1 Introduction

This pilot study aims to find evidence of syntactic complexity development throughout the revision process using longitudinal data. The analysis in this section explores syntactic complexity development in English writing in longitudinal L2 data collected in the US, focusing on two SC indices operationalised to measure the following constructs: the adjectival and prepositional modification indices processed by TAASSC (Kyle, 2016).

This study aims to find differences in syntactic complexity between 1) first, second and third drafts for the same topic written by the same student and 2) drafts on one or more topics written by the same student throughout at least three consecutive academic semesters. The studies in this chapter add longitudinal evidence on linguistic development in a non-experimental setting with SC indices informed by the previous analyses discussed in Chapter 4.

The analysis focuses on variations regarding the frequency measures of syntactic complexity indices at phrasal level (1) in texts among first, second and third versions of texts and (2) in all the texts produced by the same learner throughout three consecutive academic semesters. The meaningfully salient indices in the data will be interpreted in terms of their functions within a sentence. Based on this, it aims to find the frequently found syntactic features in the text revision process that can provide evidence on functional dimensions of the syntactic features indicative of quality writing.

As discussed in Chapter 2, many studies show that using syntactic measures proposed by Biber et al. (2011) provides meaningful interpretations of L2 writing quality. However, some of the findings are conflicting due to the differences in corpora and statistical methods. For example, attributive adjectives were interpreted as indicating lower levels of development in Parkinson and Musgrave (2014). This feature was the marker of high-scoring written integrated texts part of the TOEFL by Biber et al. (2014). In part, this can be explained by the different writing purposes and maturity of authors represented in the corpora used in each study. However, adding a functional dimension to syntactic development research may complicate text feature identification, causing difficulty in interpreting or generalising the results. While this functional approach to syntactic complexity at a more fine-grained level potentially provides better explanations for L2 syntactic complexity development, it should be noted that it needs cross-validations by replicating all or partial data or methods to generalise it to the larger target population.

The motivation for including this additional analysis is as follows. The MDA in Chapter 4 also included features related to these two features, and they were also significant in terms of the factor loadings. These features are related to phrasal compression, which was the key construct underlying the variables for

SC in MDA. Therefore, using a different set of measures exploring the same construct might add complementary information on this assumption of phrasal compression for L2 writing development.

5.2.2 Method

5.2.2.1 Corpus data

This pilot study utilizes a selection of texts from the PELIC corpus, which offers a valuable resource for examining writing revisions. Notably, over 11% of the corpus contains multiple versions of student texts with revisions based on teacher feedback. For the current analysis, 326 texts produced by 55 students were chosen. This selection includes 148 first drafts and 178 drafts revised either once or twice based on formative teacher feedback. The data was selected to complements the primary data, the PELIC longitudinal corpus used for the final analyses reported in Chapters 4, 5, and 6. The data selection was based on data selection considerations discussed in section 3.2.1. Table 5-1 below shows the number of texts collected throughout academic semesters, which indicate the course of study taken by 55 students.

Table 5-1. Number of texts throughout the semesters (course of study) in which students submitted drafts

Course of study	1st draft	2nd draft	3rd draft	Total
Course 1	51	51	8	110
Course 2	53	53	16	122
Course 3	40	40	5	85
Course 4	4	4	1	9
Total	148	148	30	326

The bar chart plotted using the ggplot function in the tidyverse package (Wickham et al., 2019) in Figure 5-1 shows the average length of texts produced along the course of study by each student and the revision index.

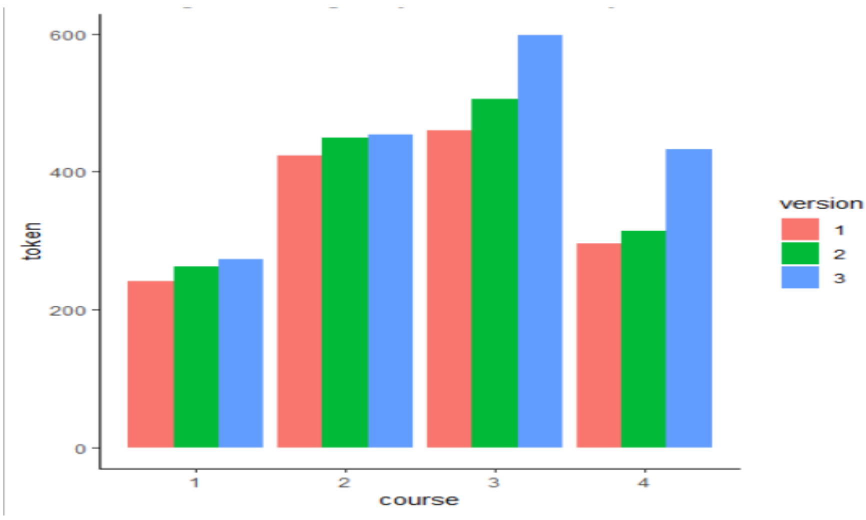

Figure 5-1. Average text length (word count per text) by course of study

Overall, the texts became longer as student revised their first drafts, and the average length of texts increased as their study progressed. This correlation between length and revision - the exception is the fourth unit of observation, to which only one student contributed their drafts - seems to reflect partial improvement made while dwelling on the constructive feedback on their first draft. Other extralinguistic influences also should be considered regarding the length variation, such as the requirement of writing tasks including the word limit, as specified in the rubric of the writing placement test. The objectives specified at the placement level, intended to be a proxy of writing class tasks, reflect the length of essays at each level (Juffs, 2020).

5.2.2.2 Procedures

Previous studies support the use of phrasal indices, incorporating the functional dimensions of syntactic complexity features (e.g., Larsson & Kaatari, 2020; Kyle & Crossley, 2018). Defining language structure in terms of its form and function leads to considering the issue of consistency for defining a language structure. Let's consider an example to illustrate the benefit of using software for tagging linguistic features. Imagine a grammatical construction where a noun modifies another noun phrase (e.g., *the teacher's kindness*). Defining this construction requires considering both its form (e.g., the presence of a noun suffix like "-ness" or a plural "-s") and its function (acting as a modifier within the noun phrase).

Manually defining such constructions can be a two-step process and can lead to inconsistencies across researchers, even if they agree on the overall concept. On the other hand, automatic taggers enable reliable tagging by applying the same pre-defined rules to every instance of text. This ensures consistency in how features are identified, regardless of the researcher. Imagine multiple researchers analyzing the same text – with an automatic tagger, they can be confident that each researcher is identifying the noun modifying the noun phrase (*teacher's kindness*) in the same way. This consistency allows for more reliable comparisons across studies, compared to potentially subjective manual tagging approaches. This present study uses the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016) to obtain syntactic complexity frequencies to explore syntactic development. I aimed to conduct the pilot analyses based on the features used for the MDAs discussed in Chapter 4, as these features have been noted as markers of a written academic register in previous literature and key features identified from the PELIC longitudinal corpus.

Texts are first tagged for part of speech, and then the tags are used, in conjunction with the phrase structure rules, to generate several competing sentence-level parse trees (Kyle, 2021, p.7). TAASSC (ver. 1.3.8) calculates 372 indices in five categories, including clause complexity and phrase complexity, and incorporates the traditional classic syntactic complexity indices analysed by L2SCA (Lu, 2010).

As for the phrasal indices, TAASSC includes 132 indices based on seven noun phrase types and ten phrasal dependent types. These indices offer various ways to analyse noun phrase complexity. Informed by the previous literature discussed in Section 2.3.1, this analysis aimed to investigate aspects of phrasal complexity previously linked to L2 writing development. I specifically chose two indices related to adjectival and prepositional constructions, as shown in Table 5-2. These indices provide insights into how these constructions are elaborated by the average number of noun phrases within them. By examining these indices, I aimed to gain information about syntactic complexity at the phrasal level.

Table 5-2. The two phrase types processed by TAASSC (adapted from Kyle, 2016, p.57).

	Phrase Type	**Abbreviation**	**Examples**
Phrasal complexity	occurrence of adjectival modifiers per nominal	amod	*The man in the [red]$_{amod}$ hat gave the [tall]$_{amod}$ man the money*
	occurrence of prepositional dependents per nominal	prep	*The man [in the red hat]$_{prep}$ gave the tall man the money* poss *That is [her]$_{poss}$ red car*

The accuracy of tagging was manually checked with 20 random samples, as reported I n Section 3.4.3. Each of the indices produced from TAASSC is fitted with linear mixed-effect modelling (MEM). Variables violating a normal distribution were discarded. More specifically, the model's residuals were plotted to see if its residuals are approximately normally distributed. The R2-values were calculated using the r.squaredGLMM() function in the MuMIn package (Bartoń, 2017). For the key syntactic complexity measure to focus on in this chapter, texts that ranked high and low respectively on the different measures were investigated to interpret the results linguistically.

Finally, the following two models were fitted using the ANOVA() function to measure the interaction effect between the two fixed effects on prepositional modification:

full model: feature 1 ~ revision*course

reduced model: feature 1 ~ revision + course

The 'full model' includes both the main effects of Revision and Course (represented by "~") as well as their interaction effect. The interaction effect captures how the influence of one factor (e.g., Revision) on the outcome (e.g., adjectival modification) might depend on the level of another factor (e.g., Course). On the other hand, the 'reduced model' includes only the main effects of Revision and Course (represented by "+") but excludes their interaction term. By comparing these two models, I aimed to determine whether the effect of "revision" on prepositional modification depends on the "course" level.

5.2.3 Results

The frequency data of two phrase indices from TAASSC in 326 PELIC texts were counted and saved in txt. file format, which was imported to the R environment for further analysis. The scatterplot in Figure 5-2 was plotted using the ggplot function in the 'hrbrthemes' package (Rudis, 2020). The scatter plot shows the broad picture of the distribution of the two indices in PELIC revision data. The coloured dots represent the text points by the three revision indexes.

Figure 5-2. Correlation between adjectival modification and prepositional modification indices

This distribution of the two indices shows that their frequency positively correlates with each other. However, their distribution of the frequency indices did not show a clear correlation with the versions of each text, as shown in the three coloured bars representing the linear line of each revision index.

A linear mixed-effects model was run to examine the effect of revision on the linguistic variables of interest. This model did not consider the effects of L1 background and L2 proficiency to keep the model parsimonious; the individual students are nested within these factors. Table 5-3 shows the MEM model structure.

Table 5-3. Linear Mixed-effect model used in the analysis

Dependent variable	**Fixed effect**	**Random effect[1)]**
occurrence of adjectival/prepositional dependents per nominal	Version (3 levels: 1st ~ 3rd version) The course of study (4 levels: Course 2~Course 5)	Student (54 individual students) Question type (112 Types of questions)

This section reports the results of the linear mixed model analyses of the relationship between revision and frequencies of two syntactic complexity measures in the R environment (R Core Team, 2020) using the lme4 function (Bates, Maechler, Bolker & Walker, 2015). The two syntactic complexity measures, adjectival and prepositional modification, were chosen to compare the result with that from the previous studies (e.g., Larsson & Kaatari, 2020). The statistical processing procedure followed Winter (2013). First, the index of text revision and students' course of study were entered into the mixed effect model as fixed effects (without interaction term). As random effects, I had intercepts for students and question items, as well as by-student and by-question random slopes for the effect of text revision. Then, P-values were obtained by likelihood ratio tests of the full model with the revision effect against the model without the effect.

Table 5-4. MEM of adjectival modifications by versions and courses

Fixed effects	value	Std.Error	t value[16]
Intercept	0.253000	0.019471	12.993
version	-0.001446	0.003861	-0.375
course	-0.009265	0.007744	-1.196
Random effects	Variance	Std.Dev.	
question	0.004818	0.06941	
student	0.003568	0.05973	
Residual	0.00181	0.04255	

[16] The anova result between a null model and the hypothesised model : Pr(>Chisq) : >0.05

Table 5-5. MEM of prepositional modifications by versions and courses

Fixed effects	value	Std.Error	t value [17]
Intercept	0.119560	0.015721	7.605
version	0.006722	0.003080	2.183
course	0.012640	0.006331	1.997
Random effects	**Variance**	**Std.Dev.**	
question	0.003681	0.06067	
student	0.001875	0.0433	
Residual	0.001151	0.03393	

Table 5-4 shows that the revision effect did not prove to be significant on the frequency of adjectival modification ($\chi^2(1)=0.14$, $p< 0.70$), decreasing it by -0.001 words ± 0.003 (standard errors). On the other hand, the modest significance on the frequency of prepositional modification is shown in Table 5-5 ($\chi^2(1)=4.71$, $p< 0.03$), increasing it by 0.007 words ± 0.003 (standard errors). The ANOVA result comparing the significance between the two models, with revision effect and null effect, showed that the variable 'revision' affects the frequency of prepositions per noun phrase and is not likely to be by chance. However, the increase was minimal.

Overall, these two indices did not show adequate evidence of being indicative of syntactic development, unlike previous reports (e.g., Larsson & Kaatari, 2020). The p-value (=0.03) of the difference between the null model and interaction model in Table 5-5 shows that the revision process and course of study are significantly interdependent. Albeit the magnitude gauged by their slopes is minimal (+0.007), the prepositional modifications increased across revised versions.

[17] The anova result between a null model and the hypothesised model : Pr(>Chisq) : 0.03

5.2.4 Conclusions

This study used a linear mixed-effects model to explore the relationship between the revision process and adjectival and prepositional modifications in L2 student writing. The result indicates that one of these two indices showed a moderately significant correlation with draft revision in this study. One of the possible explanations for the moderate significance may be that revised text groups are too similar because texts were produced by the same student at relatively short intervals (within the same academic semesters). Another reason may be related to the fact that only two of all the indices were selected, while any of the other indices are potentially indicative of writing quality. As the SC features explored in this pilot study did not provide meaningful linguistic insight regarding the overarching research question (elucidating the communicative functions associated with the linguistic features), no further analysis was pursued. While the current analysis focused on the overall frequency of prepositional modification, the variations within these constructions also warrant further investigation. These variations, particularly those associated with specific communicative functions, can offer valuable insights into L2 writing development.

Section 5.3 delves deeper into this aspect by analysing the PELIC longitudinal corpus data. This analysis goes beyond the elaboration of prepositional constructions (examined here) and focuses on the prepositions themselves. By examining the frequency of the prepositional constructions over time, I aim to determine whether and how prepositional use evolves as L2 writers develop their skills.

5.3. Pilot Study 2: Mixed-effect models of preposition distribution in L2 longitudinal writing drawing on a Bayesian approach

5.3.1. Introduction

This second pilot study is motivated by the result of the multidimensional analyses in Chapter 4. Pilot MDA study 2 (Section 4.4) found the distinction between of-phrases (loaded in Dimension 2) and other prepositional phrases (loaded in Dimension 1). This leads to expecting potential functional differences between of-phrases and other prepositional constructions in context, which invites further investigation.

The final MDA study (Section 4.5) also found linguistic features with high Dimension 1 (D1) scores in the L2 texts. These features include spoken elements like second-person pronouns, contractions, mental verbs, and to-infinitive dependents, which engage the reader and make the writing interactive, contributing to a persuasive purpose. Conversely, texts with negative D1 features exhibited descriptive tones, providing abstract conceptual illustrations or factual descriptions, and frequently employed phrasal features such as nouns, prepositional phrases, and 'of' prepositions at a clausal level.

The analysis of the Dimension 1 scores confirms the hypothesis that spoken features emerge earlier in writing development, while more literate features emerge later. Specifically, there is an overall decrease in the frequency of positively loaded features (e.g., to-infinitives and mental verbs) over time, replaced by negatively loaded features (e.g., promotional phrases modifying clauses and noun phrases). Mixed-effects models revealed a significant increase in phrasal complexity features, particularly in the final time point (Course 5), supporting previous longitudinal studies' findings (e.g., Gray et al., 2019). This suggests a growing frequency of phrasal complexity features associated with literate and informational writing over time, partially confirming the prediction that phrasal and informational density features gain prominence.

Unexpectedly, there was a sharp increase in the Dimension 1 score in Course 2, indicating relatively higher usage of features associated with personal involvement compared to other time points. This nonlinear developmental trajectory is difficult to interpret based on a unitary theory-informed hypothesis (e.g., the usage-based perspective on syntactic development or formal SC elaboration). The findings from the analysis in Chapter 4 reveal the significant role of writing topics and L2 level in explaining variance, while the L1 background variable does not significantly influence the results. To further analyse the data, this chapter employs a mixed-effects model, which differs from the frequency model used previously.

While the analysis reported in Chapter 4 does not trace individual constructions' frequencies, some of the identified constructions need further investigation. As seen in Chapter 4, *verbs controlling to-infinitives* are likely to be frequently used in the PELIC longitudinal writing, considering the high correlation between the mental verbs and to-infinitives on Dimension 1. This result indicates the high likelihood of their co-occurrence for a shared communicative function. However, their co-occurrences are not directly explored in this chapter. The prepositional phrases, which were significant in the MD analysis in Chapter 4, are the focus of this section. This focus on prepositions is due to the perceived importance of observing the emergence of the prepositional construction for enquires regarding development in academic writing. Biber et al. (2021a) argue that infinitives controlled by verbs should appear earlier (stage 3) than prepositions as postmodifiers (stages 4 and 5). This specific frequency distribution information can help predict the development of prepositions as postmodifiers in L2 learner writing, providing insights into the constructional aspects of writing literacy. While this type of information helps predict the relative order of emergence of the features, it does not necessarily indicate a decline in the frequencies of early emerging features. Rather, early acquired and therefore frequently used constructions are likely to be used in later stages when appropriate.

The study discussed in this section utilises a mixed-effects model drawing on Bayesian theory. The strengths and suitability of using this method for this study are discussed in Section 5.3.2.

5.3.2 Method

5.3.2.1 Linguistic Variables

The PELIC longitudinal data is collected from a specific situational context, and therefore it may benefit from using a statistical model that allows more theory-informed inferences that do not cohere with random distribution. In the context of L2 English writing, sub-grouping individuals based on L1 background and English proficiency allows for investigating potential differences in results. In this chapter, the focus is on descriptive and inference analyses of key constructions and syntactic features identified in previous analyses.

The key dependent variable in this analysis is the frequency of prepositional phrases functioning as phrasal modifiers. These phrases consist of a preposition (e.g., *in, on*, or *with*) followed by a noun phrase (e.g., *the book on the table*). I used the Stanford parser via TAASSC to identify these prepositional phrases in the writing samples.

The key independent variables are time and L2 proficiency. I aim to understand how the frequency of these prepositional phrases changes over time as students develop their L2 writing skills (time) and how it relates to their overall L2 proficiency level.

The analysis employs a mixed-effects model, similar to the final MDA study described in Section 4.5. This model accounts for potential random effects, such as individual student variation in their writing styles and the influence of different writing topics. By considering these random effects, we can obtain a more accurate picture of the relationship between prepositional phrase frequency, time, and L2 proficiency.

5.3.2.2 Bayesian MEM

There is a growing interest in using linear mixed-effects models, also known as hierarchical or multilevel models, for dealing with hierarchical data structures and accounting for correlated observations within the same grouping levels (Gries, 2015b). Several advantages of using this method are discussed in the following paragraphs.

A Bayesian approach is particularly valuable when there is a hierarchical data-generating process (Winter & Bürkner, 2021). Hierarchical models incorporate the idea that inferences about one quantity affect inferences about another. Unlike general linear models, which assume independent observations, hierarchical models group observations into clusters and consider shared attributes and similarities among the clusters. In a more practical term, the *brm* package (Bürkner, 2017) using a Bayesian theory as a modelling parameter incorporates different data distribution more flexibly, (e.g., Poisson distribution for count variables), it is suitable for exploring datasets that are not randomly sampled by nature.

Bayesian inference offers several advantages for the current study. It is particularly suited for handling hierarchical data, where variables are interdependent. Bayesian inference allows for adjustments of assumptions based on the posterior distribution, rather than relying solely on confirming or rejecting assumptions. Moreover, it enables the incorporation of prior knowledge into the prediction of data distribution. Priors, which capture assumptions about plausible parameter values, can be informed by domain knowledge and existing literature (Winter & Bürkner, 2021).

By utilizing simulation-based methods like Markov Chain Monte Carlo (MCMC), Bayesian inference provides flexibility in modelling complex hierarchical structures such as random intercepts and slopes. This is especially beneficial for small or complex datasets, where frequentist methods are limited to relying on asymptotic approximations (Hansen, 2022). Additionally, Bayesian inference allows researchers to incorporate prior knowledge, enhancing estimation accuracy and reducing the impact of noisy or incomplete data (Gelman, Carlin, Stern, & Rubin, 1995). In contrast, frequentist methods do not typically allow for the integration of prior knowledge. McElreath (2020) also hghlights the difference between point estimates and confidence intervals in frequentist approaches compared to posterior distiributons in Bayesian inference. Bayesian inference naturally provides estimates of uncertainty in model parameters and predictions through posterior distributions, whereas relying on point estimates and confidence intervals may not fully capture the range of possible values.

The current analyses employ several regression models, including Poisson and negative binomial regression. The Poisson distribution models the count of event occurrences in a fixed period, assuming equal mean and variance. Poisson regression, a generalized linear model, uses the Poisson distribution as the response variable's probability distribution. As the frequency data of prepositions in written texts may not follow a normal distribution, Poisson regression is an appropriate choice. Additionally, a binomial distribution is employed to handle overdispersion, by incorporating a shape parameter (Winter & Bürkner, 2021).

To begin, a prior distribution for the model is specified. In order to estimate the effect of time (as a continuous variable) on preposition frequencies, a prior for the slope parameter (beta) is set using the prior argument in the *brm* function. In this case, a normal distribution with a mean of 0 and a standard deviation of 1 is utilized. Priors for other model parameters, such as the intercept or the variance of the random effects, are also specified. Subsequently, the model is fitted with the previously defined formula and priors. After fitting the model, posterior probabilities for each parameter are obtained using the summary function. This allows us to assess how different each parameter, especially the effect of time on preposition frequencies, is from zero. Additionally, the effect size for the time variable is calculated to determine its practical significance. One commonly used effect size measure for mixed effects models is the conditional R-squared, which quantifies the proportion of variance explained by the fixed and random effects in the model. Furthermore, alternative hypotheses are formulated to test against the null hypothesis, which posits that the effect of time on preposition frequencies is zero.

Bayesian models allow for the incorporation of prior assumptions into data distribution estimation. Priors represent plausible parameter values based on domain knowledge, often drawn from existing literature (Lemoine, 2019). Weakly informative priors are more conservative choice due to their role in reducing type 1 error rates with small sample sizes (Lemoine, 2019, p.917).

For random effects, a hierarchical prior is recommended, which allows for shrinkage towards a shared mean (McElearth, 2020). The specific form of the hierarchical prior depends on the model structure, with a common choice being a normal distribution with a mean of zero and a standard deviation (tau). The prior distribution for tau, such as a half-normal distribution with mean zero and standard deviation of 1 or 2, is often used.

As for general principles of the Bayesian approach, Norouzian et al. (2019) propose the following three steps: (1) specifying alternative hypotheses to explore various possibilities for the effect size, (2) obtaining a comparative measure (Bayes factor) by comparing likelihoods under alternative and null hypotheses, and (3) interpreting the comparative measure using a classificatory scale for evaluating the support of data for different hypotheses.

Lastly, this study uses normalized frequency data to address uneven text lengths and enable feature comparison across sub-corpora. Given that Poisson regression requires non-negative integer counts as input data, the normalized frequencies are rounded to zero decimal points.

5.3.3. Results & Discussion

This section presents the analysis to make inferences about the frequency variation of prepositions over time considering the influence of writer-related and situational variables. I used MAT to extract the frequency measure of prepositions, following the method used in Chapter 4. As MAT does not distinguish the frequencies of the prepositions by their syntactic functions, the analysis results based on this measure were interpreted based on this limitation. Four regression models are conducted, starting with a simple Poisson regression model and progressing to two mixed models with a Poisson distribution, and one mixed model with a negative binomial distribution. This progression is for exploration purposes, and the decision-making relevant to the process of expanding the model will be discussed in later drafts, focusing on posterior predictive simulations (Kruschke, 2013).

5.3.3.1 Simple Poisson regression model

A simple Poisson regression model is fitted to the data using the *brm* function in the R Studio (version 2023.06.0+421). This model examines prepositions as the response variable and five courses as the time predictor. Uninformative prior distributions are used for the parameters in this model, and the posterior distributions are estimated using Bayesian methods. Table 5-6 presents the results of the Poisson regression model predicting preposition frequency across the five courses.

Table 5-6. Poisson regression predicting preposition frequency across five courses (using the brm function)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	2.34	0.02	2.31	2.38	1	3640	2795
Course 2	-0.05	0.02	-0.1	-0.01	1	3862	3035
Course 3	0	0.02	-0.05	0.05	1	3787	2878
Course 4	0.06	0.05	-0.04	0.17	1	3865	2705
Course 5	-0.03	0.12	-0.28	0.2	1	4138	2708

The 95% credible intervals (l-95% CI and u-95% CI) in Table 5-6 indicate that the slopes of the predictor variables, only Course 2 is significant while the rest of Courses 3, 4, and 5, are not significant in terms of their increase or decrease compared to Course 1.

Another Poisson regression model is presented in Table 5-7, utilizing the same predictor variable "course" and dependent variable "preposition," but using the "*glm*" function in R, which employs maximum likelihood estimation (MLE) instead of Bayesian methods. Table 5-7 showcases the results of the Poisson regression model predicting preposition frequency across the five courses.

Table 5-7. Poisson regression predicting preposition frequency across five Courses (using the glm function)

	(Intercept)	course2	course3	course4	course5
Coefficients:	2.341454	-0.0539327	-0.0007969	0.0637853	-0.0246842
Degrees of freedom:		1114 Total (i.e. Null)		1110 Residual	
Null deviance:		827.5			
Residual deviance:		817.4		AIC:	5445

Both models yield similar results, but their interpretation and estimation methods differ. In both Tables 5-6 and 5-7, it is observed that the estimated effect sizes for Course 2, Course 3, and Course 5 are negative, suggesting that students in their second, third, and fifth semesters tend to use fewer prepositions compared to their first semester (the reference level), on average, except for the fourth semester. However, the standard errors are relatively large for some of these estimates, indicating that the differences between courses may not be statistically significant (Norouzian et al., 2019, p.253).

Additionally, this study considers the alternative hypothesis and incorporates a weighting scheme for effect sizes. The alternative sizes of effects are based on previous findings and are specified within the range of -1 and +1, as commonly expected in L2 research (Plonsky & Oswald, 2014). A Cauchy distribution with a scale of ".707" is employed to weight effect sizes, with the range between -1 and +1 receiving more weight (Ly, Verhagen, & Wagenmakers, 2016; Morey, Romeijn, & Rouder, 2016).

Table 5-8 provides the conditional effects of the simple Poisson regression model presented in Table 5-6, displaying the estimated values of the dependent variable (preposition) for each level of the predictor variable (Course), along with their standard errors and the lower and upper bounds of the 95% credible intervals.

Table 5-8. Conditional effects of simple Poisson regression model

	Course	preposition	estimate	SE	lower	upper
1	course 1	10.21525	10.398367	0.1838778	10.048221	10.76377
2	course 2	10.21525	9.846011	0.1555127	9.553583	10.15611
3	course 3	10.21525	10.386354	0.1739242	10.050412	10.73877
4	course 4	10.21525	11.075836	0.5484054	10.020915	12.20175
5	course 5	10.21525	10.07137	1.159591	7.891595	12.72803

The estimated effect sizes and standard errors differ across the three models presented so far due to the different methods (MLE vs. Bayesian) and software packages used (R vs. Stan). Table 5-8 offers additional insights into the conditional effects of the model, presenting the estimated preposition frequencies for each level of "Course" at a particular value of "prepositions" (10.21525). The estimates suggest that Course 2 has a lower frequency compared to other courses, while Course 4 has a higher frequency. This is additional evidence of the nonlinearity observed in previous analyses; as we would expect to see the gradual increase of the prepositional constructions over time, the decrease at Course 2 requires further inspection.

However, the wide confidence intervals for these estimates in some of the Courses (especially Courses 4 and 5) indicate a significant level of uncertainty in the results (see Figure 5-3). Figure 5-3 visually represents the estimated number of prepositions for each level of the predictor variable course, with error bars indicating the lower and upper bounds of the 95% credible intervals. This visualization allows for easy comparison of the estimated preposition frequencies across different courses.

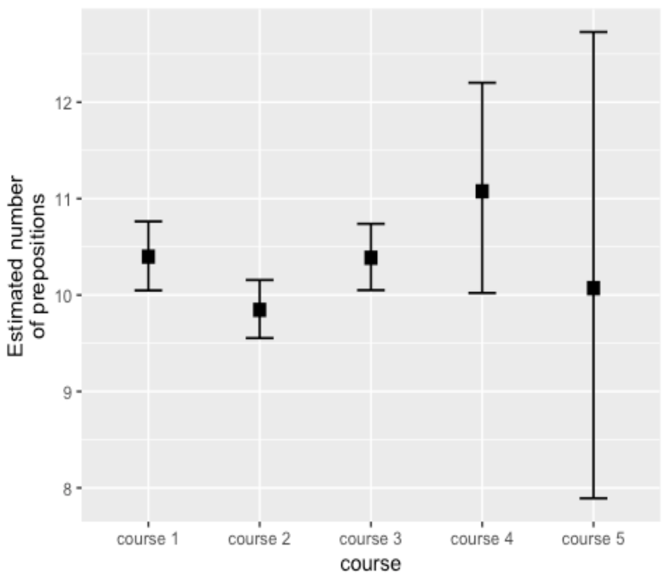

Figure 5-3. Conditional effects of simple Poisson regression model

5.3.3.2 Poisson regression models with random effects

To account for variation in preposition frequencies across time points, two Poisson regression models with random effects are fitted using the *brm* function. These models introduce random effects and priors into the analysis. The first model incorporates a random effect for individual students, capturing the variation in preposition usage between different students. The second model includes a random effect for question types, accounting for the variation in preposition usage across different types of questions.

Table 5-9 presents the results of the mixed-effects Poisson regression model. It estimates the conditional effect of the course on preposition counts while considering the grouping structure of the data based on individual students. Table 5-9 is divided into two sections: Group-Level Effects and Population-Level Effects.

Table 5-9. Mixed effects Poisson regression model predicting preposition counts over time

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.09	0.01	0.07	0.12	1	1371	2122
Population-Level Effects:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	2.34	0.02	2.3	2.37	1	4818	3650
course2	-0.04	0.02	-0.09	0.01	1	5416	3402
course3	0	0.02	-0.05	0.05	1	5834	3314
course4	0.07	0.06	-0.04	0.18	1	6349	3050
course5	0.06	0.13	-0.2	0.3	1	5777	3151

The Group-Level Effects provide estimates for the random intercept parameter, which quantifies the variation in preposition counts between different students. The estimated standard deviation of the intercept is 0.09, with a 95% credible interval ranging from 0.07 to 0.12. This suggests that there is some variability in preposition counts among individual students, although the magnitude of this variation is relatively small.

The Population-Level Effects present estimates for the fixed effects of the predictor variable “course” on preposition counts. The estimates are reported as log-rate ratios, where a value of 0 indicates no effect, and positive or negative values indicate an increase or decrease in the expected count of prepositions, respectively. For example, the estimate for Course 2 is -0.04, indicating a decrease of 0.04 units in the expected log count of prepositions compared to Course 1.

Table 5-10 represents the result of testing a hypothesis that the expected value of preposition count in Course 2 is equal to 0 and is modelled using an exponential function of the intercept and the coefficient for “Course2.” Table 5-10 includes the estimated coefficient for “Course2” (9.33), the estimated standard error (1), and the lower and upper bounds of a 90% confidence interval for the coefficient (CI.Lower and CI.Upper, respectively). The Bayes factor, which quantifies the support for the hypothesis, is not calculated in this case as the hypothesis is a point null hypothesis.

Table 5-10. Hypothesis Tests for model predicting course effect on preposition frequency

Hypothesis	Estimate	Est.Error	CI.Lower[18]	CI.Upper	Evid.Ratio	Post.Prob	Star[19]
Intercept + course2 = 0	9.33	1	7.53	11.39	NA	NA	*[20]

Furthermore, the "Post.Prob" column presents the posterior probability of the null hypothesis being true given the data. In this instance, the posterior probability cannot be calculated as the prior probability was not specified. Lastly, the "Star" column indicates whether the hypothesis is supported by the data based on a 95% confidence interval. In this case, the asterisk (*) indicates that the null hypothesis is rejected at the 95% level of significance because the value being tested against (i.e., the expected value of "prep" for Course 2 being 0) falls outside the 95%-CI.

5.3.3.3 A negative binomial model

The analyses discussed in this subsection draw on a binomial distribution, which is employed to handle overdispersion. This is done by incorporating a shape parameter (Winter & Bürkner, 2021). In the Poisson distribution, the mean is equal to the variance (Winter & Bürkner, 2021, p.11). The negative binomial distribution differs from the Poisson distribution by including an additional parameter, often referred to as 'phi' or 'shape' in the *brms* package, which directly estimates dispersion from the data.

In the initial step, a prior distribution was specified for the model. For the effect of time on preposition frequencies, the slope parameter (beta) was assigned a prior using the prior argument in the *brm* function, with a normal distribution having a mean of 0 and a standard deviation of 1. Priors for other model parameters, such as the intercept or the variance of the random effects, were

[18] 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses
[19] Posterior probabilities of point hypotheses assume equal prior probabilities.
[20] '*': For one-sided hypotheses, the posterior probability exceeds 95%; for two-sided hypotheses, the value tested against lies outside the 95%-CI

also specified. Subsequently, the model was fitted using the formula used in the previous stage and the specified priors. Posterior probabilities for each parameter were obtained using the summary function, providing evidence of how different each parameter, particularly the effect of time on preposition frequencies, is from zero.

To interpret the practical significance of the time variable, the effect size was calculated using the conditional R-squared, a commonly used measure for mixed-effects models that quantifies the proportion of variance explained by the fixed and random effects in the model. Additionally, alternative hypotheses were formulated to test against the null hypothesis that the effect of time on preposition frequencies is zero, aiming to determine if there are differences in preposition frequencies across courses. To evaluate the model fit and assess whether the negative binomial model provides a better fit to the data, leave-one-out cross-validation (LOO-CV) was performed using the Poisson model (mixed_mdl) and the Negative Binomial model, following Vehtari et al. (2017).

In Bayesian analysis, prior distributions play a crucial role in incorporating prior knowledge or assumptions about model parameters. Prior distributions represent the researcher's beliefs about the plausible range of parameter values before observing the data. The choice of prior distribution can significantly influence the posterior distribution of the model parameters and consequently affect the analysis results.

The Poisson regression models and the negative binomial model did not specify priors, so the default priors of the brms package were used. For Poisson models in brms, the default prior for the intercept and fixed effects is a Student's t-distribution with degrees of freedom (df) set to 3 (Bürkner, 2017). For the random effects, the default prior is a normal distribution with mean 0 and standard deviation 1 (Bürkner, 2017). For the Poisson family, the prior for the dispersion parameter is a half-normal distribution with a scale parameter of 1 (Bürkner, 2017). The default priors in brms are relatively weak, meaning that they allow for a wide range of parameter values. This was appropriate choice considering little prior knowledge about the parameters and for the purpose of avoiding strong prior assumptions that might bias the results. However, the prior specification for a normal distribution exemplified in Figure 5-4 could be used to specify priors in the next analyses.

To assess the model's goodness of fit, Figure 5-4 presents a plot for model checking of a negative binomial regression model using posterior predictive checks. It compares the observed data with data simulated from the posterior distribution of the model parameters, using the empirical cumulative distribution function (ECDF). A good fit is indicated when the observed data fall within the range of simulated data, and the ECDF curves closely match each other.

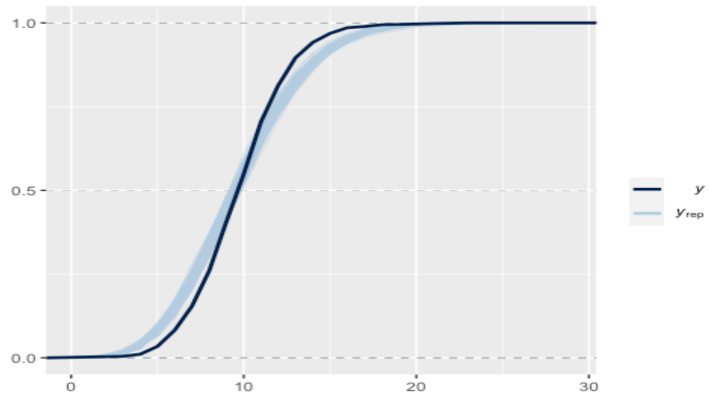

Figure 5-4. Plot for model checking of a negative binomial regression model using the posterior predictive checks**

To evaluate the hypothesis regarding the coefficient for the interaction between course 1 and course 2, a hypothesis test was conducted on the negative binomial model. Table 5-11 displays the results of the hypothesis test, indicating that the posterior probability of the hypothesis (course2 < 0) is 0.91. This suggests a 91% chance that the true coefficient value is negative. Although the evidence is moderate, it does not provide strong support to reject the null hypothesis.

Table 5-11. Hypothesis Tests for class b

Hypothesis	Estimate	Est.Error	CI.Lower[21]	CI.Upper	Evid.Ratio	Post.Prob[22]
course2 < 0	-0.03	0.02	-0.07	0.01	9.62	0.91

Figure 5-5 depicts a density plot of the posterior distribution of the slope

[21] 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses
[22] Posterior probabilities of point hypotheses assume equal prior probabilities.

parameter for Course 2. The plot indicates a very small but moderately significant distribution for the slope, with the null value (representing no difference between the conditions) indicated by a vertical dashed line at x = 0.

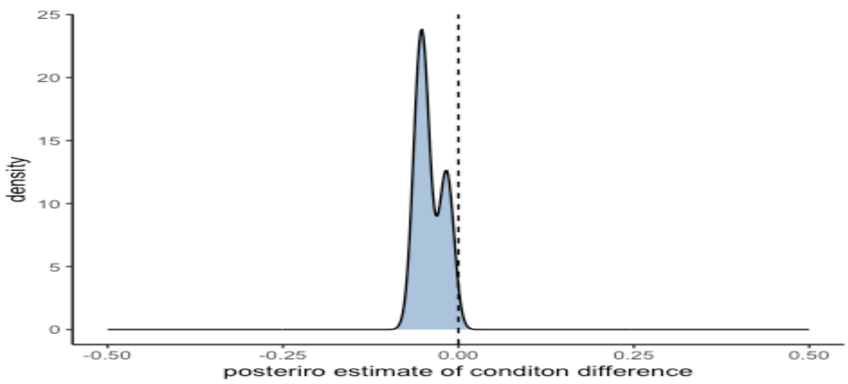

Figure 5-5. Plot of the posterior distribution of the slope of Course 2

Consequently, the analyses on preposition use in the data indicate some moderate significance in the variation over time. However, the significance is not strong enough to definitively confirm or reject an increase or decrease in the variable over time.

5.3.4 Conclusions & Next Steps

The analysis of prepositional phrase distribution in this study aimed to contribute to a deeper understanding of L2 syntactic development. I have expected to observe more frequent use of prepositions across Courses. This expectation was based on the previous findings that showed the use of the preposition 'of' is often associated with higher levels of language proficiency (e.g., Pérez-Paredes and Díez-Bedmar, 2018). However, there have been also findings that suggest low-proficiency students relied on using postmodifying PP (of) more than high-proficiency students (e.g., Lan et al., 2019), so the findings needed to be interpreted within the context of the data collected.

While the findings from this study indicated some variation in preposition use over time, the results were inconclusive regarding a clear-cut developmental trajectory. This aligns with the complexities highlighted by Pérez-Paredes and Díez-Bedmar (2018) in their research on the preposition 'of'. Additionally, the study partially supports Lan et al.'s (2019) findings regarding the relationship between prepositional phrase use and proficiency levels. However, this interpretation is limited in that the prepositional constructions are not as distinguished as in the previous studies, particularly for prepositional functions such as postmodifying prepositions.

The subsequent section (5.4) focuses on noun and adjective complement clauses, building upon the insights of Biber et al. (2021) regarding their prevalence in academic writing and their association with epistemic and evaluative stance functions. To overcome the limitations of the present study and to gain a more nuanced understanding of the relationship between linguistic use and L2 lexico-grammatical development, the accuracy of the structural and syntactic properties of these constructions are checked and reported. Additionally, the prior distributions within the mixed-effects models is reconsidered for more accurate and robust findings. By examining the frequency and distribution of these constructions, I aim to identify patterns related to L2 writing development and explore their functional dimensions.

5.4 Final study: predicting variation of noun-that clauses and adjective-that clauses constructions

The first pilot analysis focused on phrasal features, specifically examining the frequency variations of two constructions identified in Chapter 4: adjectival modification and prepositional modification of noun phrases. These pilot analyses aimed to investigate the development of phrasal complexity; a characteristic often associated with academic writing. Second pilot analysis narrowed down the scope to prepositional phrases, directly focusing on all prepositional phrases rather than examining the elaboration by nouns. Similar to the first pilot analysis, this analysis focused on phrasal features and examined a construction commonly found in academic registers.

This final analysis shifts the focus from general phrasal complexity to constructions that provide insights into the writer's attitude or viewpoint, by considering constructions to express stance. I utilized the frequency measures for nouns and adjectives controlling that-complement clauses (THNC and THAC, respectively) established in Chapter 4. THNCs are nouns followed by a "that" clause (e.g., *The fact that she arrived late surprised everyone*). ATHCs are adjectives followed by a *that*-clause (e.g., *He was happy that he passed the exam)*. These constructions can directly express the writer's stance through the choice of nouns and adjectives used concerning the *that*-complement clause.

This chapter aims to describe the frequency patterns of two syntactic constructions to complement the MD analysis discussed in Chapter 4 to evaluate the hypothesis of syntactic development in L2 learner writing: noun-clause complements and adjective-clause complements. The motivation for this selection of features is based on the observation that analysing multiple features' co-occurrences complicated identifying unique patterns of key syntactic constructions. While this functional approach to syntactic complexity has provided explanations for L2 syntactic complexity development in Chapter 4, it lacked information about frequency patterns of individual syntactic constructions of interest.

As discussed in Chapter 4, these two constructions were considered markers of grammatical complexity, as they are syntactically defined forms that are often referred to as frames for stance (Biber, 2006). These two constructions, along with the verb-that clauses analysed in Chapter 5, are associated with stance functions. Predicative adjectives are used to convey evaluative or emotive stance in conversations, typically without complements (Biber et al., 2021b, p.515). On the other hand, academic prose uses predicative adjectives for epistemic or evaluative stance, often followed by complements including to-infinitives and that-clauses (Biber et al., 2021, p.516). Moreover, many of these stance adjectives have noun counterparts (e.g., *importance*, *certainty*; Biber et al., 2021b, p.962).

As discussed in Chapter 2, the stance function is prevalent in most registers (Biber, 2021), and the typical difference associated with academic written register, particularly with student written register, in comparison to other registers, renders it worth investigation. Despite its recognised importance in academic writing functions, Wray (2018) notes that L2 learners less use these constructions compared to verb-based bundles. In that regard, it is an important area for further research in L2 writing development. However, they were lost during the feature reduction process in the MD analysis in Chapter 4, and therefore they are separately discussed in this chapter.

The main linguistic variable of interest in this section is the frequency of two key syntactic constructions, particularly those related expressing stance. Stance refers to the writer's attitude or viewpoint towards the information being presented. This section focuses on how the frequency of these stance-related constructions changes over time as L2 learners develop their writing skills.

This section explores the following questions about the two syntactic constructions: How do the frequencies of the two constructions (noun-that SV; adjective-that SV) vary over time? To that aim, the mixed-effects models explore the frequency patterns of two key linguistic features over time: that-adjective clauses and that-noun-clauses. The frequency data is based on the MAT-tagged 1,115 texts and a mixed-effects model is run for each feature drawing on a Bayesian approach and integrating priors based on information from the Pilot study 2 (Section 5.3). Once the analysis is reported, the significance of the patterns regarding the hypothesis from previous literature is discussed, which will be synthesised and summarised in Section 5.4.4.

5.4.1 Method

Two linguistic features tagged by MAT are THACs and THNCs. For the noun-controlled that-clauses, I used the manually re-tagged texts to two features of that relative clause and that-noun complements, based on the MAT-tagged texts, as MAT does not distinguish these two constructions (see Section 3.4.3 for a discussion about the procedures and reports on the tagging accuracy regarding the manual tagging). So only that-noun complement constructions were used while that-relatives were excluded in this analysis. As for the random effects, individual students and writing topics were entered as random effects as well as the random slope of the course (time) effects for individual writers. These choices were based on the previous mixed-effects models of dimensions scores and individual variables, as reported in Chapters 4 and 6.

Two linguistic features tagged by MAT were considered for analysis: THAC and THNC. As discussed in Section 3.4.3, I manually re-tagged the texts to differentiate these two constructions. However, for this particular analysis, I decided to exclude that-relative clauses and only include that-noun complement constructions (THNC). This decision aligns with the research objective of investigating stance expression. That-noun complements often play a role in conveying the writer's attitude or viewpoint towards the information being presented. That-relative clauses, on the other hand, primarily function to provide additional information about a noun and might not directly contribute to stance expression.

I also included individual students, writing topics, and the random slope of the course (time) effects for individual writers as random effects in the mixed-effects model. These choices were informed by the previous mixed-effects models of dimensions scores and individual variables, as reported in Chapters 4 and 6.

As Poisson distribution deals with count variables, the raw frequency data of these two linguistic variables were used as input data. In other words, the normed frequency used in the MD analysis, as reported in Chapter 4, was unnormalized for this analysis. Instead, an offset term was included for running mixed-effects models to control for differences in the length of texts. As a Poisson family assumes that the response variables represent the average rate of occurrence, the offset term adjusted the results relative to the number of tokens in each text by dividing it by the log-transformed tokens. Finally, the prior was specified as a normal distribution with a mean of 0 and a standard deviation of 0.5, as the normal distribution is a common choice for prior distributions due to its flexibility and ease of interpretation. This specified prior represents a narrower prior compared to the default prior in the 'brms' package, which is the Student's t-distribution (Student, 1908) with four degrees of freedom. This choice also reflected no assumption of strong prior belief in any particular direction of effect and allowed the data to have a stronger influence on the posterior distribution, as the prior information does not heavily constrain the analysis. This approach was considered appropriate because the analysis was meant to estimate the parameters of interest rather than feeding a strong hypothesis for checking.

5.4.2 Results 1: predicting variation of adjective-that clauses

Table 5-12 and Figure 5-6 show the descriptive statistics of adjective-that clauses across five courses. The average frequencies tend to increase over time, but the frequencies overall are very low, all below one occurrence per text across all sub-corpora. Albeit very scarce in the corpus, the average frequencies increase over time.

Table 5-12. The average frequencies of adjective-that clauses across five courses

	course 1	course 2	course 3	course 4	course 5	Total
Average raw frequency per text	0.10	0.19	0.23	0.30	0.57	0.18

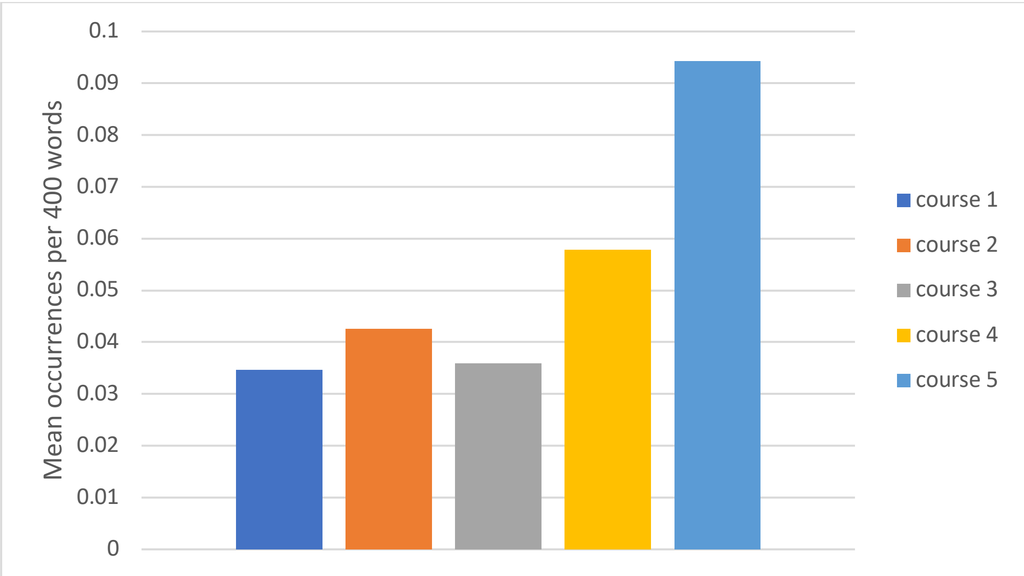

Figure 5-6. The average frequencies of adjective-that clauses across five courses

Table 5-13 shows the result of a mixed-effects model of adjective-that clauses across the five sub-corpora.

Table 5-13. The Mixed-effects model of adjective-that clauses across five courses

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.49	0.26	0.04	1	1	1267	3247
sd(Course2)	0.58	0.31	0.04	1.21	1	1034	3381

sd(Course3)	1.09	0.33	0.4	1.72	1	1034	2323
sd(Course4)	0.69	0.5	0.03	1.88	1	4008	6382
sd(Course5)	1.59	1.28	0.07	4.96	1	4227	4696
cor(Intercept, Course2)	-0.11	0.41	-0.79	0.71	1	2816	5613
cor(Intercept, Course3)	0.17	0.36	-0.53	0.8	1	1048	1558
cor(Course2, Course3)	0.26	0.37	-0.59	0.82	1	1159	2790
cor(Intercept, Course4)	-0.04	0.4	-0.75	0.73	1	8943	10893
cor(Course2, Course4)	0.01	0.4	-0.74	0.75	1	8323	10786
cor(Course3, Course4)	0.19	0.38	-0.62	0.81	1	7267	10515
cor(Intercept, Course5)	-0.17	0.4	-0.84	0.64	1	7308	8848
cor(Course2, Course5)	-0.01	0.41	-0.76	0.75	1	9550	12154
cor(Course3, Course5)	-0.18	0.38	-0.83	0.61	1	13109	11623
cor(Course4, Course5)	-0.1	0.4	-0.8	0.68	1	10648	12373
Population-Level Effects:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-7.94	0.22	-8.39	-7.55	1	4989	6456
Course2	0.03	0.27	-0.51	0.55	1	5527	10169
Course3	-0.56	0.36	-1.31	0.12	1	4424	6819
Course4	0.1	0.53	-1.1	1	1	9190	8194
course5	0.2	1.32	-3.23	1.97	1	5814	3154

To assess the model's goodness of fit, Figure 5-7 presents a plot for model checking of a negative binomial regression model using posterior predictive checks. As seen in Figure 5-7, the observed data represented by *y* shows a relatively close match to *yrep*, which indicates the range of simulated data, and the ECDF curves tend to closely match each other.

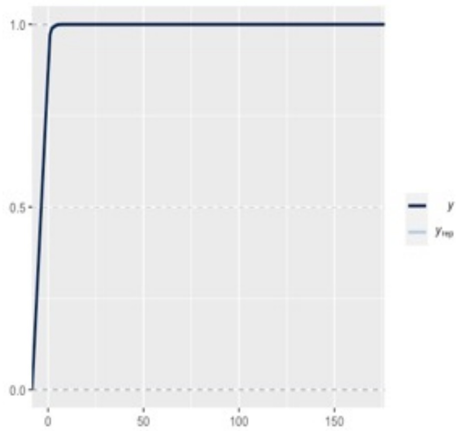

Figure 5-7. plot for model checking of a negative binomial regression model using the posterior predictive checks

Figure 5-7 indicate that the estimation of the model converges with the sample data. However, it also shows that the converging point is around zero on the x-axis. This means that the adjective-that clause constructions are very infrequent in the corpus overall, most of the counts fall within near zero per text.

To evaluate the hypothesis regarding the coefficient for the interaction between course 1 and course 2, a hypothesis test was conducted on the negative binomial model. Table 5-14 displays the results of the hypothesis test, indicating that the posterior probability of the hypothesis (course2 < 0) is 0.38. This suggests a 37% chance that the true coefficient value is negative, providing little support to reject the null hypothesis.

Table 5-14. Hypothesis Tests for class b

Hypothesis	Estimate	Est.Error	CI.Lower[23]	CI.Upper	Evid.Ratio	Post.Prob
course2 < 0	*0.07*	*0.22*	*-0.31*	*0.43*	*0.6*	*0.38*

Figure 5-8 depicts a density plot of the posterior distribution of the slope parameter for Course 2. The plot indicates a very wide but insignificant distribution for the slope, due to its inclusion of the null value at x = 0 within the distributional scope.

[23] 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses

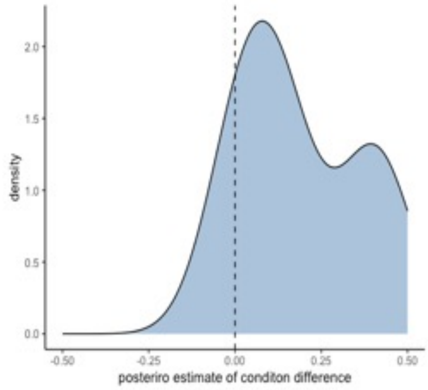

Figure 5-8. Plot of the posterior distribution of the slope of Course 2

Consequently, the analyses on adjective-clause use in the data indicate statistically insignificant variation over time. This may surprise readers, given the apparent upward tendency in Figure 5-6. There are a few possible reasons why the results might not be statistically significant despite the visual suggestion of an increase. The first speculation involves the wide posterior distribution, encompassing a range of possible values for the slope, including values close to zero. This width can indicate uncertainty in the estimate and can lead to non-significant results even if there's a true underlying trend. However, other factors in the model besides course (time) might be influencing adjective-clause use, potentially masking the effect of time or introducing noise into the analysis.

5.4.3 Results 2: Predicting the variation of noun-that clause construction

Table 5-15 and Figure 5-9 show the descriptive statistics and visualised frequency distribution of noun-that clauses across five courses, respectively. The average frequencies tend to increase over time, but the frequencies overall are very low, ranging from approximately 1 to 2 across all sub-corpora. Albeit very scarce in the corpus, the average frequencies increase over time, except for Course 5.

Table 5-15. The average frequencies of noun-that clauses across five courses

	course 1	course 2	course 3	course 4	course 5	Total
Average raw frequency	1.18	1.39	1.42	1.7	1.33	1.38

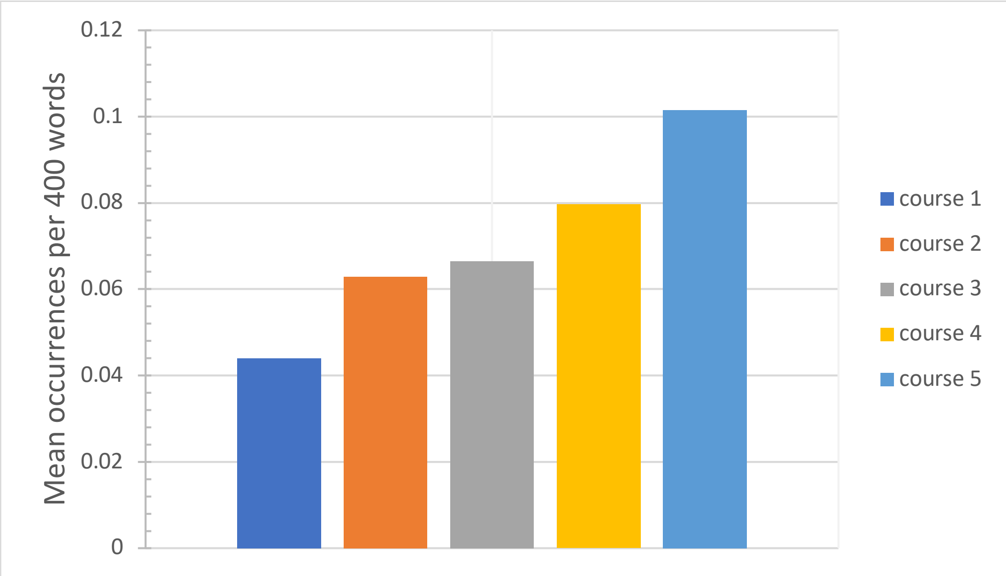

Figure 5-9. The average frequencies of noun-that clauses across five courses

Table 5-16 shows the result of a mixed-effects model of noun-that clauses across the five sub-corpora.

Table 5-16. a mixed-effects model of noun-that clauses across the five sub-corpora

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ ESS	Tail_ ESS
sd(Intercept)	0.08	0.06	0	0.23	1	10418	7887
sd(Course2)	0.12	0.09	0.01	0.34	1	10001	8337
sd(Course3)	0.12	0.09	0	0.35	1	8717	7080
sd(Course4)	0.36	0.29	0.01	1.08	1	9132	8036
sd(Course5)	1	1	0.03	3.72	1	9086	9190
cor(Intercept,Course2)	-0.07	0.41	-0.8	0.73	1	21616	10806
cor(Intercept,Course3)	-0.07	0.41	-0.79	0.72	1	21092	11819
cor(Course2, Course3)	0.01	0.41	-0.74	0.77	1	16341	12234
cor(Intercept,Course4)	-0.02	0.41	-0.76	0.74	1	19616	11659
cor(Course2,Course4)	0	0.41	-0.75	0.75	1	15451	13160
cor(Course3,Course4)	0.02	0.41	-0.74	0.76	1	13343	11845
cor(Intercept,Course5)	-0.01	0.41	-0.76	0.75	1	22099	12530

cor(Course2,Course5)	0	0.41	-0.76	0.76	1	17976	12347
cor(Course3,Course5)	0	0.41	-0.75	0.76	1	13951	12774
cor(Course4,Course5)	0	0.41	-0.76	0.76	1	12504	13192
Population-Level Effects:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ ESS	Tail_ ESS
Intercept	-5.65	0.16	-5.98	-5.34	1	13801	11899
Course2	-0.32	0.19	-0.68	0.06	1	13567	12995
Course3	-0.55	0.19	-0.91	-0.18	1	13201	12434
Course4	-0.4	0.34	-1.1	0.24	1	14016	10663
Course5	-0.79	0.97	-2.81	1.11	1	8806	7557

Figure 5-10 presents a plot for model checking of a negative binomial regression model using posterior predictive checks. A relatively good fit is indicated as the observed data represented by y falls within the range of simulated data represented by yrep, and the ECDF curves reasonably closely match each other. Again, the both observed and estimated counts are very low, almost all the observed counts fall beneath a count of five (on the x-axis).

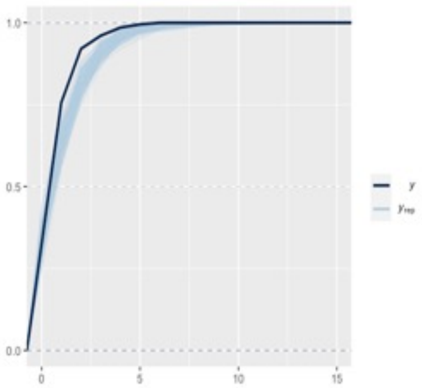

Figure 5-10. plot for model checking of a negative binomial regression model using the posterior predictive checks

To evaluate the hypothesis regarding the coefficient for the interaction between course 1 and course 2, a hypothesis test was conducted on the negative binomial model. Table 5-17 displays the results of the hypothesis test, indicating that the posterior probability of the hypothesis (course2 < 0) is 0.87. This suggests an 87% chance that the true coefficient value is negative. As the evidence is not significant, it does not provide strong support to reject the null hypothesis.

Table 5-17. Hypothesis Tests for class b

Hypothesis	Estimate	Est.Error	CI.Lower[24]	CI.Upper	Evid.Ratio	Post.Prob
course2 < 0	-0.2	0.17	-0.48	0.09	6.73	0.87

Figure 5-11 depicts a density plot of the posterior distribution of the slope parameter for Course 2. The plot indicates a relatively large and moderately significant distribution for the slope, with the null value (representing no difference between the conditions) indicated by a vertical dashed line at x = 0.

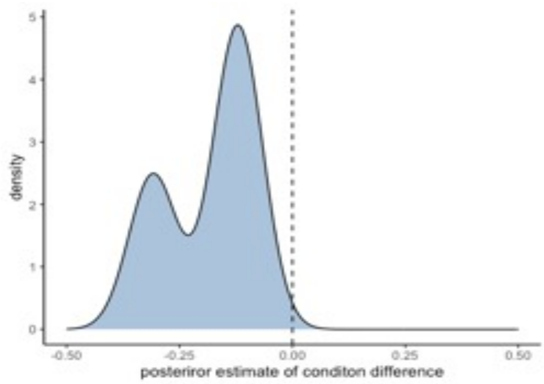

Figure 5-11. Plot of the posterior distribution of the slope of Course 2

In conclusion, the frequency variation of noun-clause construction in the data indicates some moderate significance in the variation over time. However, the significance is not strong enough to definitively confirm or reject an increase or decrease in the variable over time at a significant probability value of 0.05.

5.4.4 Discussions & Conclusions

This study aimed to unravel the developmental trajectories of that-complement constructions (THACs and THNCs; that-clauses controlled by nouns and adjectives) in L2 writing, considering the interplay of grammatical complexity and contextual factors. While the analysis revealed moderate increases in the

[24] 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

frequency of noun-that-complement constructions, the patterns for adjective-that-complement constructions were less clear. The methodology of this chapter, drawing from Poisson distribution and mixed-effects models, allowed for a meticulous examination of THAC and THNC frequencies. However, by investigating the raw frequencies of these constructions, this study encountered challenges due to the complexities in drawing robust conclusions from such data, particularly for dealing with low frequency constructions, which are often associated with short text lengths (see p.157 in Section 4.5.1 for a relevant discussion).

Previous studies such as Gardner et al. (2019), Pan (2018) and Gray et al. (2019) have emphasised the importance of understanding the contextual and grammatical influences on linguistic constructions. Gray et al.'s (2019) work shed light on how L2 proficiency levels might affect the use of complex grammatical structures, while Gardner et al. (2019) explored the impact of genres and disciplines on syntactic choices. On the other hand, Pan (2018) elucidated the effect of study level on the variations of grammatical complexity.

Comparative analysis between adjective- and noun-that-complement constructions, similar to the approach taken by Biber et al. (2021), highlighted intriguing distinctions in frequency patterns. While they have noted that adjective-related constructions are likely to be acquired and therefore produced at earlier stages than those of nouns, the finding showed that adjectives showed a steady increase while nouns did not. These nuances prompt questions regarding the structural and pragmatic differences between adjective- and noun-controlled clauses, urging a deeper exploration of grammatical and contextual determinants.

Analyzing adjective-that-complement constructions revealed statistically insignificant variations over time. Conversely, the examination of noun-that-complement constructions showcased moderate significance in variations, resonating with the observations made by Larsson & Kaatari (2020) in their exploration of noun-controlled constructions.

These findings resonate with previous research suggesting the complex nature of L2 syntactic development (Biber et al., 2011; Pérez-Paredes & Díez-Bedmar, 2020). The insights gained, although constrained by the limitations of low frequencies, align with the observations of Biber (1992) in emphasizing the complexity of language use, necessitating a nuanced understanding of contextual influences and writing development. The significance of premodifying slots in NP complexity has been discussed in relation to the phenomenon of nominalisation in Chapter 4. In this chapter, I have investigated the phenomenon of one aspect of the postmodification by nouns. These findings in combination indicate that both noun phrase premodification and noun phrase elaboration increase over time. However, this increase of noun phrases controlling that-clauses is an aspect of phrasal elaboration, which is not consistently discussed as a significant marker of L2 writing development (e.g., Díez-Bedmar & Pérez-Paredes, 2020). Further research on this should follow to elucidate on this aspect of linguistic complexity in L2 writing development.

5.5 Conclusions

The two mixed-effects models demonstrated the estimations of the distribution of noun-clauses and adjective-clauses as among the key syntactic features in the present thesis but were removed from Study 1. The raw frequencies of the two key constructions tend to increase over time. Still, overall frequencies are extremely low, which might be affecting the insignificance of their variations over time, as shown in the two mixed-effects models discussed in this section. Overall, the analyses reported on the noun-clause use in the data show little significance of the variation over time. On the other hand, the patterns of adjective-clause use show some moderate significance of the variation over time. However, the significance is not strong enough to provide definite confirmation or rejection for the increase or decrease of the variable over time.

The two features analysed in this chapter are both similar to and different from those in Chapter 4. Overall, the analysis did not find enough evidence to support the hypothesis of an increasing pattern over time. It may be understandable that from a one-year to a one-year-and-a-half period of time is not likely to be enough length of time to observe a salient syntactic variation. In this regard, rather than discarding the results based on the statistical significance, a more fine-grained approach may provide some insight into what is happening in the data during the observation period. For that aim, both qualitative and more refined quantitative subsequent analyses may prove to be fruitful, considering situational variables, a writing topic, which has proven to be significant as previously discussed in Chapter 4. In addition, writer-specific variables, not systematically investigated in this thesis but still proven to be significant in earlier discussions in Chapter 4, are also due more careful consideration in this stage. Among many possible ways to address this, the first thing to be conducted is a qualitative analysis of some of the samples classified with different topics and time points. If this step does not provide a feasible answer to this, it may be worth retagging the features for a more accurate representation of the frequency data of the feature and rerunning a mixed-effects analysis. The third possible step is to include more features for mixed-effects analyses to obtain some comparable data, from which to estimate the pattern in this particular feature in a broader picture of linguistic variations in the data. The updated measures taken and their results are discussed in Chapter 6.

Chapter 6 Collostructional Analysis of Stance features in L2 writing

6.1 Introduction

This chapter focuses on how verb selection within that-clauses evolves with writing development, providing insights into how adult writers of English as a second language (L2) express stance. Chapters 4 and 5 analyzed a range of linguistic features using frequency measures. This chapter, however, focuses on verb-argument constructions (VACs) and employs association measures. This shift is necessary because verbs differ fundamentally from nouns and prepositions.

Nouns and adjectives conveying the meaning of stance can have more diverse argument structures. They can appear in that-clauses (e.g., *the necessity of classes*), prepositional phrases (e.g., *with strong evidence*), or even stand-alone (e.g., *necessary*). This makes it difficult to classify the specific structures used for stance.

On the other hand, stance verbs like *argue, claim*, or *oppose* often require *that*-clauses to express a viewpoint. This predictability makes association measures more informative for these verbs. This makes VACs distinct from constructions controlled by nouns and prepositions. Verbs can appear in various constructions, including stance-related *that*-clauses (e.g., *argue that classes are necessary*). The verb withing a *that*-clauses can have different strength of the association.

This study examines two hypotheses about stance forms and functions in the PELIC longitudinal corpus. To achieve this, I used a method called collostruction analysis (Gries et al., 2005). This technique is particularly useful for investigating more sophisticated constructions because it focuses on the strength of association between words in a construction. Association strength data is crucial for the second hypothesis, which focuses on more sophisticated constructions.

Collostruction analysis differs from the methods based on frequency measures used in Chapters 4 and 5. Chapters 4 and 5 employed a multidimensional analysis method. This method focuses on how often different constructions co-occur. Collostruction analysis, on the other hand, hones in on the relationship between the verb and its dependent within a specific stance construction. In other words, it analyzes how strongly a particular verb is associated with the following element (e.g., *that*-clause) in expressing stance.

Using collostruction analysis helps gain a deeper understanding of how verb choice evolves within stance constructions as writing skills develop. This will shed light on how writers refine their ability to express stance over time under three key themes.

This chapter bridges the gap between stance research and formulaic language acquisition, from a usage-based perspective of L2 writing development. By employing collostruction analysis, I consider both the frequency of VACs and the strength of the association between verbs and their dependent clauses. This multifaceted approach allows us to investigate both formulaic and sophisticated constructions, providing a more comprehensive picture of how L2 writers develop their stance repertoire.

This chapter investigates how verb preference within stance constructions evolves in student writing over time. By analyzing verb choices, we gain insights into how L2 writers develop their ability to utilize stance for various purposes. Additionally, examining verb preferences across different text groups can illuminate how language use becomes more sophisticated (Kyle, 2016).

6.2 Literature Review

This section reviews the theoretical underpinnings of stance constructions, formulaicity, and sophistication. Section 6.2.1 examines how *that*-clause constructions following verbs function as markers of stance (writers' personal attitudes and evaluations) within academic writing (e.g., Hyland & Tse, 2005; Biber & Gray, 2016). These constructions are particularly frequent in student register, offering valuable insights into how writers position themselves relative to their arguments (Gardner et al., 2019; Hardy & Romer, 2013). Based on this, Section 6.2.2 explores the concept of formulaic language as prefabricated chunks that learners acquire through exposure (Durrant, 2018; Ellis, 2002). I discuss the hypothesis that L2 writers initially rely on more frequent VAC constructions, potentially reflecting the influence of their language input. This analysis sheds light on how formulaic language acts as a springboard for developing stance expression skills. In Section 6.2.3, I explore the concept of sophistication as a measure of L2 writing development. Sophisticated stance constructions, characterized by less frequent verb usage and potentially stronger association with academic registers, will be explored. I conclude this section by presenting two research questions linked to these constructs.

6.2.1 Stance in Student Registers

Apart from conveying factual information, individuals who speak or write often share their personal emotions, opinions, value judgments, or evaluations. This is referred to as expressing a *stance* (Biber et al., 2021b, p.958). While stance is generally less frequent in academic writing compared to other registers (Biber, 2021), it still serves a vital communicative function (Hyland & Tse, 2005).

The discourse within this section centres on investigations that utilize MDA methods to delve into stance, following on from Chapter 4.

Stance in academic writing has been researched with different conceptualisations (e.g., *hedging*, Hyland, 1994, 1996; *evaluation,* Hunston & Thompson, 2000). A range of terms has been employed across various analytical frameworks. See Table 6-1 for related stance functional categories of this construction.

Table 6-1. Stance parameter, function and linguistic devices related with complement that-clause constructions[25]

category	meaning	term	description	semantic verb class (example)	grammatical devices
attitudinal	status of knowledge (Gray & Biber, 2012)	evidentiality (Chafe, 1986; Hyland, 2005)	the degree of reliability of knowledge; writer's commitment to the reliability of a proposition	likelihood (appear, believe)	noun+that-clause constructions, stance noun+prepositional phrase constructions (Biber, 2006)
		status (Thompson & Hunston, 2000)	the degree of certainty the speaker/writer has regarding information		
attitudinal	personal feeling/ attitude (Gray &	affect	attitude toward the content	non-factive (say) factive (acknowle	
		Evaluation (Hunston &	expression of a speaker or writer's attitudes, feelings, and		

[25] Sources: 1) Gray & Biber (2012, pp.15-16, p.18) 2) Biber (2004, pp.133-135) 3) Hyland (2005, p.177) 4) Hyland & Tse (2005)

	Biber, 2012)	Thompson, 2000)	values towards a topic, which reflects the speaker/writer's value system and community	dge, realize, see, show) attitudinal (expect, feel)	
interaction	engagement (Hyland, 2005)	self-mention (Hyland, 2005)	the extent to which an author places themselves in a text		First-person pronoun, possessive pronoun, third-person pronouns with reference to the author
		engagement (Hyland, 2005)	dimension where writers acknowledge the readers' presence and involve them in the writing process.		Reader pronoun, directive
		expectedness (Hunston, 2000)	how obvious or expected the information is to the hearer/reader		

Stance, a multifaceted concept in writing, involves how a writer positions themselves and their ideas. Gray and Biber (2012) note that stance is broadly categorized as *attitudinal* and *epistemic*. Attitudinal stance reflects the writer's personal feelings, evaluations, and judgments about a topic (e.g., *I strongly disagree with this approach*). It conveys the writer's emotional perspective and subjective opinion. On the other hand, epistemic stance focuses on the writer's stance regarding the certainty or credibility of a proposition (e.g., *It is likely that further research is needed*). It conveys the level of knowledge, belief, or justification associated with the information presented.

That-complement constructions serve as powerful tools for expressing stance (Hyland & Tse, 2005). These constructions involve embedding a clause within a larger sentence structure to convey the author's perspective. This framework allows the writer to highlight their evaluation, making the attitudinal significance the foundation through which the embedded information is understood (Hyland & Tse, 2005). The versatility of that-clauses lies in their ability to be introduced by various predicates (e.g., communication verbs like *suggest* or mental verbs like *think).* Attitudinal and epistemic stance can both be expressed using that-complement constructions.

Interestingly, research suggests that epistemic stance holds greater importance in academic research writing compared to attitudinal stance (Gray & Biber, 2012). While stance plays a more prominent role in conversation, it is also crucial in academic writing to persuade readers to accept the author's viewpoint or research findings (Hunston & Thompson, 2000; Hyland, 2005; Charles, 2007). Some specialized grammatical devices frequently used in academic writing to express epistemic stance include extraposed complement clauses, stance nouns followed by that-clauses, and stance nouns followed by prepositional phrases (Biber & Gray, 2013). Interestingly, the PELIC corpus shows a stronger focus on attitudinal stance, as has been discussed regarding Dimension 1 in Section 4.6.

In adddtion to these two types of stance, Hyland (2005) proposed a stance framework for academic writing, incorporating a dimension of writer-oriented features of interaction. This framework examines how scholars mark their texts to indicate the potential accuracy or credibility of a claim (evidentiality, expressed through hedging and boosting), their level of commitment to the claim and the attitude they want to convey towards a subject, proposition, or reader (including markers of affect or attitude and presence, reflected by the use of first-person pronouns and possessive determiners).

The term *evaluation* is often used in stance discussions to explain the meaning conveyed by stance markers. These markers represent the writer/speaker's evaluation of a proposition or entity. Hunston (2000) outline four parameters, or meanings, of evaluative language: value and expectedness. These correspond to the attitudinal and engagement stance categories in Table 6-1, respectively.

Research suggests that stance features act as markers of registers (Biber, 2006). In analyses of university spoken and written registers, stance features were strongly associated with dimensions reflecting spoken/written differences (Biber, 2006). More generally, stance features were linked to ‘involved production’ and ‘persuasion’ in corpora with broader registers (Biber, 1988).

Interestingly, narrative dimensions appear consistently across languages and discourse domains, regardless of spoken or written mode (Biber, 2019).

Hyland's research found stance markers to be about five times more prevalent than engagement features in academic writing, with hedges being the most common aspect of writer perspective. This highlights the importance of distinguishing fact from opinion and presenting claims cautiously, considering colleagues' viewpoints (Hyland, 2005, p. 186).

While previous research emphasizes stance markers in academic prose, the multidimensional analysis findings suggest a stronger focus on engagement features in the PELIC longitudinal written texts, as discussed in Chapter 4 (Section 4.5).

Studies consistently report that student writing leans heavily on features associated with attitudinal stance, contrasting with the stance typically found in professional writing (Nesi & Gardner, 2017; Gardner, Nesi & Biber, 2019; Hardy & Romer, 2013). Stance in writing can be shown through verbs like 'think' or 'believe' with 'to' or 'that' clauses, often using 'I'. These verbs (mental or communicative acts) often appear with 'that-deletion', 'I', and past tense verbs. This suggests a personal stance (Gardner et al., 2019).

For instance, a study by Hardy & Romer (2013) found a high prevalence of adjectival features in student writing, which are uncommon in published academic texts (p. 193). Similarly, examining the BAWE corpus through a multidimensional analysis, Gardner, Nesi and Biber (2019) identified two distinct dimensions: ‘stance toward the work of others’ and ‘personal stance.’ Similarly, studies on British university student essays (Staples, Biber & Reppen, 2018) revealed co-occurring linguistic features performing specific functions. The dimension labelled ‘Personal Stance’ includes stance features like stance verbs followed by *that*-clauses or *to*-clauses, often accompanied by first-person pronouns.

Research also suggests that personal stance features are characteristic of student writing for L2 writers to a different degree (e.g., Weigle & Friginal, 2015; Pan, 2018). Several studies have shown how stance helps identify L2 student-specific variations in academic writing (e.g., Staples, Biber, & Reppen, 2018; Weigle & Friginal, 2015; Pan, 2018).

Studies examining TOEFL iBT responses revealed that higher-proficiency test-takers relied less on the stance dimension (Weigle & Friginal, 2015; Staples et al., 2018). Both Weigle and Friginal (2015) and Staples et al. (2018) found that a stance dimension was used less by higher-proficiency test-takers in TOEFL writing tasks.

Pan's (2018) study also demonstrates that stance devices differentiate L1 and L2 student writing. They compared Dimension 1 in prior studies by Biber (1988; 2006) and found that the positive features in this dimension highlight the importance of stance expressions in L1 and L2 academic writing. This suggests that using stance is crucial for distinguishing L1 from L2 writing within a specific academic register (applied linguistics). Notably, all three proficiency levels of L1 writing were more attitudinal than the corresponding L2 levels.

L1 writers, beyond conveying information, utilize various stance expressions to evaluate their own and others' work, along with expressing "judgements, opinions, and commitments" (Hyland, 2005, p.176). In contrast, L2 writers tend to be more reserved in critically evaluating others' research and expressing their own affective stance towards the information they present.

Chapter 4 classified the PELIC texts using Biber's (1989) typology, identifying ‘involved persuasion’ as the dominant text type. This type is more characteristic of spoken genres. Interestingly, while that-complement constructions can be used for both epistemic and attitudinal stances, they also function as a register marker (Biber et al., 2021b). They are generally more frequent in spoken registers, but within academic prose, the verb type controlling the that-clause influences the functional stance and differentiates student from expert writing.

Hyland's (2005) concept of engagement helps us understand the dominant PELIC text type (involved persuasion) discussed in Chapter 4 (Section 4.3). Hyland’s (2005) research found stance markers to be more frequent than engagement features, with hedges being the most common marker of a writer perspective. However, the PELIC corpus exhibited a stronger correlation with features marking engagement (e.g., second-person pronouns, directives).

These features, while indicating engagement rather than stance, can still work together to convey the writer's opinion or feelings.

Dimension 1, identified in the PELIC corpus analysis, likely relates to this persuasive function. Text samples with frequent features positively loaded on Dimension 1 (e.g., second-person pronouns, conditional clauses, infinitives, mental verbs) aim to persuade and involve the reader. These features have also been interpreted as spoken-like, including contractions and mental verbs with to-infinitive complements.

The MDA research discussed so far suggests that student writing tends to use linguistic devices associated with attitudinal stance rather than epistemic stance. These devices express the author's personal viewpoint on a statement or fact within a sentence. They often involve verbs, nouns, or adjectives followed by complement clauses or infinitive phrases. While the PELIC corpus also showed separate sets of lexico-grammatical features for personal involvement (Dimension 1) and persuasive explanation (Dimension 3), these dimensions did correlate with each other (see Section 4.5).

6.2.2 Formulaicity

This section explores formulaic language (FL) in L2 writing development. I introduce definitions of FL and discuss its potential benefits for L2 learners, drawing on existing research on FL.

Formulaic language refers to prefabricated sequences of words or elements that learners acquire and use as wholes (Wray, 2018). These sequences can be single words (e.g., *hello)* or multi-word expressions (e.g., *it goes without saying*). Siyanova-Chanturia and Pellicer-Sánchez (2018) define FL as an umbrella term encompassing both single words and multi-word expressions. They highlight its multifaceted nature, including aspects like familiarity, conventionality, and predictability. The concept of formulaicity can also be applied to grammatical structures with open slots (where a verb can be inserted). From a usage-based perspective, where the distinction between lexis (vocabulary) and grammar is blurred, these constructions can also be considered formulaic.

There are two main approaches to studying FL from a frequency-based approach: a linguistic approach (relying on linguistic criteria to define formulaic sequences, e.g., Granger & Paquot, 2008 and a frequency-based Approach (using quantitative measures like frequency of occurrence to identify formulaic sequences, e.g., Biber et al., 2004). Within the frequency-based approach, researchers typically distinguish frequency from association strength. Frequency refers to how often a particular sequence appears in a corpus (collection of text) while association Strength involves how strongly the individual words within a sequence tend to co-occur.

Biber et al. (2004) is one of the examples which focuses on the frequency of lexico-grammatical combinations. They argue that frequent combinations like *think that* or *want to* represent a reliance on semi-formulaic language, associated with earlier stages of development. This is because learners might be using these pre-built structures without fully understanding the underlying grammatical complexities.

While these combinations may appear equally complex from a theoretical standpoint (they all have the same number of elements), a usage-based perspective reveals a crucial difference. Highly frequent combinations require less processing power, making them easier to learn and use compared to less frequent ones. Therefore, studying these patterns can provide insights into the nature of complexity itself (Biber et al., 2021a).

The second perspective focuses on the strength of association between elements within a formulaic construction. This is measured by how often specific elements co-occur, regardless of overall frequency. Similar to the frequency-based approach, the prediction is that learners acquire stronger associations first. For example, a highly associated verb-construction combination like *admit that* might be learned before a less associated one like *conclude that*.

The key difference between the two approaches lies in their focus. The frequency-based approach looks at overall usage, while the collostructional approach examines the inherent attraction between specific elements. This can lead to infrequent but highly associated word combinations, particularly in specific registers like academic writing. These combinations might be challenging for learners to acquire due to their rarity in everyday contexts, yet they are valuable for achieving target communicative goals. In this sense, formulaicity aligns with communicative competence, the ability to use language effectively in different situations (Granger, 2018).

Both frequency and association strength likely play a role in L2 writing development. Research suggests that measures based on association strength are better predictors of learner preferences (Gries et al., 2005), acquisition order (Ellis & Ferreira-Junior, 2009), and overall formulaic language proficiency (Schmitt, 2004). These findings highlight the importance of considering both perspectives when studying formulaic language and its impact on L2 writing development.

Notably, studies have shown the significant role that FL plays in research on academic writing and communication (Durrant, 2018). A key characteristic of FL is that it promotes fluency. By learning and using formulaic constructions, learners avoid having to build sentences from scratch, leading to smoother and more natural-sounding language production (Durrant & Mathews-Ayadinli, 2011). Similarly, Biber et al. (2004) observed the usefulness of frequently occurring word strings (called lexical bundles) for L2 learners. They distinguish between formulaic expressions (completely fixed phrases) and semi-formulaic expressions (where some components can vary). Both types offer benefits for fluency development.

Despite the fluency benefits, research suggests that L2 learners may underuse formulaic expressions (Wray, 2018). This is puzzling because these expressions offer a clear path to more confident and efficient communication. Wray (2018) observes that formulaic sequences identified by frequency and association strength might have different learning trajectories. Lexical bundles (highly frequent word strings like *such as*) and more semantically complex formulaic expressions (like phrasal verbs) may be learned and used differently.

Furthermore, research suggests that formulaic language chunks as "prefabricated elements that learners acquire through exposure (Ellis, 2002)" can be particularly useful for L2 learners (Biber et al., 2004). The frequent use of stance features, which can be considered semi-formulaic, may offer valuable insights into L2 writing development. However, a register-specific analysis of formulaicity in stance constructions remains relatively unexplored (compared to genre analysis; see Chapter 2, Section 2.2 for a discussion on register vs. genre). By adopting a register perspective, this chapter builds upon the findings of Chapter 4 on stance function and explores how formulaicity relates to development in L2 writing.

6.2.3 Syntactic sophistication

This section explores how formulaic language (FL) can be a marker of development in L2 writing, aligning with usage-based research on lexical sophistication (Kyle et al., 2021; Romer & Berger, 2019).

Kyle et al. (2021) investigated the development of verb-argument construction (VAC) sophistication in adolescent EFL learners' writing over two years. They found a correlation between writing quality (holistic scores) and VAC sophistication measures (e.g., main verb frequency and verb-VAC frequency). This suggests that learners increasingly use less frequent controlling verbs as their writing improves. Importantly, collostructional strength, a measure used to assess formulaicity, is calculated based on language data representing a specific type of discourse (e.g., academic writing). Therefore, what constitutes *sophisticated* formulaic language may depend on the register of the data used for analysis.

Usage-based theories argue against a strict separation of form and meaning (Goldberg, 2003). They propose that grammatical forms, like lexical items, carry meaning. Collostructional analysis, which examines the association strength between words and grammatical patterns, reflects this view (Gries & Ellis, 2015). It assumes that words and grammatical patterns function similarly in language acquisition and use.

Previous research suggests that VACs are psychological units acquired and produced according to Zipf's law (Ellis et al., 2014). This law predicts that a small number of constructions will be very frequent, while a larger number will be less frequent. From this perspective, learners likely acquire highly frequent VACs first, followed by less frequent ones.

The analysis discussed in this chapter proposes a hypothesis: VACs with stronger association strength specific to academic writing (compared to general language) may emerge later in development, signifying greater sophistication. Additionally, analyzing the collostructional strength of less frequent verbs in VACs can reveal information about formulaic constructions particularly useful for academic writing. These constructions may include verbs that convey specific meanings relevant to academic contexts.

By investigating formulaic language through the lens of collostructional strength and register variation, we can gain valuable insights into the development of lexical sophistication in L2 writing.

6.3 Pilot study: Do stance constructions become more formulaic in L2 academic writing through the revision process?

The analysis results discussed in Chapter 4 revealed the salience of personal stance in student writing, but the associated features (e.g., second-person pronouns, mental verbs, conditional adverbial subordinators, to-infinitives) were not those conventionally pointed out (e.g., first-person pronouns, mental verbs, that-complement clauses). This necessitates an investigation of the stance markers further to compare the L2 learners' writing with the target student registers. The objective of this pilot study is to explore potential similarities or differences between the corpus of the students’ essays written in universities in an English-speaking country and the selection of the L2 writing data.

In English as a second language education programs, students receiving instruction are highly likely to receive feedback directed towards tasks and level-specific achievement goals set by their teaching instructors, prompting them to reflect on and incorporate this feedback into their language usage. Moreover, the educational programs primarily employ process writing as the main teaching method. Therefore, there is a high likelihood that manuscript revisions will occur in alignment with the feedback provided by the students' instructors, which also serves as a basis for predicting the direction of language development based on the possibility that students will strive to use language encouraged in the register targeted by the English education program.

This serves as a basis for supporting the significance of investigating language variation in essays written by students in such English education programs. In other words, studying essays written by students in educational environments using English as a second language has already been emphasized as an important foundation for language development in Chapters 1 and 2. For these reasons, the variation in language usage that arises from multiple revisions based on a single topic and task is also valuable as reference data in such language usage studies. Examining whether the results of such research show patterns of consistency or inconsistency with longitudinal study data will be a useful comparison. The pilot study introduced in this section shares the fundamental concerns and aims of the research questions introduced in the preceding and introductory sections of this chapter. However, it is based on process writing data and employs slightly different research techniques thereafter. The difference in research techniques stems from the relatively small scale of the dataset, which will be detailed further in the methodology section.

This study aims to answer the following question:

Q1. What are the verbs that are highly formulaic in the *stance* constructions in the L2 written revisions?

Q2. What stance functional dimensions are associated with *highly sophisticated* verbs (VACs that are infrequent but with high collostruction strength in academic registers)?

Q3. How do the sophisticated constructions' association strengths vary over time?

This study uses the association strength measure to explore the variation of VACs associated with stance function, verb-that clauses, and *verb-to infinitives* in L2 writing.

6.3.1 Corpus Data

The primary data is the University of Pittsburgh English Language Institute corpus (PELIC; Juffs et al. 2020), an L2 longitudinal learner corpus collected in an intensive English programme. A subset of the PELIC corpus data was used, consisting of written essays with one or two revised versions, produced by learners of 8 linguistic backgrounds who contributed written essays throughout three or more academic semesters.

In addition, a reference corpus representative of the target student register is the Michigan Corpus of Upper-Level Student Papers (MICUSP; O'Donnell & Römer, 2012; Römer & O'Donnell, 2011), the collection of the university students' written assignments. Table 6-2 shows the two corpora used for the Pilot study.

Table 6-2. Data composition for pilot collostructional analysis

	PELIC (main corpus)	MICUSP (reference corpus)
description	the University of Pittsburgh English Language Institute corpus (PELIC; Juffs et al., 2020) - an L2 longitudinal learner corpus collected in an intensive English programme in the U.S.	the Michigan Corpus of Upper-Level Student Papers (MICUSP; O'Donnell & Römer, 2012; Römer & O'Donnell, 2011) - a target student register
No of texts	326 (1st, 2nd and 3rd drafts)	20 (senior undergraduate assignments in Biology, Psychology, and Sociology)
No of words	125,894 (386 per text)	77,729 (3,886 per text)

The 326 PELIC texts were organized in an Excel spreadsheet for analysis. Each row in the spreadsheet represents a single student, which contains information about multiple texts produced by that student in a long data format. This format allows us to track each student's development over time across personal course study (the chronological order of the semesters in which the student contributed writing samples) as well as draft versions (whether the text is the first, second, or third draft written by the student on the same topic).

As a result of this coding scheme, each student's entry can be linked to up to nine different time points based on the number of texts they contributed, reflecting their writing journey across drafts and semesters.

There were several texts on different topics produced in the same semester and revised by the same writer. The texts were processed by the order of 1) semester 2) version 3) text id. In other words, ‘version’ was prioritised to ‘text id’ in ordering texts. This ordering makes sure that the revised texts on the same topic are ordered subsequently. It ensures that the revision based on the teacher’s feedback can be tracked.

As discussed in Section 3.2, the writing rubric that came from the intensive language programme shows that the writing activities aimed to prepare students for the main writing tasks in their disciplines. What differentiates this study from the studies drawing on the main longitudinal data is that this study focuses on the draft revision process, to see if there are any variations in the stance constructions across the revised drafts. The selected data is relatively smaller than the main data, about a third, because they were included only when they had revised versions in the longitudinal data.

The MICUSP corpus is used as reference data for the sophisticated construction of the student register. Randomly selected 20 undergraduate texts were used in this study. I chose to use part of the whole corpus to facilitate the processing considering that it is a pilot analysis aiming to observe the soundness of the methodology for the final study. The main reason to have this reference data is twofold: to have it as a benchmark for which the increasing or decreasing trend of the key construction’s collostructional scores across revised texts and to select target-like constructions so that such constructions could be investigated further in the main L2 data.

6.3.2 Method

I took the following steps for the collostruction analysis of stance constructions.

First, I extracted occurrences of the two relevant constructions, *that* clauses and *to*-infinitives (i.e., the frequency a + c in Table 3-10. Contingency table for collostructinal analysis of word W1 and construction C (Gries et al., 2005, p.644)), from the corpus using MAT (Nini, 2019). Table 6-3 shows the structural configuration tagged to the four features:

Table 6-3. Four *features* tagged by MAT (reference: MAT manual)

feature	tag	Example
That verb complements	THVC	I hope that you come.
Infinitives	TO	I want to come.
Subordinator that deletion	THATD	I hope you come.
Split infinitives	SPIN	I want to really come.

The second step was to lowercase and lemmatize all the 107 verbs occurring in these 687 instances using WordNetLemmatizer in Natural Language Toolkit (Bird et al., 2009) in the Python environment; this also included the manual lemmatization of phrasal verbs as well as the correction for spelling variants such as -ize vs. -ise, etc.

The third step consisted of determining the overall frequencies of these verb lemmas in the corpus. I identified the frequency of each verb controlling the construction (i.e., frequency *a* in Table 3-10). Then, the frequency of each such word was determined in the entire corpus (i.e., the frequency *a* + *b* for W1 in Table 3-10).

Fourth, I estimated the number of constructions in the corpus (i.e., (a + b) + (c + d) = N), which was approximated by using the token frequency of all verbs. As a result of these four steps, the frequencies in bold type in Table 3-10 were obtained so that the remaining frequencies could be computed by means of subtraction. For example, b results from subtracting the frequency of the word in the construction (i.e., a) from the word's overall frequency (i.e., a + b).

The final step of analysis involved computing for each verb its collostruction strength using the Fisher-Yates Exact test to the that-clauses and to-infinitives, respectively (Gries et al., 2005; refer to the formula of this in Section 3.3.6.2). These steps are then repeated for every verb occurring in the construction within the two corpora, respectively.

In order to identify the presence of stance constructions within the corpora, a comprehensive list of mental verbs governing both *to*-infinitives and *that*-clauses were used as referenced in Biber et al. (2021, p.369). The analysis employed a total of 91 mental verbs, which have been recognized as highly frequent in general corpora. More specifically, the aim of the pilot study was to explore the variation of the strongly formulaic stance constructions through the revision process, rather than finding overall patterns of all the stance constructions (see Appendix 6-1 for a list of the verbs used in the study).

Subsequently, one linear and one nonlinear Linear-Generalised Model (LGM) models (Preacher & Hancock, 2015) were implemented using the R package

n (Rosseel, 2012). These models examined the individual and overall trajectories of 1) the absolute frequencies of the stance constructions and 2) converted frequencies weighted with formulaicity scores. The patterns among the scores were used to test the two aforementioned hypotheses, as shown in Table 6-4. The sum value of all the association measures in each text was then used to fit a non-linear growth model.

Table 6-4. Two non-linear growth models

Hypothesis	Independent variable	Dependent Variable
Hypothesis 1: There will be a decrease in the use of formulaic language (general formulaicity) across revisions of the student writing samples.	Revision order	Sum value of collostruction scores
Hypothesis 2: There will be an increase in the use of target-like stance constructions across revisions of the student writing samples	Revision order	collostruction scores of 'verbs most preferred in reference corpus'

To investigate the first hypothesis noted in Table 6.4, a growth model was employed utilizing the sum value of the association strength scores as a variable. The objective was to assess if an over-reliance on specific verbs in the target stance constructions is more pronounced in the early drafts. To examine the second hypothesis, an additional growth model was conducted, utilizing a different variable: the six verbs with the highest ranks in the MICUSP corpus. The rationale behind selecting these six verbs was to investigate whether these verbs, which exhibit a higher degree of target-like usage, become more strongly associated with the later stages of the revision process in the L2 writing data.

For this pilot analysis, I extracted occurrences of the two relevant constructions, *that* clauses and *to* infinitives (i.e., the frequency a + c in Table 3-9), from the corpus using MAT (Nini, 2019). All the 107 verbs occurring in these 687 instances were lowercased and lemmatized using WordNetLemmatizer in Natural Language Toolkit (Bird et al., 2009) in the Python environment. Once obtaining the remaining frequencies of these verb lemmas in Table 6-4, I computed for each verb its collostruction strength to the that-clauses and to-infinitives, within the two corpora, respectively. In order to identify the presence of stance constructions within the corpora, the analysis employed a total of 91 mental verbs, which have been recognized as highly frequent in general corpora (Biber et al., 2021b, p.369). Finally, one linear and one nonlinear Linear-Generalised Model (LGM) models (Preacher & Hancock, 2015) were implemented using two R packages: *lavaan* (Rosseel, 2012). These models were used to examine the individual and overall trajectories of collostructional scores to test the two aforementioned hypotheses, as shown in Table 6-4.

6.3.3 Results & Discussion

The rank of the collostruction scores computation results is reported in Tables 6-5 and 6-6. Table 6-5 presents the ten most strongly associated verbs for "that" constructions in the MICUSP corpus and the scores and ranks of these verbs while Table 6-6 shows those for '*to*' constructions. The order of the controlling verbs has been arranged based on their ranks in the MICUSP corpus, which serves as the reference corpus, enabling direct comparison of the scores between the PELIC and MICUSP corpora. The score ranking provides a comparison point of the relative association strengths of the verbs in the construction.

Table 6-5. Collostruction analysis score of verb-that clauses

PELIC			MICUSP		
rank	verb	score	rank	verb	score
8	find	9.76	1	find	42
5	believe	18.3	2	believe	25.4
1	think	101	3	think	18.9
16	conclude	1.92	4	conclude	16.9
6	know	18.3	5	know	11.1
2	mean	38.1	6	mean	10.4
3	realize	22.3	7	realize	6.57
4	learn	20.7	8	predict	6.36
7	feel	12.9	9	perceive	5.63
9	see	9.73	10	recognize	4.33

The verb 'find' in Table 6-5 exhibits the highest score in MICUSP, while taking up the 8th place in the PELIC corpus. This indicates that the verb 'find' is notably more strongly favoured and attracted to ‘that’ constructions in MICUSP compared to PELIC. Conversely, the verb 'think', which holds the top position in PELIC, ranks third in MICUSP. Although 'think' demonstrates a strong association with the constructions in both corpora, the score in PELIC suggests an over-reliance on this particular verb in the PELIC revision corpus.

Table 6-6. Collostruction analysis score of verb-to clauses

PELIC			MICUSP		
rank	verb	score	rank	verb	score
2	need	67.81	1	need	42
1	want	186.83	2	want	19
N/A			3	compare	18
21	find	-11.00	4	find	10
15	expect	-2.00	5	expect	5.3
16	choose	-2.06	6	choose	4.7
6	suppose	10.25	7	suppose	4.1
4	decide	38.76	8	decide	3.7
3	like	52.84	9	wish	3.1
5	prefer	21.62	10	afford	1.7

A similar trend is observed in relation to to-infinitives in Table 6-6. For instance, the verbs 'need' and 'want', ranked first and second, respectively, exhibit notably higher association strength scores. Although these verbs show a strong association with *to* infinitives in both corpora, the scores are relatively moderate in the MICUSP corpus. The exceptionally high scores of these two verbs in the PELIC corpus indicate a potential over-reliance on them in L2 writing. It should be noted that the score comparison may be influenced by text length, as the sub-corpora have similar sample sizes but uneven text lengths (refer to Table 6-2).

Finally, the results of the two Growth Models are presented in Table 6-7. The slope of each model of the stance construction is interpreted as follows. A negative slope in the first model and a positive slope in the second model is interpreted as a development in the ability to produce more target-like constructions. In other words, the stance constructions with the high association strength scores are expected to be more saliently observable in later stages of the writing process, as evidence of development.

The results of the two Growth Models in Table 6-7 show evidence of ‘highly formulaic’ and ‘sophisticated constructions (operationally defined as less frequent, but highly associated in academic writing).’

Regarding the first hypothesis, a negative slope in the first model suggests that as revisions progress, constructions tend to become less reliant on very frequent, general-purpose verbs. This shift indicates development from overusing high-frequency combinations to a wider, more contextually appropriate vocabulary.

As for the second hypothesis, a positive slope in the second model means that VACs with sophisticated verbs (less frequent, but highly associated in academic writing) become more prominent across revisions. This aligns with the expectation that learners incorporate more specialized, register-appropriate verb usage over time.

These findings suggest L2 writers expand their VAC vocabulary and that their verb-construction associations become more target-like over the revision process. This supports prior observations about L2 learners initially relying heavily on high-frequency constructions (Granger, 2018).

Table 6-7. The results of the (non)-linear models

Variable	Model	Intercept	slope	Goodness of model fit	
Sum of all the collostruction scores	Linear model	1.157	-0.014	Test statistic	2.243
				Degrees of freedom	1
				P-value (Chi-square)	0.134
				Comparative Fit Index (CFI)	0.992
Six target-like verb scores	Nonlinear model	2.676	0.177	Comparative Fit Index (CFI)	1.000
				Tucker-Lewis Index (TLI)	1.000

The linear Growth model regressed the collostruction scores based on the sum value of all 53 mental verb frequencies to predict the trajectories of the collostruction score in the texts across nine time points. The result suggests a minimal decrease in the score over the revision process, with a decrease of 0.014 for each revision point. This indicates that the overall formulaicity score decreased during the revision process.

For the Nonlinear Growth Model in the second row, the collostruction score of the MICUSP was referenced to decide on six the most sophisticated verbs in the constructions. The trajectories of the collostruction score in the PELIC revised texts across nine-time points were regressed. The results indicate that the initial score, starting at 2.676, increased by 0.177 at each revision point. This suggests that these verbs became more strongly associated with that-constructions in the later stages of the revision process in the PELIC corpus.

The results showed that the more target-like constructions in constructions associated with the more sophisticated constructions were more saliently observable in the later stages of the writing process. Albeit the trend was significantly inconsistent among the constructions, the results indicate that writers expand their VAC repertoire and productivity over time, and verb-VAC associations move closer to native usage.

The results overall add evidence for the assumption that writers expand their VAC repertoire and productivity over time, and verb-VAC associations move closer to target-like usage. The study result expands the observation on collocation to the realm of collostruction that L2 learners use a large number of constructions filled with high-frequency verbs, and they tend to use a limited range of high-frequency combinations (Granger, 2018, pp.240-241).

However, the overall fit of the models is not satisfactory, potentially due to the trend being significantly inconsistent among the constructions. This is interpreted in relation to the strong random effects (individual differences and writing topics) observed in the mixed-effects models reported in Chapter 4 (4.6), which are likely to interact with the nature of the longitudinal data with relatively short time points mixed with revised versions.

6.3.4 Conclusions and Next Steps

The collostruction score rankings discussed in Section 6.3.3 revealed specific verb preferences within each corpus. I further examined individual verb-construction patterns ('find that', 'think that', 'need to', 'want to') to understand potential tendencies of preferences toward specific verbs within the target construction. The preferred type of dependent construction is the that-deleted construction, which implicates a spoken nature of the linguistic use in the corpus, as Biber (1988) noted. Moreover, the overuse of ‘think’ constructions decreased over time, indicating the variation of the construction was consistent with the hypothesis of sophistication.

While the pilot study offers initial insights, methodological limitations of the analysis should be noted. First, a larger L2 dataset and a more comparable reference corpus would enhance the reliability of comparisons. Considerations for the final study (Section 6.4) include both the longitudinal L2 writing corpus and a reference corpus closely aligned to the main L2 dataset in terms of context and text length.

In addition, this pilot study result suggests the need that collostruction strength measures should be refined. Instead of using a sum of all verbs for a construction, focusing on the trajectories of a selected subset of sophisticated verbs could provide a more targeted analysis for testing the first hypothesis. This aligns with the individual verb differences observed in the score rankings.

6.4. Final study: Analysis of stance Verb-Argument Constructions in the L2 longitudinal writing

6.4.1 Introduction

This chapter builds on the introduction of stance features in Chapter 2 (Section 2.4) by applying a collostructional analysis. This analysis focuses on one specific stance construction feature (verb-that-clause) and aims to answer the following research questions:

Question 1: What verbs co-occur most frequently (have the strongest association strength) with the chosen stance construction in L2 writing?

Question 2: What stance functional dimensions are typically associated with verbs that are both generally infrequent and have high collostructional strength within academic writing registers (highly sophisticated verbs)?

Question 3: Do the association strengths of these sophisticated constructions (VACs with infrequent but highly associated verbs) change over time in the L2 writing samples?

The analysis aims to investigate how the set of strongly attracted verbs within stance constructions varies across the writing samples. This will provide evidence for or against the hypothesis that the variety of stance functions is associated with both the frequency and the sophistication (semantic categorization) of the verb within the VAC.

Furthermore, the analysis aims to identify the main communicative purposes served by the target stance constructions. Ultimately, by examining the longitudinal variations in stance features, we can investigate if the constructions associated with stance become more sophisticated over time. In other words, this analysis will explore how the grammatical complexity and overall sophistication of stance features change across the writing samples.

Methodologically, I compared collostruction scores between the main longitudinal dataset and the Level 5 texts in the PELIC corpus. This comparison aimed to track if verb rankings within VACs in the longitudinal data become more similar to the Level 5 patterns over time. I selected Level 5 as a reference point for several reasons:

While the longitudinal data includes some Level 5 writers, the sample size is smaller compared to the entire Level 5 subset of the PELIC corpus. Using this larger dataset increases representativeness.

Level 5 usage serves as a benchmark for investigating target language variation and verifying my hypotheses. It's not about judging whether the longitudinal data perfectly matches Level 5, but about observing developmental trends and tendencies.

There are potential limitations to using any single corpus as a reference point. A future discussion could address this by comparing the Level 5 benchmark to other reference corpora like MICUSP or general academic prose.

It's important to emphasize that the actual results of this analysis will further justify the use of Level 5. Our findings will either support or challenge its suitability as a reference point for this particular analysis.

6.4.2 Results & Discussion: Collostruction score across five courses

Preliminary analysis interpretation considered ten verbs with the highest collostruction score, and their semantic categories in the corpus. Typical stance complement clauses controlled by verbs (Biber et al., 2021b, p.366, p.661, pp.961-962, pp.659-661, p.664) were considered in interpreting the score ranking.

Table 6-8 shows the ten most strongly attracted verbs to the that-complement clauses in the PELIC corpus. The analysis result suggests that *think* and VAC are most strongly attracted, while *find* and VAC are most weakly attracted among the ten verbs. See Appendix 6-2 for a list of all the verb-VAC combinations by collostruction strength score rank and verb frequency in VACs.

Table 6-8. Verbs most strongly attracted to the That-complement clauses in the longitudinal PELIC data based on the collostruction scores (V frequency in the target construction)[26]

Collo-struction score rank[27]	Verb	collostruction score			V frequency in VAC			Rank of verb in VACs (freq)	Academic register freq	Semantic category
		V that S V	V () S V	V (that) S V	V that S V	V () S V	V (that) S V			
1	think	240	139	379	246	147	393	2 (860)	**	mental V
2	see	79	74	153	82	77	159	5 (614)	****	mental V
3	believe	47	84	131	49	86	135	9 (213)	-	mental V
4	know	46	40	86	55	49	104	1 (909)	***	mental V
5	mean	46	33	79	49	37	86	7 (360)	*	mental V
6	feel	22	38	59	28	44	72	4 (627)	-	mental V
7	learn	48	-6	42	56	3	59	3 (823)	-	mental V
8	realize	7	33	41	8	34	42	10 (66)	-	mental V
9	say	20	18	38	26	24	50	8 (309)	***	communication V
10	find	19	13	32	24	19	43	6 (543)	****	mental V

6.4.3 A qualitative analysis of stance constructions

The purpose of this subsequent analysis is to find functional and/or semantic categorization of the target construction used in the L2 writing. The verbs explored in this chapter have several different meanings and usages, and they appear in various constructions as well as that-complement constructions. The focus of this analysis is to look into the specific use of these verbs in *that*-complement clauses, more specifically, their stance functions in this construction.

This qualitative analysis aims to provide contextual information to make sense of the collostruction analysis result; more specifically, it aims to provide further information to interpret the collostruction analysis score of the ten verbs in that

[26] academic register frequency (from Biber et al., 2021b)

[27] the collostruction score rank is based on the third measure among the three scores in the table, which is the collostructional score calculated based on both V that SV and V (zero that) SV constructions.

complements and the ranking of these verbs across three Course groups, following the process as discussed below.

The first round analyzed ten random text samples for the ten verb-controlled constructions referring to Table 6-8. However, this didn't reveal enough about how these constructions function differently in real communication, where the constructions were generally infrequent (the average occurrence of each construction was zero or one). To understand communicative function, I analyzed whole texts, rather than focusing on the concordance lines including the target constructions. This allowed me to see how a construction contributes to the overall topic and message of the text. Additionally, I included texts with multiple occurrences of both that-deleted and that-included constructions (e.g., 'find SV' and 'find that SV') for more context.

The 'say SV' example initially served to show how text length and course level might have skewed the initial analysis. To make this example more relevant, we need to explain why *say* is problematic from a stance perspective. *Say* is primarily a reporting verb and might not strongly convey the writer's stance or attitude. This makes it less useful for our analysis, which focuses on how writers express their viewpoints through VACs.

Once all of these texts were examined, I focused on identifying verbs central to expressing the writer's stance or attitude. Verbs like 'see' and 'learn', while are likely directly linked to conveying viewpoints, were therefore excluded, as ‘learn’ was not used in the stance function in the data. To illustrate, while the 'say SV' construction was initially frequent, it became clear that its function was mostly to report information. This didn't align with my goal of examining stance expression.

In addition, ‘See SV / See that SV’ constructions were also removed due to samples not associated with stance function.

Table 6-9 shows a summary of the qualitative analysis of stance constructions analysed in this chapter (See Appendix 6-2 for a full list of all the verbs in that-complement clauses in the longitudinal PELIC corpus).

Table 6-9. Features and functions associated with the key stance constructions

verb	number of texts of each construction occurrence	total occurrence	semantic category[28]	semantic category [29]	associated linguistic features[30]
think SV/ think that SV	277	393	likelihood verb	mental V	First-person pronoun, present tense
say SV/ say that SV	115	159	non-factive verb	communication V	Third-person pronouns, past tense
believe SV/ believe that SV	105	135	likelihood verb	mental V	
know SV / know that SV	95	104	factive verb	mental V	You (second-person pronouns), present tense / You (second-person pronouns), present tense, interrogative
mean SV / mean that SV	79	86	factive verb	mental V	Third-person pronoun, present tense
feel SV/ feel that SV	56	72	attitudinal verb	mental V	First-person pronoun, present tense / Third-person pronoun, present tense
learn SV/ learn that SV	35	59	factive verb	mental V	
see SV/ see that SV	50	50	factive verb	mental V	
find SV/ find that SV	39	43	factive verb	mental V	First-person pronoun, that deletion, past tense / third-person

[28] the stance verbs classified by semantic category (Biber, 2004, pp.133-135).
[29] verb semantic types are classified following Biber et al. (2021)
[30] Separately presented only when a formal difference exists between the two constructions headed by the same verb

					pronoun
realize SV/realize that SV	36	42	factive verb	mental V	First-person pronoun, past tense, that deletion

Table 6-9 shows that mental and communication verbs controlling *that*-clauses were the main patterns, which did indicate the stance function actively employed in this construction. As mental verbs with *that*-clauses are considered an important device to express stance (Biber et al., 2021), the result shows that most verbs that are strongly attracted to the focus VACs are semantically categorised as mental verbs. These verbs all seem to be associated with the thoughts or feelings of a human subject. This association contrasts with typical stance use in academic prose reported by Biber et al. (2021), where the authors associated the use of communication verbs such as *show, ensure, and indicate* with the function of reporting with a non-personal subject. As for the verb scores of different dependent types, all the verbs were more frequent in *that*-deleted construction.

As the same that-clause construction can be used for epistemic and less often, for attitudinal stance function, it is possible that 'find' can be used for either of the two types of stance (attitudinal and epistemic). Quantitative counting of this construction would not differentiate such nuanced functional differences, and therefore I sampled texts to examine how each construction was used in context. I intended to find the typical and representative use of these constructions, and therefore, the targeted sample texts consisted of the ones with a high frequency of the target construction. To that aim, Sections 6.4.3.1 to 6.4.3.8 present the qualitative analysis based on the framework of stance function in Table 6-1. Stance parameter, function and linguistic devices related with complement that-clause constructions.

6.4.3.1 Construction 1: Find SV/ Find that SV

(1) I **found** each country has eight coins with different face value.
text id_11470, Course 3

(2) I **found** it was not true. *text id_ 29779, Course 3*

I have defined the function of the 'find that SV' construction in the sentences (1) and (2) as stance, but the use of the linguistic device to convey the stance function is not in the way typically considered as attitudinal or epistemological for describing the writer's experience. For example, Sentence (2) from Text 29779 uses personal experiences to support the writer's point, but rather than for persuasive purposes, this essay intends to narrate past personal experiences. Notably, it is not a usual category reported in previous research (Hyland, 2005). Rather, it seems to be associated with the personal stance with persuasive function in Dimension 1 in Section 4.5.

(3) It **found** that all aspects of the energy system have various degrees of vulnerability. *text id_ 29779, Course 3*

Text 29779 is presented to compare the difference between 'find SV' and 'find that SV' constructions. This text uses third-person pronouns, probably past tense (grammatical error makes it hard to be sure about the tense), which are used to convey an epistemological stance, delivering an impersonal tone to the statement. These sentences were used to support the argument in an argumentative essay.

(4) However Barca was not far behind on the race looking back at the history of the team we find that most of the best players in the world have played some parts of their careers in Barca starting with the legendary goalkeeper Zamora all the way to Johan Cruijff and Maradona. *text id_ 37724, Course 3*

Another example of 'find that SV' construction, text 37724, uses an epistemic stance, which is typically expected of the verb 'find', often associated with impersonal, scientific fact provision. This argumentative essay uses 'find' construction to support the writer's argument based on empirical evidence to support the argument. It uses a combination of features not typically found in either expert academic writing or student essays - but it mixes the persuasive function by 'we' and neutral tone by 'find' - using this device to persuasively present empirical evidence. The use of a plural form of the first-person pronoun seems to be intended to involve the reader and give the impression of general acceptance of the writer’s claim.

6.4.3.2 Construction 2: Feel SV / Feel that SV

(5) I **<u>feel</u>** that the soldier is fair. *text id_ 3448, Course 2*

Text 3448, a persuasive argumentative essay employing personal narrative by using ‘feel that SV’ construction, makes a great deal of use of ‘feel’ with first-person pronouns. In order to argue the writer’s point, they make use of personal experience. Typical of the personal narrative found in Dimension 2 in this corpus as discussed in CH4.

(6) I **<u>feel</u>** that I have never succeeded in my life. *text id_ 37679, Course 4*

Text 37679, another argumentative essay, makes use of personal narrative employing ‘feel that SV’ constructions, mainly presenting personal opinion based on personal experience. In general, this construction was often used in a personal narrative to present opinion on a personal topic - for that purpose, personal stance was extensively used. In general, the constructions using ‘feel’, a very general and frequent mental verb, are used to express subjective verification of an experience, opinion or statement not always the most suitable choice. The texts represented here seem to use it for many different situations where more specific verbs might have presented the ideas more accurately.

(7) some people <u>feel</u> that the government should not do that and let them have completed freedom to access the Internet. *text id_ 12215, Course 3*

Text 12215 is an argumentative essay that presents both the pros and cons of a social issue before taking a side. Interestingly, the writer uses ‘feel’ extensively instead of verbs like ‘think’ to convey an impersonal tone when stating social phenomena.

This combination of an emotional verb *feel* aiming for an impersonal tone is unusual. The topic itself is neutral and objective, and the writer isn't expressing personal opinions or feelings, but rather presenting facts and social observations. Despite aiming for impartiality, the use of ‘feel’ introduces a personal stance marker.

It's important to note that this is not a grammatical error, but rather a choice of verb that might be considered ineffective or inappropriate. Using ‘feel’ weakens the persuasive effect of the statements. This could potentially be evidence of the writer being in an **earlier stage of development** in their writing skills, where they are using a grammatically correct structure but haven't yet mastered the stylistic nuances of choosing more formal or impactful verbs for specific situations.

6.4.3.3 Construction 3: realize SV / realize that SV

(8) I <u>**realized**</u> that I do not like my major. *text id: 4102,1 Level 5, Course*

Text 12215, a personal narrative, uses ‘realize’, for attitudinal stance function. In general, this construction was very infrequent in the corpus, with only eight total occurrences with only one in each text. Moreover, they are all found in Course 2 or 3, which means they were not used in Course 1. Text 41021 is also associated with a similar pattern. . The data is inconclusive about whether the absence of this attitudinal stance construction in Course 1 reflects a deliberate communicative choice by the learners or simply writing topics that didn't call for it. Further investigation is needed to determine if the increased use of this construction in later courses signifies a development in writing sophistication.

6.4.3.4 Construction 4: Say SV / Say that SV

(9) About one-third of those hit by viruses <u>said</u> the infection forced them to reformat their hard drives, 16% <u>said</u> they lost important data, and 8% had to replace hardware. *text id_ 25874, Course 3*

Text 25824 makes extensive use of *say* construction (e.g., *said that*). Unlike direct quotations, these phrases don't indicate the source of the information explicitly. While these sentences might not directly reflect the writer's own opinions or evaluations, they effectively report the views or findings of others.

In this particular argumentative essay, the writer employs this construction to report a statistic. Doing so contributes to an impersonal and neutral tone, which is appropriate for the objective presentation of factual information in an argumentative essay. The impersonal nature of the topic itself further justifies the use of a third-person voice, reinforcing the objectivity of the reported statement.

(10) Indeed, my friends said that I had had some successes in my career, which I can not accept they were. *text id_ 3448, Course 2*

Text 3448, an argumentative essay on the topic of the definition of success, makes use of a personal narrative to support the writer's points of view. *Say that* construction is used to report personal experience, conveying the impression of empathy and persuasiveness, involving the reader in the situation. It is a rather personal topic but still uses a third-person voice to convince the reader of the writer's argument.

However, despite this initial observation, a more in-depth analysis with a larger sample size is needed to determine the function of this construction and its relationship to the distinction between impersonal reporting and personal stance. Table 6-10 shows the frequency of 'say that' and 'say that-deleted' constructions across five courses.

Table 6-10. 'SAY that SV' and 'SAY SV' constructions across Courses

	say that SV	say SV
Course 1	13 (0)[31]	10 (0)
Course 2	48 (22)	27 (11)
Course 3	52 (48)	45 (42)
Course 4	6 (6)	0
Course 5	2 (2)	0
total	121 (78)	82 (53)

As seen in Table 6-10, 'say that SV' construction is preferred to 'say SV' construction overall, which is notably observed in Level 5; out of a total of 131, the Level 5 students produced 78 'say that SV' constructions compared to 53 'say SV' constructions. Moreover, all the 8 instances of the instances were 'say that' but no *that*-deleted construction was found in courses 4 and 5. This indicates that in addition to the overall preference to 'say that' construction has persisted over time. However, heed is needed as both 'say that SV' and 'say SV' constructions across Courses 1 ~ 3 increase in frequencies. Table 6-11

[31]The values in parentheses: the numbers of the construction written in Level 5

shows the sub-construction types of ‘say that SV’ and ‘say SV’ constructions found in Level 5 only.

Table 6-11. ‘SAY (that) SV’ construction subtypes in Level 5

subtypes	say that SV	say SV
third-person pronoun + say that SV	37	46
first-person pronoun + say that SV	18	7
it is said that SV	18	0
it is [adjective] to say that	5	0
total	78	53

Table 6-11 reveals a greater variety of sub-constructions for 'say that SV' compared to 'say SV' constructions. 'Say SV' only has two sub-types, where the subject can be a third-person or first-person pronoun. In contrast, 'say that SV' has four varieties, including ‘it is said that’ and ‘it is [adjective] to say that.’

This diversity in 'say that SV' constructions suggests they can be used to express more subtle meanings. For instance, ‘it is said that’ is a more formal construction, often used in academic writing (Larsson & Kaatari, 2019; 2020). The increasing preference for 'say that SV' constructions over time, along with their variety, demonstrates a more sophisticated use of *that*-clauses introduced by ‘say.’

6.4.3.5 Construction 5: Mean SV / Mean that SV

(11) That **mean**s it is necessary and so much better to travel to the country where a language is spoken. *text id_* 42148, Course 1

Text 42148 is an argumentative essay that explores the contrasting experiences of learning new languages in one's home country versus a native-speaking country. The essay maintains a neutral tone while presenting the writer's interpretation and opinion on this social phenomenon.

The writer strategically uses the ‘say that’ construction to build their argument. This construction allows them to integrate their stance on the topic by interpreting the social phenomenon through their own perspective. In essence, the writer uses these phrases to present evidence or reasoning that supports their argument about the challenges of language learning in different environments.

6.4.3.6 Construction 6: Know SV / Know that SV

(12) Next, In the case that you know your baby has something wrong with it that is terminal and you know the baby would not live long after he/she is born. *text id_ 23877, Course 2*

Text 23877 uses the 'know SV' construction for a persuasive purpose, involving the reader in the writer's argument. It is notable that 'that' deletion, known as a marker of spoken register, is much used. A similar function has been found in Dimension 1.

(13) How do you know that family does not have covert intention? text id_ 37694, course 3

Text 37694, an argumentative essay on an ethical, social issue (Euthanasia), uses the 'know that SV' construction in a rhetorical style to put an emphasis on the statement and to involve the reader in a hypothetical situation. This is rhetorical in the way that the writer argues it is a conceivable situation which is undesirable. This use supports the writer's argument. This use may be categorised as an attitudinal stance, used in a way that is not typical but rhetoric. Notably, it uses that complementiser - typically expected in written registers

6.4.3.7 Construction 8: Think SV:+ Feel that SV

(14) I think an abortion is better than just abandoning a child for someone else to come along, hopefully, in time, to take care of and save the child. *text id_ 23877, Course 2*

Sentence (14) shows a typical use of personal stance, with the most strongly associated verb 'think' to that complements in the corpus. This construction was used to present the writer's own opinion, oftentimes in an argumentative essay. Personal stance/attitudinal stance associated with first-person pronoun, present tense and think SV construction with that deletion. The that-deletion is interpreted as more of a spoken characteristic.

6.4.3.8. Discussion & Conclusion

The analysis revealed that verb-that clause constructions, particularly those involving verbs like *think*, *see*, and *believe*, are prevalent in the corpus. A qualitative examination of these constructions indicated a strong association with the expression of personal stance, suggesting that they primarily function to convey the writer's attitudes, opinions, or feelings. This finding supports the initial hypothesis that general formulaic phrases play a crucial role in the early developmental stages. Notably, the dominance of attitudinal stance throughout the corpus is evident in the frequent use of that-complement constructions.

The increase in '*think'* construction across Courses indicates an expected developmental pattern. This construction can be used for a personal stance, in which the author can discuss non-personal, abstract subjects. The use of personal stance, as marked as a characteristic of student register (e.g., Gardner et al., 2019), is an essential skill in academic writing genres. However, the overall score patterns for different verbs exhibited inconsistencies. Moreover, the findings do not provide evidence of the emergence of language use for epistemic stance function. As Hyland and Tse (2005) noted, epistemic stance is a prevalent function used in academic papers, mostly realised using verbal predicate forms (verbs controlling that-clauses). This highlights the need for further investigation into the factors influencing verb choice within these constructions, including topic selection and the evolving nature of stance expression. While the study provides evidence for the prevalence of attitudinal stance, further research is necessary to elucidate the complex interplay between verb choice, topic selection, and the development of epistemic stance.

6.4.4 Sophisticated Constructions Across Five Courses

I have discussed in Section 6.4.3 the predominant use of the few verbs controlling *that*-clauses and the personal stance as typically associated with these constructions.

Considering the third research question of looking into the variations of potentially more sophisticated constructions over time, the collostruction scores of these verbs were examined in each sub-corpus consisting of each time point (personal order of semesters in which each writer wrote and contributed one or more drafts). This section compares the longitudinal PELIC and the reference (PELIC Level 5) corpus regarding the selected verbs as markers of sophistication.

By comparing the collostruction score rankings between the longitudinal data and the Level 5 corpus, I aim to observe whether certain patterns of verb-that-clause construction preference emerge as writers progress. If scores become more similar, it suggests a potential shift towards usage patterns found in the target register. Mismatches can highlight unexplained factors in the developmental trajectory and distinguish these from overall trends expected at higher proficiency levels.

As noted previously, stance is one of the functions L2 writers have difficulty fully realising with diverse and appropriate expressions (e.g., Pan, 2018; Hyland & Milton, 1997). Accordingly, the assumption was that strong attraction between the highly frequent verbs and the construction would be more salient at the early stages of writing development but less so at later stages, as the reliance on these verbs in the VACs gets reduced. Therefore, I assumed that constructions headed by *more sophisticated (less frequent in general but more frequent in the target domain) head verbs* would get more frequent at later stages.

This analysis examines whether the rankings of each verb construction in the longitudinal corpus (Courses 1 to 5) and the reference corpus (Level 5) are similar. If the ranking of a construction in the Courses are similar to that of Level 5, it strengthens the likelihood that the observed variation patterns in the longitudinal data reflect expected trends. If the ranking of a construction in Courses 1 thoughout 5 changes and the direction of the variation conforms to that of Level 5, a similar conclusion is drawn. In contrast, a mismatch between the ranking of each construction in the two corpora suggests unexplained variations between these two corpora.

I selected verbs for sophistication analysis based on both their collostruction scores in the longitudinal data and their general frequency in academic writing (as underlined verbs in Table 6-12; the academic frequency of these was considered following Biber et al., 2021b). Table 6-12 comprises the top ten collostruction scores of the reference corpus, consisting of Level 5 writing in the PELIC corpus. The semantic categories of the most strongly associated verbs with each VAC exhibited a similar pattern across the different dependent types (*that* complement clauses and *that*-deleted clauses) to that of the longitudinal data presented in Table 6-8. In other words, the Level 5 corpus also showed a consistent preference for the *that*-deleted construction and mental verbs in the list of the ten most strongly verbs to the VACs, as shown in Table 6-10. This is roughly translated as that the PELIC written corpus, regardless of their English levels, tends to prefer that-deleted construction to *that*-complement construction. As it has been noted as a characteristic of spoken registers, the preference for *that*-deleted counterpart of the prototypical construction indicates the nature of this spoken register in this written learner corpus.

Table 6-12. Controlling verbs most strongly attracted to that-clauses in the Level 5 corpus

Collostruction score rank	Verb	V that S V	V () S V	V (that) S V	Academic register frequency[32]	Semantaic category
1	think	445	335	781	*	mental V
2	believe	107	253	360	*	mental V
3	<u>say</u>	138	207	345	***	communication V
4	mean	114	129	243	*	mental V
5	<u>know</u>	97	130	226	**	mental V
6	realize	23	92	115	*	mental V
7	<u>find</u>	61	54	115	**	mental V
8	feel	37	69	107	*	mental V
9	<u>see</u>	61	34	95	**	mental V
10	argue	11	70	81	-	communication V

[32] The general frequency of the most common verbs controllling a complement that-clause in academic registers, adopted from Biber et al. (2021b, p.661)
***: over 200, **: over 100, *: over 20 (occurrences per million words)

Nine verbs were selected from verbs in Table 6-8, based on their collostruction score in the longitudinal corpus. These verbs are in general the most strongly attracted verbs to the target VACs (verbs controlling *that-complement* clauses) in the longitudinal data. In addition, the four underlined verbs are selected due to their potential significance in terms of sophistication in reference to the Level 5 data and general academic registers: *think, say, mean, know, find*. All the nine most strongly attracted verbs in the longitudinal corpus were also among the most strongly attracted verbs in the target construction in the reference data, except for the relative rank order of these verbs, with the exception of 'argue'.

The collostruction scores were computed based on the four sub-groups of the data: three Courses in the longitudinal corpus (Courses 3-5 were merged to balance the sample size between Courses 1, 2 and 3-5) and one Level 5 corpus. Table 6-11 comprises collostruction score rankings of nine selected verbs in VACs across the four sub-corpora: Course 1, Course 2 and Course 3-5 in comparison to that of Level 5 data (consisting of all the texts written at Level 5, including the non-longitudinal texts written by students with entry-levels of 4 and 5). It also shows unnormalized (raw) frequencies of each verb in VACs and other constructions, respectively.

Table 6-13. Collostruction scores by course and level

ranking	Course 1	Course 2	Course 3-5	Level5
1	realize (3/8)[33]	realize (9/14)	realize (30/44)	think
2	say (19/40)	say (56/135)	say (84/134)	say
3	mean (13/59)	mean (32/148)	mean (41/153)	mean
4	find (6/88)	see (9/228)	think (138/245)	know
5	learn (9/113)	find (14/253)	feel (27/178)	realize
6	know (24/137)	feel (32/296)	find (23/202	find

[33] (V frequency in VAC/ V frequency in each sub-corpus), for example, 'realize' occurs three times in VAC in Course 1 sub-corpus; and it accurs eight times in total in Course 1 sub-corpus.

7	see (8/133)	know (33/343)	see (33/253)	feel
8	think (85/182)	think (170/433)	learn (30/334)	see
9	feel (13/153)	learn (20/376)	know (47/429)	learn

Table 6-13 show that many constructions fluctuate in scores, making it difficult to conclusively confirm our initial hypothesis that very general verbs would dominate early, followed by more sophisticated verb choices later. In addition, verbs like 'realize', 'find', 'know', and 'see' are common in academic registers, yet their scores in the PELIC corpora as in Table 6-13 show mixed patterns.

While not all the verbs' collostruction scores show a clear pattern of increase or decrease across the sub-corpora, some verbs do (e.g., *think* and *realize*). While some verbs' collostruction scores don't display clear trends across the sub-corpora, others show interesting patterns that align with scores in the Level 5 texts. This suggests a potential shift in semantic function over time. Both 'think that SV' and 'realize that SV' can express an attitudinal stance. Consider the following examples.

(1) I think that physical, mathematical and numerical model with the experience could give us knowledge of our reality.
Text id: 41339 (Level 5, not included in longitudinal data)

(2) I realized that I do not like my major. *Text id: 41021 (Level 5, Course 4)*

Sentence (1) employs a first-person pronoun and *think* as a controlling verb in *that*-complement clause construction. The sentence seems to be associated with an attitudinal stance function, which was used to express a personal stance toward a non-personal proposition/argument. On the other hand, the VAC with *realize* was used to describe personal attitude about a personal event Sentence (2).

Taken together, both examples above could be associated with an attitudinal stance, employing the [V-that SV] construction. However, the difference between the constructions controlled by *think* and *realize* seems to arise from the distinction between personal vs non-personal topics in *that*-clauses. Considering the increase of the think construction and the decrease of the realize construction over time and at the Level 5 corpus, the writing seems to head toward expressing a personal stance toward more non-personal topics. The increasing use of 'think' and decreasing use of 'realize' aligns with their Level 5 ranks. This shift might indicate writers moving towards expressing personal stances on more non-personal topics.

The difference between the constructions controlled by *think* and *realize* seems to arise from the distinction between personal vs non-personal topics in *that*-clauses. Table 6-13 reveals a consistent increase in the score rank of 'think' constructions across Courses in the longitudinal corpus. This suggests a clear preference for using 'think' throughout the writing samples. While the reasons for this trend are not entirely clear, it aligns with findings presented in Chapter 4 regarding dimensional variations. There are two potential explanations for this. One possibility is that students develop a growing awareness of the attitudinal stance function (expressing attitude or opinion) associated with 'think' constructions. This could lead to their more frequent and deliberate use across courses. Another possibility is that the writing prompts themselves might influence verb choice. As students progress through the courses, the topics might increasingly require the use of attitudinal stance, favouring 'think' constructions.

6.5 Conclusions

The analysis in Section 6.4.3.9 suggests a link between verb choice (e.g*., think* vs*. realize*) and topic type (personal vs. non-personal) in formulaic language constructions. While some verbs, like 'think,' show an expected increase in use with development, the overall score patterns reveal inconsistencies. This necessitates a deeper analysis to explore the influence of factors like topic selection on verb-argument construction choices. Notably, attitudinal stance remains dominant throughout the longitudinal corpus, evidenced by the frequent use of 'that-complement' constructions.

Despite the inconsistent score trends, a qualitative analysis could be conducted to see if there's a subtle shift towards using constructions that express epistemic stance (certainty or knowledge).

This chapter focused on the highly frequent attitudinal stance function found in verb-that complement constructions within the PELIC written texts. ‘Think’ was the most frequent verb in this construction, while ‘realize’ or ‘find,’ which are generally more frequent in academic writing, were used less often. Consistent with the increasing use of ‘think’ across courses, the collostructional score of this construction was the highest for Level 5 texts. "Find" also showed a similar pattern of increasing use with development.

However, many constructions did not exhibit a consistent increase in score rank. Notably, the score rank of the ‘feel’ construction steadily increased, indicating a preference for it across Courses. Interestingly, this trend reversed at Level 5, where the ‘feel’ construction ranked 5th for Courses 3-5 but dropped to 7th at Level 5. This suggests a potential shift in writing style at the higher level. ‘Feel’ constructions are typically associated with attitudinal stance, and topic selection might influence the increased preference for them across Courses, but a decrease at Level 5 could indicate a move towards more objective language.

It's important to note that Course and Level variables are likely intertwined. Since student level generally increases across courses, the observed trend could be primarily driven by students' overall language development rather than specific course content.

Despite the limitations, this analysis provides valuable insights. It suggests that ‘think’ constructions become increasingly favored as students progress through the writing program. Further research that disentangles the effects of course content and student development could provide a more nuanced understanding of this phenomenon.

Many of the top 9 collostructions have attitudinal functional uses in context. This use of that-complement construction contrasts the findings regarding this construction in academic writing. However, this pattern reinforces the attitudinal stance function as a dominant communicative function found in the PELIC longitudinal corpus discussed in Chapter 4. This construction has been used along with personal pronouns and that deletion frequently, as exemplified in context.

The decrease in rank observed for constructions like 'realize' doesn't necessarily reflect a decrease in their overall frequency. As highlighted in the analyses of Chapters 4 and 5, topic selection and other unidentified factors can influence score rankings. This makes it challenging to definitively interpret the variations in these constructions over time based solely on rank changes.

This chapter explored the two hypotheses presented earlier regarding formulaic language development in L2 writing. The analysis employed collostructional analysis to examine verb-that clause constructions across the writing samples.

The results offer some support for the first hypothesis, suggesting that general formulaicity plays a significant role in the early stages of development. This is evident in the high association strength of frequent verb-that constructions (e.g., *think that*) in Course 1.

However, the variations observed in the rank of these constructions challenge our expectations. Previous research suggested a pattern of increasing importance for very general verbs in early stages. While some verb-that constructions, like 'think,' showed consistent use, others exhibited unexpected fluctuations. This makes it difficult to definitively conclude whether very general verbs such as 'think' are solely favored in the early stages, giving way to more sophisticated constructions later. This suggests the existing theory might require refinement to better explain the dynamics of formulaic language development within specific datasets.

The second hypothesis, focusing on register-specific formulaicity markers, also yielded mixed results. The chosen verbs (say*, find, know, see)* generally showed inconsistent patterns of increase or decrease compared to Level 5.

Despite the challenges in interpreting score variations for grammatical sophistication, a more nuanced qualitative analysis revealed a potential trend towards a stronger use of epistemic stance markers (verbs expressing certainty or knowledge) over time.

This study highlights the complexity of formulaic language development in L2 writing. While frequency plays a role, other factors like topic selection and register variation can significantly influence verb choice. Future research should consider more nuanced approaches that combine quantitative and qualitative analyses to gain a deeper understanding of this dynamic process.

Chapter 7 Conclusions

This longitudinal study investigated the development of linguistic complexity and functional features in L2 academic writing, focusing on a less explored register: adult learners of English in an Intensive English Programme (IEP). The study employed a learner corpus of academic writing collected over three to five semesters at a U.S. university. Multidimensional (MD) analysis, collostructional analysis, and mixed-effect modelling were used to analyze patterns of syntactic forms and their connection to communicative functions in the learners' writing. The study results provided evidence of how L2 learners' lexico-grammatical choices become more sophisticated for communicating functions such as stance and aligned with academic linguistic choices over time.

The findings of the MD analysis (Chapter 4) suggest that L2 learners develop greater phrasal complexity as they progress in their studies. This highlights the importance of incorporating writing tasks that encourage the use of these features in L2 writing instruction. The observed influence of writing topics on syntactic variation underlines the need for a balanced curriculum design. While learner-centered approaches that allow for topic selection can foster engagement, it's crucial to ensure exposure to writing styles that require a wider range of syntactic features.

The major features associated with functional dimensions informed by the result of the exploratory factor analysis provided information on inherent communicative purposes in texts that played key roles in these linguistic variations. In the first dimension, for instance, we have seen features such as second-person pronouns, mental verbs and infinitives associated with persuasive involvement while the negatively loaded features were associated with scientific descriptions.

In addition, the findings regarding linguistic variations marking syntactic complexity supported the assertion that L2 academic writing development is register-specific in nature. In other words, the analysis has found evidence to suggest that phrasal features associated with literate and informational writing functions become more salient over time. In addition, we have seen the significant topic effects in the linguistic variations, which I have associated with the nonlinear patterns of dimensions 1 score across time points. This may be unpacked in further research to provide more interesting insights into how these linguistic variations interact with the programme's learner-centred approach to writing and process writing approaches.

The key linguistic forms with stance functions were further analysed. The complementary analyses focusing on *that*-complement clauses analysed the variation of the stance functions associated with these constructions in the L2 writing over time. These analyses were based on the initial findings that emerged from exploratory factor analysis, including the influence of situational variables, operationalised as writing topics for L2 academic writing has provided contextual knowledge about the effects of this variable.

The collostruction analysis discussed in Chapter 6 has provided insight into the variation of head verbs in verb-that clause constructions in reference to register-appropriate writing devices reported in previous empirical findings (e.g., Biber, 2006). The shift toward non-personal writing types, for example, has shown evidence of transitioning toward syntactically sophisticated writing over time. In addition, a quantitative analysis of the noun- and adjective-clause constructions informed us about the estimation of the distribution of these two constructions in L2 writing developmental trajectories. The findings showed that these constructions were extremely infrequent in L2 writing, and albeit statistically insignificant, the frequencies increased over time. A qualitative inspection of texts with reference to English-level descriptors may help understand the quantitative findings reported in the thesis. For example, written texts grouped by the same courses may be cross-checked with reference to the L2 level, to understand the inherent writing purposes in line with the assigned levels.

As has been discussed regarding 'say that-clause' and 'say S V' constructions, these structurally elaborate constructions (as discussed in Section 6-N) tend to increase in frequencies across courses. While a preference for 'say that-clause' to 'say SV' construction was clear throughout the courses, evidence for the differences as to when and why one construction over the other is preferred is limited.

The analysis consistently revealed, across Chapters 4, 5, and 6, that writing topics significantly influence the linguistic choices L2 learners make. For instance, the link between verb choice and topic type highlighted in Section 6.5 underscores the importance of integrating instruction on how verb selection can impact the tone and formality of writing. The observed influence of topic selection on linguisctic variation highlights the need for a balanced curriculum design that incorporates both learner-centred approaches and exposure to a wider range of writing styles requiring diverse syntactic features.

Similarly, the observed shift towards a more objective writing style at the later courses (e.g., Course 5) in Section 6.5 further emphasizes the need for incorporating writing tasks that encourage the use of a neutral and detached tone. These findings highlight the importance of curriculum design that considers the interplay between topic selection and the development of appropriate linguistic features in L2 writing.

Due to the small size of the samples for this construction, the found patterns regarding these constructions in this L2 writing data should be interpreted with caution. However, two points can be made, at least for this L2 writing corpus. First, the preference for some of the constructions analysed qualitatively (e.g., 'say that-clause' construction) is in line with the previous literature marking this construction as an academic written register compared to that-deletion. Second, the analyses showed more various sub-construction types in later courses, indicating more sophisticated linguistic use regarding this construction (e.g., see Section 4.4.2 and Section 6.4.3.4).

In summary, the research reported in this thesis contributes empirically based linguistic descriptions of university-level English language writing development for adult L2 writers. The empirical findings from large data have provided insight for patterning register-specific language use made in writing. Although the patterns were not always statistically significant, the analyses provided evidence that the patterns generally aligned with the previous findings. Specifically, the analysis results suggest that form-function mapping is a useful framework for identifying linguistic patterns in L2 writing, as a unique and stand-on register. Previous research (e.g., Vyatkina, 2013; Díez-Bedmar & Pérez-Paredes, 2020) did not explore the hypotheses based on functional perspectives. Even Biber et al. (2021a), proposing the written register complexity framework, did not provide a functional framework for these complexity features. I propose that a functional framework is useful in understanding a corpus collected within a specific situational context, providing the information for the shared communicative purposes linked to the main writing topics and in this regard enabling more usage-based perspective interpretation of the corpus analysis. It is believed that the series of studies discussed in this thesis has contributed to adding to our knowledge of L2 writing as a unique register and its developmental course as the writers progress to learn writing in English.

The findings also show the significance of our understanding of the topic effects and individual differences (including intra- and inter-variabilities) in writing over time. Therefore, the findings have added further support for the claim that the syntactic variation in longitudinal L2 writing is strongly influenced by various factors, including writer-specific factors and situational contexts. The writers make strategic choices to accomplish tasks while being economical and efficient in terms of linguistic choice. The evidence of language use has offered information for unpacking the complex nature of language development, which contributes to making principled hypotheses for L2 writing development.

Moreover, the linguistic descriptions of large textual data provide essential information for making principled pedagogical decisions. The empirical data about linguistic development can inform the design and timing of intervention and the refinement of developmental theories from an integrated systemic perspective (Harring & Houser, 2016, p.291). The longitudinal analyses presented in this thesis could be added to shed light on the importance of considering these factors in understanding writing development for L2 learners.

Understanding L2 register variations aids in the development of syllabi and materials aligned with L2 learners' communicative goals at different developmental stages. Moreover, comprehending the relationship between form and function in learner writing is essential for effective EFL instruction. This knowledge empowers educators to tailor teaching strategies to the specific needs of learners acquiring English as a foreign language (Ortega, 2015. For example, in the classroom, teachers may decide to provide learners with linguistic language forms used for a core communicative function in academic writing. This knowledge allows educators to tailor their teaching strategies to the specific needs of learners acquiring English as a foreign language (Ortega, 2015).

While admitting that this reconceptualisation of writing development is beyond the scope of this thesis, this seems to require careful consideration as to assessing the relative importance of different aspects of language variations observed in the data, if wishing to make relevance to the current pedagogical concerns. For example, future studies should consider incorporating seemingly conflicting hypotheses and findings to disentangle the interaction between language use and writer-specific and situational variables. Specifically, many previous studies have found that noun complexity is a marker of l2 writing development (e.g., Vyatkina, 2013; Pérez-Paredes & Díez-Bedmar, 2018; Díez-Bedmar & Pérez-Paredes, 2020). However, the findings were not always compatible with one perspective among possible competing theories. For example, Pérez-Paredes and Díez-Bedmar's (2018) findings suggest that while older learners write sentences that are more complex, syntactic complexity at the clause level may require further years of instruction to develop. This means that the usage-based expectation of the noun complexity development is not likely to be generalised for L2 learners, probably because the use of this aspect of language would be also influenced by the complexity of concepts, tasks and cognitive processes of the writer. This multidimensionality of linguistic complexity in English academic writing, as has been noted in previous literature (e.g., Pérez-Paredes & Díez-Bedmar, 2018; Biber et al., 2021a) and reiterated in this thesis, should be addressed by incorporating more comprehensive linguistic selections and better-informed linguistic hypotheses that can capture these phenomena without imposing single-sided theories.

Finally, the analysis presented in this study is constrained by several factors for its methodological limitations. Firstly, the relatively small sample size necessitates further investigation with larger and more diverse datasets to enhance the generalizability of the findings. It is essential to acknowledge the limitations posed by the extremely small sample size of Course 5 (consisting of data provided by only one writer). The findings and interpretations related to this course were limited due to its potential lack of representativeness. Future research with a larger sample size for all sub-corpora is necessary to draw more definitive conclusions. Specifically, the limited increase in noun- and adjective-clause use observed in Chapter 5 warrants additional exploration with expanded sample sizes or extended observation periods to corroborate the findings.

Additionally, the adequacy of text length for robust analysis is a critical consideration. While short texts can be analysed using methods used in this study (e.g., MD analysis), the reliability of normalised rates and the occurrence of zero values can be problematic, particularly for infrequent features. Although MD analysis can accommodate smaller text lengths, the optimal text length for the employed methodologies remains a subject for further investigation. Given that the majority of texts in our corpus fall below the recommended 1000-word threshold for reliable MDA, the potential impact on the study's findings deserves careful attention.

One crucial limitation of the longitudinal analyses in this thesis is that they did not trace the linguistic variation in individual trajectories of academic writing. Although the results suggest the importance of situational and writer-specific factors in syntactic variations in the data, these were not clearly modelled to map the language production in the English programme onto the academic writing registers. In this regard, the analysis results of this thesis open many areas for investigating learner writing with more granular linguistic variables. For example, the generally insignificant results of prepositional constructions analysed in Chapter 5 may be worth further investigating with respect to whether the predominant use of adjectives becomes replaced with more nominal options as students increase their academic literacy. More granular measures that reflect different prepositional constructions indicative of functional markers may prove to be helpful for understanding a variation of this particular construction in L2 writing development. The final note for further research from the observations made in these analyses is that the notion of 'elaboration' proved a significant factor. Notable interaction effects, including individual differences and writing topics with syntactic variations, call for further investigations within a systematic framework to elucidate how the individual factors moderate the variations and how the writing topics mediate the syntactic variations.

References

Ädel, A. (2015). Variability in learner corpora. . In G. G. S. Granger, & F. Meunier (Ed.), *The Cambridge Handbook of Learner Corpus Research (Cambridge Handbooks in Language and Linguistics)* (pp. 401-422). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.018

Aguado-Jiménez, P., Pérez-Paredes, P., & Sánchez, P. (2012). Exploring the use of multidimensional analysis of learner language to promote register awareness. System, 40(1), 90-103.

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. Corpora, 4(1), 71-83.

Amante-Nochefranca, G., Orbase-Sandal, O., Alieto, E. O., Laput, I. U., Albani, S. E., Lucas, R. I., & Tanpoco, M. (2023). AI-Assisted English Language Learning and Teaching in a Developing Country: An Investigation of ESI Student's Beliefs and Challenges. The International Conference on Artificial Intelligence and Smart Environment, pp.281-289.

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research*, *30*(2), 141-161.

Atak, N., & Saricaoglu, A. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect. Assessing Writing, 47, 100506. https://doi.org/https://doi.org/10.1016/j.asw.2020.100506

Bartoń, K. (2017). MuMIn: Multi-model inference. R package version 1.40.0. Available online at https://cran.r-project.org/package=MuMIn

Bartoń, K. (2017). MuMIn: Multi-model inference. R package version 1.40.0. Available online at https://cran.r-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Berzak, Y., Kenney, J., Spadine, C., Wang, J. X., Lam, L., Mori, K. S., Garza, S., & Katz, B. (2016). Universal dependencies for learner English. arXiv preprint arXiv:1605.04278.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (1989). *A typology of English texts*. Linguistics, 27(1), 3–43.

Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. Literary and Linguistic Computing, 5, 257–269.

Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis*. Discourse Processes*, 15(2), 133-163. https://doi.org/10.1080/01638539209544806

Biber, D. (1993). Representativeness in corpus design. Literary and Linguistic Computing, 8, 243–257.

Biber, D. (1995). Dimensions of register variation: A cross-linguistic comparison. Cambridge University Press.

Biber, D. (2004) 'Historical patterns for the grammatical marking of stance: a cross-register comparison', *Journal of Historical Pragmatics,* 5, 107–136.

Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=623257

Biber, D. (2019). Multi-Dimensional Analysis: A Historical Synopsis. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-Dimensional Analysis : Research Methods and Current Issues* (1 ed., pp. 11-26). Bloomsbury Academic. http://www.bloomsburycollections.com/book/multi-dimensional-analysis-research-methods-and-current-issues/ch1-multi-dimensional-analysis-a-historical-synopsis/

Biber, D. (2021). Stance and Grammatical Complexity*: An Unlikely Partnership Discovered through Corpus Analysis. In *The Register-Functional Approach to Grammatical Complexity* (pp. 113-128). Routledge.

Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=471982

Biber, D., & Gray, B. (2010[2021]). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *In Biber, Reppen, Staples & Egbert (Eds). The Register-Functional Approach to Grammatical Complexity: Theoretical Foundation, Descriptive Research Findings, Application*. Routledge. Doi:10.4324/9781003087991-7

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. English Language and Linguistics, 15(2), 223-250. https://doi.org/10.1017/S1360674311000025

Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL ibt® test: A lexico-grammatical analysis. *ETS Research Report Series*, *2013*(1), i-128. https://doi.org/10.1002/j.2333-8504.2013.tb02311.x

Biber, D., & Gray, B. (2016). Phrasal versus clausal discourse styles: A synchronic grammatical description of academic writing contrasted with other registers. In *Grammatical Complexity in Academic English: Linguistic Change in Writing* (pp. 67-124). https://doi.org/10.1017/CBO9780511920776.003 (Cambridge: Cambridge University Press)

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge Approaches to Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511804489

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. TESOL Quarterly, 36(1), 9-48.

Biber, D., & Hared, M. (1992). Dimensions of register variation in Somali. Language Variation and Change, 4(1), 41-75.

Biber, D., Gray, B., & Poonpon, K. (2011). Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development? *TESOL Quarterly*, 45(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Biber, D., Gray, B., & Staples, S. (2014). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, *37*, 639–668.

Biber, D., Gray, B., Staples, S., & Egbert, J. (2020a). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. Journal of English for Academic Purposes, 46, 100869. https://doi.org/10.1016/j.jeap.2020.100869

Biber, D., Gray, B., Staples, S., & Egbert, J. (2021a). *The Register-Functional Approach to Grammatical Complexity: Theoretical Foundation, Descriptive Research Findings, Application.* (1st ed.). Routledge. https://doi.org/https://doi-org.uoelibrary.idm.oclc.org/10.4324/9781003087991

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021b). Grammar of Spoken and Written English. John Benjamins. https://doi.org/https://doi.org/10.1075/z.232

Biber, D., Reppen, R., Staples, S. & Egbert, J. (2020b). Exploring the longitudinal development of grammatical complexity in the disciplinary writing of L2-English university students. *International Journal of Learner Corpus Research*, 6(1), 38-71. https://doi.org/10.1075/ijlcr.18007.bib

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: investigating language structure and use . Cambridge: Cambridge University Press.

Bird, S., Klein, E. & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media Inc. https://www.nltk.org/book/

British Educational Research Association [BERA] (2018) Ethical Guidelines for Educational Research, fourth edition, London. https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018

Brown, P., & Fraser, C. . (1979). Speech as a marker of situation. In K. R. Scherer. & H. Giles (Ed.), *Social markers in speech* (pp. 33-62). Cambridge University Press.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 23-46).

Bulté, B., & Housen, A. (2014). Conceptualising and measuring short-term changes in L2 writing complexity. Journal of Second Language Writing, 26, 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Soft- ware, 80(1), 1–28.

Carter, R. (2004). Grammar and spoken English. *Applying English grammar: Functional and corpus approaches*, 25-39.

Casal, J. E., & Lee, J. J. 2019. Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing, 44*, 51–62.

Chafe, W. (1986). Evidentiality in English conversation and academic writing. Evidentiality: The linguistic coding of epistemology, 20, 261-272.

Charles, M. (2007). Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction. English for Specific Purposes, 26(2), 203-218. https://doi.org/https://doi.org/10.1016/j.esp.2006.08.004

Chen, D., & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. In Y. Marton (Ed.), The 2014 Conference on Empirical Methods in Natural Language Processing (pp. 740–750). Stroudsburg.

Chomsky, N. (1988). Generative grammar. *Studies in English linguistics and literature*.

Christie. (2012). A Language Theory for Educational Practice. *Language learning*, *62*(s1), 1-31. https://doi.org/https://doi.org/10.1111/j.1467-9922.2012.00678.x

Clarke, I., McEnery, T., & Brookes, G. (2021). Multiple correspondence analysis, newspaper discourse and subregister: A case study of discourses of Islam in the British press. Register Studies, 3(1), 144-171.

Council of Europe and Association of Language Testers in Europe. (2011). Manual for language test development and examining: For use with the CEFR. Strasbourg, Council of Europe. https://www.alte.org/resources/Documents/ManualLanguageTest-Alte2011_EN.pdf. Accessed May 14, 2021.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). To Aggregate or Not? Linguistic Features in Automatic Essay Scoring and Feedback Systems. *Grantee Submission*, 8(1).

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. Written Communication, 28(3), 282-311.

Crosthwaite, P. (2016). A longitudinal multi-dimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes*, *22*, 166-178.

Crystal, D (1995) The Cambridge Encyclopedia of the English Language, Cambridge: Cambridge University Press.

De Marneffe, M. C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., & Manning, C. D. (2014). Universal Stanford dependencies: A cross-linguistic typology. LREC Vol. 14, pp. 4585-4592.

Díez-Bedmar, M.B. and Pérez-Paredes, P. (2020) Noun phrase complexity in young Spanish EFL learners’ writing. Complementing syntactic complexity indices with corpus analyses. International Journal of Corpus Linguistics, 25,1,1-33.

Durrant, P. (2018). Formulaic language in English for academic purposes. in Siyanova-Chanturia & Pellicer-Sánchez (Eds.) *Understanding formulaic language: A second language acquisition perspective*, 211-227. London: Routledge.

Durrant, P. (2022). *Corpus Linguistics for Writing Development: A Guide for Research* (1st ed.). Routledge. https://doi.org/https://doi-org.uoelibrary.idm.oclc.org/10.4324/9781003152682

Durrant, P., & Brenchley, M. (2022). Development of noun phrase complexity across genres in children's writing. Applied Linguistics. https://doi.org/10.1093/applin/amac032

Durrant, P., Brenchley, M., & Clarkson, R. (2020). Syntactic development across genres in children's writing: the case of adverbial clauses. Journal of Writing Research, 12(vol. 12 issue 2), 419-452. https://doi.org/10.17239/jowr-2020.12.02.04

Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding Development and Proficiency in Writing: Quantitative Corpus Linguistic Approaches*. Cambridge: Cambridge University Press. doi:10.1017/9781108770101

Durrant, P., & Mathews-Ayadinli. J. (2011).A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, *30*, 58–72.

Durrant, P., & Schmitt, N. (2009).To what extent do native and non-native writers make use of collocations? *IRAL*, *47*, 157–177.

Egbert, J., Biber, D., & Gray, B. (2022). Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness. Cambridge University Press. https://doi.org/DOI: 10.1017/9781316584880

Egbert, J., & Staples, S. (2019). Doing Multi-Dimensional Analysis in SPSS, SAS, and R. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-Dimensional Analysis : Research Methods and Current Issues* (1 ed., pp. 125-144). Bloomsbury Academic. h

Ellis, N. C. (2002). Frequency effects in language processing. Studies in Second Language Acquisition, 24, 143–188.

Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal*, *93*(3), 370–386.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. In: De Gruyter Mouton.

Ellis, N. C., Römer, U., & Brook O'Donnell, M. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Wiley-Blackwell.

Ervin-Tripp, S. (1972). On Sociolinguistic Rules: Alternation and Co-occurrence. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 213–250). London: Holt, Rinehart and Winston.

Explosion AI. (2018). *SpaCy language models* [Version 3.0]. Retrieved from https://spacy.io/models/en#en_core_web_sm

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological methods, 4(3), 272.

Fisher, R. A. (1934). Two New Properties of Mathematical Likelihood. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 144(852), 285–307.

Friginal, E., & Mustafa, S. S. (2016). A comparison of US-based and Iraqi English research article abstracts using corpora. Journal of English for Academic Purposes, 25, 45-57.

Friginal, E., & Weigle, S. (2014). Exploring multiple profiles of L2 writing using multi-dimensional analysis. Journal of Second Language Writing, 26, 80-95.

Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, Level, Genre: Integrating Situational Perspectives in a New MD Analysis of University Student Writing. *Applied Linguistics*, 40(4), 646-674. https://doi.org/10.1093/applin/amy005

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219-224. https://doi.org/10.1016/s1364-6613(03)00080-9

Gorsuch, R. L. (2013). *Factor analysis*. (2nd ed.) L. Erlbaum Associates.

Goulart, L. (2021). Register variation in L1 and L2 student writing: A multidimensional analysis. *Register Studies*, *3*(1), 115-143.

Granger, S. (2018) Formulaic sequences in learner corpora. In Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds). *Understanding Formulaic Language: A Second Language Acquisition Perspective* (1st ed.). NY: Routledge. https://doi.org/https://doi-org.uoelibrary.idm.oclc.org/10.4324/9781315206615

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 27–49). Amsterdam and Philadelphia: John Benjamins.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. Journal of second language writing, 9(2), 123-145.

Gray, B. E. (2011). Exploring academic writing through corpus linguistics: When discipline tells only part of the story [Ph.D dissertation, Northern Arizona University].

Gray, B. (2019). Tagging and Counting Linguistic Features for Multi-Dimensional Analysis. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-Dimensional Analysis : Research Methods and Current Issues* (1 ed., pp. 43-66). Bloomsbury Academic.

Gray, B., & Biber, D. (2012). Current conceptions of stance. In Hyland, K., Guinda, C. S., & Sancho Guinda, C. (Eds). *Stance and Voice in Written*

Academic Genres. 15-33. Palgrave Macmillan UK. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=1039413

Gray, B, Cotos, E., & Smith, J. (2020). Combining rhetorical move analysis with multi-dimensional analysis Research writing across disciplines. In U. Römer, V. Cortes & E. Friginal (Eds.), Advances in Corpus-based Research on Academic Writing(pp. 139-168). John Benjamins Publishing Company.

Gray, B., Geluso, J., & Nguyen, P. (2019). *The Longitudinal Development of Grammatical Complexity at the Phrasal and Clausal Levels in Spoken and Written Responses to the TOEFL iBT® Test (TOEFL Research Report No. RR-90)*. Educational Testing Service.

Gray, B., Staples, S., & Egbert, J., & Biber, D. (2021). Investigating lexico-grammatical complexity in L1 and L2 university student writing across genres, disciplines, and levels. Presentation at American Association of Applied Linguistics Conference (online)

Gregg, N., Coleman, C., Stennett, R. B., & Davis, M. (2002). Discourse complexity of college writers with and without disabilities: A multi-dimensional analysis. Journal of Learning Disabilities, 35(1), 23-38.

Gries, S. T. (2014). Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics. Theory and Data in Cognitive Linguistics, 67, 15.

Gries, S. T. (2015a, cited in Kyle, 2016). Personal Communication.

Gries, S. Th. (2015b). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, *10*(1), 95-125.

Gries, S. T. (2023). Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. Cognitive Semantics, 9(3), 351-386.

Gries, S. T., & Durrant, P. (2020). Analyzing Co-occurrence Data. In (pp. 141-159). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_7

Gries, S. T., & Ellis, N. C. (2015). Statistical Measures for Usage-Based Linguistics. *Language learning*, *65*(S1), 228-255. https://doi.org/10.1111/lang.12119

Gries, S. T., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *16*(4), 635-676. https://doi.org/doi:10.1515/cogl.2005.16.4.635

Guadagnoli, E. and Velicer, W. F. (1988) Relation to sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), pp. 265-275.

Guo, R., & Ellis, N. C. (2021). Language Usage and Second Language Morphosyntax: Effects of Availability, Reliability, and Formulaicity [Original Research]. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.582259

Halliday, M. A. K. (1979). *Linguistic Studies of Text and Discourse*. London: Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. (2014). An introduction to functional grammar. Routledge.

Halliday, M. A. K., & Webster, J. J. (2004). *The Language of Science: Volume 5*. Bloomsbury Publishing Plc.

Hansen, B. (2022). Probability and statistics for economists. Princeton University Press.

Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. Journal of English for Academic Purposes, 22, 119-131.

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). Corpora, 8(2), 183–207. http://dx.doi.org/10.3366/cor.2013.0040

Harring, J. R., & Houser, A. (2016). Longitudinal models for repeated measures data. In A. Rupp, Leighton, JP (Ed.), *The Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications* (1st ed., pp. 267-296). John Wiley & Sons, Inc.

Housen, A. (2002). A corpus-based study of the L2 acquisition of the English verb system. In S. Granger, J. Hung, S. Petch-Tyson, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=622338

Hudson, R. (2009). Measuring maturity. In Beard, R., Riley, J., Myhill, D., & Nystrand, M. (Eds.). *The SAGE handbook of writing development (pp.349-362)*. Sage Publications, https://www.doi.org/10.4135/9780857021069.n24

Hughes, R., & McCarthy, M. (1998). From sentence to discourse: Discourse grammar and English language teaching. *Tesol Quarterly*, *32*(2), 263-287.

Hunston., S. (1993) ‘Evaluation and ideology in scientific writing’ in Ghadessy, M. (ed.) *Register analysis: theory and practice* (London: Pinter).

Hunston, S. (2000). Evaluation: an introduction. In Hunston, S. and G. Thompson (eds). Evaluation in text: Authorial stance and the construction of discourse. Oxford University Press.

Hunston, S. and G. Thompson. (2000). *Evaluation in text: authorial stance and the construction of discourse.* Oxford: Oxford University Press.

Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.

Hunt, K. W. (1966). Sentence structures used by superior students in grades four and twelve, and by adults (Report No. CRP-5-0313). Tallahassee: Florida State University. (ERIC Document Reproduction Service No. ED 010 047)

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the society for research in child development*, *35*(1), iii-67.

Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, 13(3), 239-256.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied linguistics*, *17*(4), 433-454.

Hyland, F. (2000). ESL writers and feedback: Giving more autonomy to students. *Language Teaching Research*, 4(1), 33-54. Sage Publications.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, *7*(2), 173-192.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of second language writing*, *6*(2), 183-205.

Hyland, K., & Tse, P. (2005). Hooking the reader: a corpus study of evaluative that in abstracts. English for Specific Purposes, 24(2), 123-139. https://doi.org/https://doi.org/10.1016/j.esp.2004.02.002

Hymes, D. (2003). *Foundations in sociolinguistics: An ethnographic approach*. Psychology Press.

Irvine, J. T. (1979). Formality and informality in communicative events. *American Anthropologist*, *81*(4), 773-790.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, *4*(1), 51-69. https://doi.org/10.1016/1060-3743(95)90023-3

Ishikawa, S. (2019). The ICNALE: The international corpus network of Asian learners of English. Retrieved from http://language.sakura.ne.jp/icnale/.

Issitt, S. (2017). Evaluating the impact of a professional English for academic purposes programme: A corpus based study [Doctoral dissertation, University of Birmingham]

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. Journal of Second Language Writing, 12(4), 377-403.Jiang, J., Bi, P., & Liu, H. 2019. Syntactic complexity development in the writings of EFL learners: Insights from a dependency syntactically-annotated corpus. *Journal of Second Language Writing, 46*, 100666.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065). doi:10.1098/rsta.2015.0202

Juffs, A. (2020). Aspects of Language Development in an Intensive English Program (1st Ed.). Routledge. https://doi.org/10.4324/9781315170190

Juffs, A., Han, N-R., & Naismith, B. (2020). The University of Pittsburgh English Language Corpus (PELIC) [Data set].

http://doi.org/10.5281/zenodo.3991977. Retrieved from: https://eli-data-mining-group.github.io/Pitt-ELI-Corpus. Accessed 24 June 2021.

Kecskes, I. (2018). Formulaic language and its place in intercultural pragmatics. In Siyanova-Chanturia & Pellicer-Sánchez (Eds.) *Understanding formulaic language: A second language acquisition perspective*, 132-149. London: Routledge.

Kim, J. E., & Nam, H. (2019). How do textual features of L2 argumentative essays differ across proficiency levels? A multi-dimensional cross-sectional study. *Reading and Writing*, *32*(9), 2251-2279.

Kim, S. (forthcoming) Multidimensional analysis of linguistic development in L2 learner writing in an Intensive English programme in a U.S university. *The Corpora and Language in Use.* Presses Universitaires de Louvain.

Kim, Y.-J. (1990). Register variation in Korean: A corpus-based study [Ph.D dissertation, University of Southern California].

Kobayashi, Y., & Abe, M. (2016). A Corpus-Based Approach to the Register Awareness of Asian Learners of English. *Journal of Pan-Pacific Association of Applied Linguistics*, *20*(2), 1-17.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57(1), 1–44.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t-test. Journal of Experimental Psychology: General, 142(2), 573–603.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine-grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.

Kyle, K. (2021). Natural language processing for learner corpus research. International Journal of Learner Corpus Research, 7(1), 1-16. https://doi.org/https://doi.org/10.1075/ijlcr.00019.int

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. Language Testing, 34(4), 513-535.

Kyle, K., & Crossley, S. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102, 333–349

Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, *43*(4), 781-812. https://doi.org/10.1017/s0272263120000546

Kyle, K., Eguchi, M., & Granger, S. (2021). Automatically assessing lexical sophistication using word, bigram, and dependency indices. Perspectives on the L2 Phrasicon: The view from learner corpora, 126-151.

Lan, G., Lucas, K., & Sun, Y. (2019). Does L2 writing proficiency influence noun phrase complexity? A case analysis of argumentative essays written by Chinese students in a first-year composition course. System, 85, 102116.

Lan, G., Zhang, Q., Lucas, K., Sun, Y., & Gao, J. (2022). A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. English for Specific Purposes, 67, 4-17.

Larsson, T., & Kaatari, H. (2019). Extraposition in learner and expert writing. International Journal of Learner Corpus Research, 5(1), 33-62. https://doi.org/10.1075/ijlcr.17014.lar

Larsson, T., & Kaatari, H. (2020). Syntactic complexity across registers: Investigating (in)formality in second-language writing. Journal of English for Academic Purposes, 45, 100850. https://doi.org/10.1016/j.jeap.2020.100850

Larsson, T, Paquot, M, & Biber, D. (2021). On the importance of register in learner writing: A multi-dimensional approach. In E. Seoane & D. Biber (Eds.), (pp. 235-258). Corpus based approaches to register variation. Amsterdam: Benjamins. https://doi.org/10.1075/scl.103.09lar

Lee, D. Y. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology*, *5*(3), 37-72.

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

Lightbown, P. M. (2013). *How Languages are Learned 4th edition*. Oxford University Press.

Llosa, L., Grapin, S. E., Friginal, E., Cushing, S. T., & Malone, M. E. (2020). Linguistic Dimensions of TOEFL iBT Essays Compared With Successful Student Disciplinary Writing in the University. *TESOL Quarterly*, *54*(1), 251-265. https://doi.org/10.1002/tesq.553

Loban, W. (1976). Language Development: Kindergarten through Grade Twelve. NCTE Committee on Research Report No. 18.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4), 474–496.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62.

Ly, A., Verhagen, J., & Wagenmakers, E.–J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32.

MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction?. Reading and Writing, 32(6), 1553-1574.

Malvern, D. D. P., Richards, B. J. P., Chipere, N. D., & Duran, P. (2004). *Lexical Diversity and Language Development : Quantification and Assessment*. Palgrave Macmillan Limited. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=358640

Man, D., & Chau, M. H. (2019). Learning to evaluate through that-clauses: Evidence from a longitudinal learner corpus. *Journal of English for Academic Purposes*, *37*, 22-33. https://doi.org/https://doi.org/10.1016/j.jeap.2018.11.007

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press.

McEnery, T., & Brezina, V. (2022). *Fundamental Principles of Corpus Linguistics*. Cambridge University Press. https://doi.org/DOI:

McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511981395

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, *45*, 499-515. Https://doi.org/10.3758/s13428-012-0258-1

Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On Determining Developmental Stages in Natural Second Language Acquisition. Studies in Second Language Acquisition, 3(2), 109-135. http://www.jstor.org.uoelibrary.idm.oclc.org/stable/44487208

Miles, J., Field, Z., & Field, A. (2012). *Discovering Statistics Using R*. SAGE Publications.

Mitchell, R. (2013). *Second Language Learning Theories*. Routledge.

Morey, R. D., Romeijn, J.–W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantifi- cation of statistical evidence. *Journal of Mathemati- cal Psychology*, *72*, 6–18.

Myhill, D. (2008). Towards a Linguistic Model of Sentenc Development in Writing. *Language and Education*, *22*(5), 271-288.

Myhill, D. (2009). from talking to writing: Linguistic Development in Writing. *BJEP Monograph Series 2, Number 6 - Teaching and learning Writing*, *27*(44), 27-44.

Myles, J. (2002). Second language writing and research: The writing process and error analysis in student texts. Tesl-ej, 6(2), 1-20.

Naismith, B., Han, N. R., & Juffs, A. (2022). The university of Pittsburgh English language institute corpus (PELIC). *International Journal of Learner Corpus Research*, 8(1), 121-138.

Nam, D. (2017). Functional Distribution of Lexical Bundle in Native and Non-Native Students' Argumentative Writing. *Journal of Asia TEFL*, *14*(4), 703.

Nelson, M. (2010). Building a written corpus. In *The Routledge handbook of corpus linguistics* (pp. 53-65). A. O'Keeffe & M. McCarthy.

Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. *British studies in applied linguistics*, *21*, 99.

Nesi, H., & Gardner, S. (2012). Genres across the disciplines: Student writing in higher education. Cambridge University Press.

Nesi, H., & Gardner, S. (2017). Stance in the BAWE Corpus: New Revelations from Multidimensional Analysis. In Corpus Linguistics 2017 Conference, University of Birmingham (pp. 25-28).

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175-220.

Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In Berber Sardinha, T. & Veirano Pinto M. (eds), Multi-Dimensional Analysis: Research Methods and Current Issues, 67-94, London; New York: Bloomsbury Academic. Retrieved from: https://sites.google.com/site/multidimensionaltagger/. Accessed 21 July 2021.

Norouzian, R., Goh, C., & Zhang, L. J. (2019). Changes over time in second language learners' use of English prepositions. Language Learning, 69(1), 249–287. DOI: 10.1111/lang.12333

Norris, J. M., & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. Applied Linguistics, 30(4), 555-578. https://doi.org/10.1093/applin/amp044

O'Donnell, M. B., & Römer, U. (2012). From student hard drive to web corpus (part 2): The annotation and online distribution of the Michigan Corpus of Upper-level Student Papers (MICUSP). Corpora, 7(1), 1-18.

Ochs, E. (1979). " Planned and Unplanned Discourse". In Syntax and Semantics. Leiden, The Netherlands: Brill. doi: https://doi.org/10.1163/9789004368897_004

Omidian, T., Siyanova-Chanturia, A., & Biber, D. (2021). A new multi-dimensional model of writing for research publication: An analysis of disciplinarity, intra-textual variation, and L1 versus LX expert writing. *Journal of English for Academic Purposes*, *53*, 101020. https://doi.org/https://doi.org/10.1016/j.jeap.2021.101020

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. https://doi.org/10.1093/applin/24.4.492

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. Journal of Second Language Writing, 29, 82-94. https://doi.org/10.1016/j.jslw.2015.06.008

Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs. Applied Linguistics, 30(4), 590-601. https://doi.org/10.1093/applin/amp045

Pallotti, G. (2015). A simple view of linguistic complexity. Second Language Research, 31, 117e134.

Pan, F. (2018). A multi-dimensional analysis of L1–L2 differences across three advanced levels. Southern African Linguistics and Applied Language Studies, 36(2), 117-131.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. Journal of English for Academic Purposes, 14, 48-59. https://doi.org/10.1016/j.jeap.2013.12.001

Pérez-Paredes, P. & Díez-Bedmar, B. (2018) Researching learner language through POS Keyword and syntactic complexity analyses. In S. Götz and J. Mukherjee (EDS.) Learner Corpora and Language Teaching. Studies in Corpus Linguistics Series. Amsterdam: John Benjamins, 101-127.

Polat, N., Mahalingappa, L., & Mancilla, R. L. (2020). Longitudinal Growth Trajectories of Written Syntactic Complexity: The Case of Turkish Learners in an Intensive English Program. Applied Linguistics, 41(5), 688-711. https://doi.org/10.1093/applin/amz034

Popper, K. (2002). The Logic of Scientific Discovery (2nd ed.). Routledge. https://doi-org.uoelibrary.idm.oclc.org/10.4324/9780203994627

Preacher, K., & Hancock, G. (2015). Meaningful Aspects of Change as Novel Random Coefficients: A General Method for Reparameterizing Longitudinal Models. Psychological Methods, 20, 84-101. https://doi.org/10.1037/met0000028

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for StatisticalComputing, Vienna, Austria. URL https://www.R-project.org/.

Rankin, T. (2015). Learner corpora and grammar. In G. G. S. Granger, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research (Cambridge Handbooks in Language and Linguistics)* (pp. 231-254). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.011

Rastgou, A., Storch, N., & Knoch, U. (2020). The Effect of Sustained Teacher Feedback on CAF, Content and Organization in EFL Writing. Iranian Journal of Language Teaching Research, 8(2), 41-61.

Reppen, R. (2007). First language & second language writing development of elementary students. In Y. Kawaguchi, T. Takagaki, N. Tomimori & Y Tsuruga (Eds.), Corpus-based Perspectives in Linguistics (pp. 147-167). John Benjamins Publishing Company.

Revelle, W. (2022). Psych: Procedures for personality and psychological research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.2.9.

Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition*, *41*(5), 1089-1110. https://doi.org/10.1017/s0272263119000202

Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6(2), 159-177.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rudis, B. (2020). hrbrthemes: Additional Themes, Theme Components and Utilities for ‘ggplot2’. R package version 0.8.0. https://CRAN.R-project.org/package=hrbrthemes

Santorini, B. (1990). Part-of-speech tagging guidelines for the Penn Treebank project (3rd revision, 2nd printing).

Sardinha, T. B., Pinto, M. V., Mayer, C., Zuppardi, M. C., & Kauffmann, C. H. (2019). Adding Registers to a Previous Multi-Dimensional Analysis. In T.

B. Sardinha & M. V. Pinto (Eds.), Multi-Dimensional Analysis: Research Methods and Current Issues (1 ed., pp. 165-186). Bloomsbury Academic.

Sarte, K. M., & Gnevsheva, K. (2022). Noun phrasal complexity in ESL written essays under a constructed-response task: Examining proficiency and topic effects. *Assessing Writing*, 51, https://doi.org/https://doi.org/10.1016/j.asw.2021.100595

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam and Philadelphia: Benjamins.

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. Language learning, 64(1), 103-131.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). *Understanding Formulaic Language: A Second Language Acquisition Perspective* (1st ed.). NY: Routledge. https://doi.org/https://doi-org.uoelibrary.idm.oclc.org/10.4324/9781315206615

Staples, S., Biber, D., & Reppen, R. (2018). Using corpus-based register analysis to explore the validity of high stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. Modern Language Journal, 102(2), 310-332.

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.

Staples, S., Gray, B., Biber, D., & Egbert, J. (2022). Writing Trajectories of Grammatical Complexity at the University: Comparing L1 and L2 English Writers in BAWE. Applied Linguistics, 44(1), 46-71. https://doi.org/10.1093/applin/amac047

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209-243.

Student. (1908). The probable error of a mean. Biometrika, 1-25.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics*, 5th ed. Boston MA: Allyn & Bacon/Pearson Education.

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What Linguistic Features Are Indicative of Writing Quality? A Case of Argumentative Essays in a College Composition Program. TESOL Quarterly, 47(2), 420-430. https://doi.org/10.1002/tesq.91

The British Association for Applied Linguistics (BAAL). (2021). Recommendations on Good Practice in Applied Linguistics. 4th Edition. Available at www.baal.org.uk. Accessed 21st July 2021

The Michigan English Placement Test (Michigan EPT), https://michiganassessment.org/pittsbur-tests/m-ept/.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=680385

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Harvard University Press.

Toutanova, K., Klein, D., Manning, C.D., & Singer, Y. (2003). *Feature-rich part-of-speech tagging with a cyclic dependency network*. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology – Volume 1 (pp. 173–180). Stroudsburg: Association for Computational Linguistics.

Trenkic, D., & Warmington, M. (2019). Language and literacy skills of home and international university students: How different are they, and does it matter? Bilingualism: Language and Cognition, 22(2), 349-365. https://doi.org/10.1017/S136672891700075X

Tyler, A. (2010). Usage-Based Approaches to Language and Their Applications to Second Language Learning. Annual Review of Applied Linguistics, 30, 270-291. https://doi.org/10.1017/S0267190510000140

Van Rooy, B., & Terblanche, L. (2006). A corpus-based analysis of involved aspects of student writing. Language Matters: Studies in the Languages of Africa, 37(2), 160-182.

Van Rooy, B., & Terblanche, L. (2009). A multi-dimensional analysis of a learner corpus. In A. Renouf & A. Kehoe (Eds.), Corpus Linguistics (pp. 239-254). Brill Rodopi.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing, 27, 1413-1432.

Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. The Modern Language Journal, 97(S1), 11-30.

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.

Weigle, S. C., & Friginal, E. (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes*, 18, 25-39.

Wells, G., & Chang, G. (1986). From speech to writing: Some evidence on the relationship between oracy and literacy. In Wilkinson, A (Ed). *The Writing of Writing*. UP: Milton Keynes. pp.109-131.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Winter, B. (2019). Statistics for Linguists: An Introduction Using R (1st ed.). Routledge. https://doi.org/10.4324/9781315165547

Winter, B., & Bürkner, P. C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. Language and Linguistics Compass, 15(11), e12439.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). Second language develop_ment in writing: Measures of fluency, accuracy, and complexity (Vol. 17). Honolulu: University of Hawaii Press.

Wray, A. (2018). Concluding question: Why don't second language learners more proactively target formulaic sequences? In *Understanding formulaic language* (pp. 248-269). Routledge.

Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. *Understanding formulaic language: A second language acquisition perspective*, 19-37.

Wulff, S. & Gries, S. Th. (2011). Corpus-driven methods for assessing accuracy in learner production. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 61-87). John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/exeter/detail.action?docID=769703

Yan, X., & Staples, S. (2019). Fitting MD analysis in an argument-based validity framework for writing assessment: Explanation and generalization inferences for the ECPE. Language Testing, 37 (2), 189 – 214.

Yates, F. (1934). Contingency Tables Involving Small Numbers and the $\chi 2$ Test. Supplement to the Journal of the Royal Statistical Society, 1(2), 217–235. http://doi.org/10.2307/2983604

Yoon, H.-J., & Polio, C. (2017). The Linguistic Development of Students of English as a Second Language in Two Written Genres. TESOL Quarterly, 51(2), 275-301. https://doi.org/https://doi.org/10.1002/tesq.296

Zhu, W. 2004. Faculty views on the importance of writing, the nature of academic writing, and teaching and responding to writing in the disciplines. Journal of Second Language Writing, 13, 29–48.

Zipf, G.K. 1949. Human behavior and the principle of least effort. Cambridge, Mass: Addison-Wesley

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with r*. Springer New York.

Appendices

Appendix 3-1: The 74 lexico-grammatical features used in the current study

source 1: 63 grammatical categories: adopted from the MAT manual, Nini (2019) pp.17-31 (Retrieved from: https://sites.google.com/site/multidimensionaltagger/, pp.18-)

source 2: The most common lexical verbs in seven semantic domains (Biber et al., 2021a, p.366-368)

source 3: The 23 grammatical complexity features, by structural form and syntactic function (Biber et al., 2021b, pp.14-15)

	Variable (stage[34])	Example
finite adverbial clause		
1	Causative adverbial subordinators (2)	I like it ***because it is blue***.
2	Concessive adverbial subordinators (2)	I like it ***although it is blue***.

[34] Hypothesised developmental stages for complexity features, adapted from Biber et al. (2021a: 278-279)

3	Conditional adverbial subordinators (2)	I may like it ***if it is blue***.
Finite complement clause		
4	That verb complements (1 & 2)	I *think* ***that you should come***.
5	That adjective complements (3)	I am *sure* ***that it is correct***.
6	That noun complement (5)	There is *a possibility* ***that it is wrong***.
7	WH-clauses (1)	I believed ***what*** *he said*.
8	Subordinator that deletion	I think ***you should come***.
Finite noun modifier clause		
9	That relative clauses on subject position (3)	I hate *the dog* ***that bit me.***
10	That relative clauses on object position (3)	I hate *the dog* ***that I saw***.
11	WH relative clauses on subject position	I saw *the man* ***who likes me.***
12	WH relative clauses on object position	I saw *the man* ***who Sally likes.***
13	Pied-piping relative clauses	He did not like *the manner* ***in which he was told****.*
14	Sentence relatives	*Bob likes fried mangoes,* ***which is disgusting.***
Non-finite complement clause		
15	Infinitives (2, 3 & 4)	I want ***to see the man***.
16	Split infinitives	He wants ***to convincingly prove it.***
Non-finite noun modifying clause		
17	Past participial WHIZ deletion relatives (4)	It is *the solution* ***produced by this process.***
18	Present participial WHIZ deletion relatives (4)	*The event* ***causing this decline*** *is today's topic.*
Adverbial phrase (a phrase modifying a clause)		
19	Place adverbials (2)	An example is found ***below****.*
20	Time adverbials (2)	I saw him ***yesterday***.
21	Prepositional clauses with 'OF' (3)	She was accused of cheating.
22	Prepositional clauses with other than 'OF' (3)	It accounts for the success.
Phrasal modifier (a phrase modifying noun, adjective or adverb phrase		

23	Attributive adjectives (2)	I saw *the* ***big*** *horse.*
24	Total adverbs (3, 4 & 5)	***Fortunately***, it was fixed in time.
25	prepositional phrase (3, 4, & 5)	I read a book ***of*** *writing development*.
Other		
26	Other adverbial subordinators	I have never seen him ***since*** *2020.*
27	Past tense	I ***saw*** him yesterday.
28	Perfect aspect	I ***have known*** him since 2020.
29	Present tense	I ***want*** to read the book.
30	First person pronouns	***I*** want to read it.
31	Second person pronouns	Do ***you*** want to read it?
32	Third person pronouns	Does ***he*** want to read it?
33	Pronoun it	***It*** is worth reading.
34	Demonstrative pronouns	***These*** are worth reading.
35	Indefinite pronouns	Does ***anybody*** want to read it?
36	Pro-verb do	You ***do*** the math.
37	Direct WH-questions	***What*** is the book about?
38	Nominalisations	This book is about ***happiness.***
39	Gerunds	***Running*** is good.
40	Total other nouns	I like ***books***.
41	Agentless passives	This book ***is sold out***.
42	By-passives	This book ***is written by*** *my friend*.
43	Be as main verb	I ***am*** very happy.
44	Existential there	***There*** *are* many books in this store.
45	Present participial clauses	***Stuffing his mouth with cookies****,* Joe ran out the door.
46	Past participial clauses	***Built in a single week****,* the house would stand for fifty years.
47	Predicative adjectives	The horse *is* ***big.***
48	Token	Total number of words in a text
49	Word length	Mean length of a word
50	Conjuncts	***Howeve***r, I do not agree with you.
51	Downtoners	It is ***nearly*** finished.
52	Hedges	It is *a* ***kind of*** exaggeration.

53	Amplifiers	It is ***absolutely*** right.
54	Emphatics	You are ***so*** kind.
55	Discourse particles	***Well***, I should get going.
56	Demonstratives	***These*** *books* are good.
57	Possibility modals	I ***can*** do it.
58	Necessity modals	I ***must*** do it.
59	Predictive modals	I ***will*** do it.
60	Contractions	It *is**n't*** right.
61	Stranded preposition	He is *the candidate that I was thinking* ***of.***
62	Split auxiliaries	*They* **are objectively shown** *that* it is good.
63	Phrasal coordination	I like *both apples* **and** *oranges*.
64	Independent clause coordination	I like apples*,* ***and I*** like oranges, too.
65	Synthetic negation	There is ***no*** *apple* in my house.
66	Analytic negation	This is ***not*** *an apple*.
67	ActivityV	I want to ***show*** you my picture.
68	Communication V	I want to ***talk*** to you about this issue.
69	mental V	I ***believe*** what you said.
70	causative V	The accident ***affected*** me and my family.
71	occurrence V	This accident ***happened*** to me.
72	existence V	This graph ***represents*** the number of oranges.
73	aspectual V	You should ***keep*** it tidy.

[1)] tag codes tagged by Stanford parser via TAASSC (ver. 1.3.2; Kyle, 2016)

[2)] variable codes adopted from MAT manual (ver 1.3; Nini, 2019)

Appendix 3-2: The seven semantic verb classes adopted from Biber et al. (2021)

67	Activity verb	make, get, go, give, take, come, use, leave, show, try, buy, work, move, follow, put, pay, bring, meet, play, run, hold, turn, send, sit, wait, walk, carry, lose, eat, watch, reach, add, produce, provide, pick, wear, open, win, catch, pass, shake, smile, stare, sell, spend, apply, form, obtain, reduce, arrange, beat,

		check, cover, divide, earn, extend, fix, hang, join, lie, pull, receive, repeat, save, share, throw, visit, accompany, acquire, advance, arrest, behave, borrow, burn, clean, climb, combine, control, defend, deliver, dig, elect, encounter, engage, exercise, expand, explore
68	Communication verb	say, tell, call, ask, write, talk, speak, thank, describe, claim, offer, suggest, admit, announce, answer, argue, deny, discuss, encourage, explain, express, insist, mention, note, propose, publish, quote, reply, report, shout, sign, sing, state, teach, warn, accuse, acknowledge, address, advise, appeal, assure, challenge, complain, consult, convince, declare, demand, emphasise, excuse, inform, invite, name, persuade, phone, pray, promise, question, recommend, remark, respond, specify, swear, threaten, urge, welcome, whisper
69	mental verb	see, know, think, find, want, mean, need, feel, like, hear, remember, believe, read, consider, suppose, listen, love, wonder, understand, expect, hope, assume, determine, agree, bear*, care, choose, compare, decide, discover, doubt, enjoy, examine, face, forget, hate, identify, imagine, intend, learn, mind, miss, notice, plan, prefer, prove, realise, recall, recognise, regard, suffer, wish, worry, accept, afford, appreciate, approve, assess, blame, bother, calculate, conclude, celebrate, confirm, count, dare, deserve*, detect, dismiss, distinguish, estimate, experience, fear, forgive, guess, ignore, impress, interpret, judge, justify, observe, perceive, predict, pretend, reckon, remind, satisfy, solve, study, suspect, trust
70	causative verb	help, let, allow, require, affect, cause, enable, ensure, force, prevent, assist, guarantee
71	occurrence verb	become, happen, change, die, grow, develop, occur, arise, bear (be born)*, emerge, fall, increase, last, rise, disappear, flow, shine, sink, slip
72	existence verb	seem, look, stand, stay, live, appear, include, involve, contain, exist, indicate, represent, deserve*, fit, locate, matter, reflect, relate, remain, reveal, sound, tend, concern, constitute, define, derive, illustrate, imply, lack, owe, own, possess
73	aspectual verb	start, keep, stop, begin, continue, complete, end, finish, cease

Appendix 3-3: Form-Function taxonomy adopted in a study exploring syntactic complexity (adapted from Biber et al., 2011, p.19)

Grammatical type	Syntactic function	index	Example
Finite dependent clause	Adverbial	Total finite adverbial clauses Because clause If clause Although clause	She won't narc on me, because she prides herself on being a gangster.
	Complement	verb + that clause verb + WH clause adjective + that clause noun + that clause	I don't know how they do it.
	Noun modifier	that relative clauses WH relative clauses	That's one thing that bothers me right now about my job.
Nonfinite dependent clause	Adverbial	to adverbial clause	To verify our conclusion that the organic material is arranged as a coating around the silica shell components, thin sections of fixed cells were also examined.
	Complement	verb + -ing clause verb + to clause adjective + -ing clause adjective + to clause noun + of + -ing clause noun + to clause	The main effect of grades has consistently been found to be the best predictor of future achievement.
	Noun modifier	nonfinite relative clause	The results shown in Tables IV and V add to the picture…
Dependent phrase (nonclausal)	Adverbial	adverbs as adverbials prepositional phrases as adverbials	Alright, we'll talk to you in the morning.
	Noun modifier	attributive adjectives nouns as nominal premodifiers	Class mean scores were computed by averaging the scores for male and female

		total prepositional phrases as nominal modifiers of as postmodifier in as postmodifier on as postmodifier with as postmodifier for as postmodifier	target students in the class.

Appendix 3-4: Accuracy Check for the TAASSC 71 features

Clausal features						Phrasal features		
	feature	Total		feature	Total		feature	Total
1	acomp	11	24	prep_as	2	1	advcl	2
	check	0		check	2		check	2
	precision rate	100		precision rate	100		precision rate	100
2	advcl	19	25	prep_at	7	2	advmod	11
	check	19		check	6		check	10
	precision rate	100		precision rate	85.71		precision rate	90.91
3	advmod	31	26	prep_because_of	1	3	amod	70
	check	30		check	1		check	68
	precision rate	96.77		precision rate	100		precision rate	97.14
4	cc	3	27	prep_before	1	4	ccomp	1
	check	0		check	1		check	1
	precision rate	100		precision rate	100		precision rate	100
5	ccomp	14	28	prep_by	2	5	conj	5
	check	12		check	2		check	5
	precision rate	85.71		precision rate	100		precision rate	100
6	conj_and	12	29	prep_for	6	6	conj_and	22
	check	12		check	6		check	20
	precision rate	100		precision rate	100		precision rate	90.909
7	conj_but	1		prep_from	6	7	conj_or	2

	check	1	30	check	6		check	2
	precision rate	100		precision rate	100		precision rate	100
8	conj_or	1	31	prep_in	17	8	dep	9
	check	1		check	17		check	0
	precision rate	100		precision rate	100		precision rate	0
9	cop	1	32	prep_in_front_of	1	9	det	98
	check	1		check	1		check	97
	precision rate	100		precision rate	100		precision rate	98.98
10	csubj	2	33	prep_into	1	10	dobj	108
	check	2		check	1		check	105
	precision rate	100		precision rate	100		precision rate	97.22
11	dep	4	34	prep_of	2	11	iobj	2
	check	4		check	1		check	2
	precision rate	100		precision rate	50		precision rate	100
12	dobj	101	35	prep_on	5	12	ncomp	20
	check	100		check	5		check	20
	precision rate	99.01		precision rate	100		precision rate	100
13	expl	7	36	prep_to	4	13	neg	2
	check	7		check	4		check	2
	precision rate	100		precision rate	100		precision rate	100
14	iobj	2	37	prep_under	1	14	nn	49
	check	2		check	1		check	49
	precision rate	100		precision rate	100		precision rate	100
15	mark	26	38	prep_with	9	15	nsubj	151
	check	26		check	9		check	145
	precision rate	100		precision rate	100		precision rate	96.026
16	ncomp	26	39	prepc_as_for	2	16	nsubjpass	9
	check	26		check	2		check	8

	precision rate	100		precision rate	100		precision rate	88.889
17	neg	14	40	prepc_for	5	17	num	10
	check	14		check	5		check	10
	precision rate	100		precision rate	100		precision rate	100
18	nsubj	150	41	prt	2	18	parataxis	2
	check	147		check	2		check	2
	precision rate	98		precision rate	100		precision rate	100
19	nsubjpass	9	42	tmod	6	19	pcomp	2
	check	9		check	6		check	2
	precision rate	100		precision rate	100		precision rate	100
20	parataxis	4	43	v	165	20	pobj	152
	check	3		check	162		check	142
	precision rate	75		precision rate	98.182		precision rate	93.421
21	pobj	2	44	vcop	42	21	poss	27
	check	2		check	42		check	27
	precision rate	100		precision rate	100		precision rate	100
22	prep_about	1	45	vmod	1	22	predet	1
	check	1		check	1		check	1
	precision rate	100		precision rate	100		precision rate	100
23	prep_after	1	46	xcomp	23	23	prep	58
	check	0		check	23		check	53
	precision rate	0		precision rate	100		precision rate	91.379
						24	rcmod	18
							check	14
							precision rate	77.778
						25	vmod	8
							check	8
							precision rate	100

Appendix 3-5: Anonymisation scheme: coding scheme of student information

(Retrieved from GitHub - ELI-Data-Mining-Group/PELIC-dataset: The University of Pittsburgh English Language Institute Corpus (PELIC) dataset

Column	Column name	Description
A	anon_id	a unique anonymous identifier for each student - two letters and one integer, e.g. *eq0*
B	gender	'Male','Female',or 'Unknown' based on student responses
C	birth year	four digit year
D	native language	students input their own first language (not from a drop-down menu)
E	language_used_at-home	language used at home in their home country, not in the U.S.
F, J, N	Non-native_language_1,2,3	the non-L1s (L2, L3, L4) with which the student feels they have the highest proficiency
G, K, O	yrs_of_study_lang1,2,3	the number of years the student has studied the non-L1s provided in columns F, J, N
H, I, P	study_in_classroom_lang1,2,3	whether or not the student studied their non-L1s in a classroom setting ('yes' or 'no')
I, M, Q	ways_of_study_lang1,2,3	students selected from a menu how they studied their non-L1s, e.g. *Practiced reading aloud*
R	course_history	a list of all the courses attended (course_id codes)
S	yrs_of_english_learning	the number of years the student has been learning English, selected from a drop-down list
T	yrs_in_english_environment	the number of years the student has lived in an English-speaking environment, selected from a drop-down list
U	age	the student's age at the time of enrollment

Appendix 4-1: Review of MDA studies[35][36]

No	Author	year	Corpus	Factor analysis	Dimensions	Linguistic features 1) 2)
1	Biber	1988	17 written (including academic prose) & 6 spoken genres mainly drawing on the LOB & LL corpus	Yes	D1: involved vs informational production D2: narrative vs non-narrative concerns D3: explicit vs situation-dependent reference D4: overt expression of persuasion D5: Abstract vs Non-Abstract Information D6: online informational elaboration	67
2	Biber	1995	spoken & written texts of English, Nuk ulaelae Tuvaluan, Korean, and Somali	no	Six dimensions established in Biber (1988)	67
3	Biber	2006	T2K-SWAL Corpus	Yes	D1: Oral vs Literate discours D2: Procedural vs Content-focused disocurse D3: Reconstructed account of events D4: Teacher-centered stance	129 ->90
4	Biber	1992	23 spoken & written genres	yes	D1: Reduced structure and	33

[35] *Asterisk indicates studies that did not state the number of linguistic features used in the study but assumed by the fact that they used Biber's 1988 with MAT, which contained 67 grammatical features

[36] Linguistic features for initially included and finally retained numbers (when only one number is available, they are final number or didn't conduct factor analysis so first number is retained)

					specificity D2: Structural elaboration of reference D3: "Framing" structural elaboration D4: Integrated structure D5: Passive constructions	
5	Aguado-Jiménez, Pérez-Paredes & Sánchez	2012	59 interview (spoken) texts with the OPI format of LINDSEI corpus / the British speakers' corpus	no	Five dimensions established in Biber (1988)	67*
6	Kim & Nam	2019	ICNALE corpora / argumentative essays/ two topics	No	Five dimensions established in Biber (1988)	67*
7	Kobayashi & Abe	2016	ICNALE corpora (International Corpus Network of Asian Learners of English)	no	correspondence analysis instead of factor analysis - Gower's distance	67 ->58
8	Moon	2016	Korean vs English native journal articles	No	Five dimensions established in Biber (1988)	67
9	Crosthwaite	2016	HKU-CAES learner corpus (Longitudinal, between three-time points, with timed test conditions for first & third points)	No	Five dimensions established in Biber (1988)	67*

10	Nesi & Gardner	2012	BAWE	no	Five dimensions established in Biber (1988)	67
11	Gardner, Biber, & Nesi	2015	BAWE	Yes	four dimensions by level of study four dimensions by disciplinary groups	N/A
12	Nesi & Gardner	2017	BAWE	no	based on established dimesnions by Gardner et al., 2019, focused on stance distinciton: D1: stance adverbials, stance nouns controlling that-clauses clustered with 3rd person pronouns, proper nouns and communication verbs associated with 'stance toward the work of others) D2: stance verbs controlling to- and that-cluses indiciting 'personal stance'	39
13	Gardner, Nesi, & Biber	2019	BAWE	yes	D1: compressed procedural information vs stance towards the work of others D2: personal stance D3: possible events vs completed events D4: informational density	150 -> 39
14	Hardy & Römer	2013	MICUSP	Yes	D1: involved vs informational D2: personal stance D3: (-)	110 -> 54

					procedural D4: possibility	
15	Hardy & Friginal	2016	MICUSP	yes	MICUSP dimensions (Hardy & Romer, 2013)	137*
16	Friginal & Mustafa	2017	RA abstracts in four disciplines from the U.S. and Iraq	No	1,3,4 Dimensions of MICUSP (Hardy & Romer, 2013)	137
17	Biber & Gray	2013	TOEFL iBT speaking & writing test answers (L2 corpora)	Yes	D1: literate vs oral responses D2: information source: text vs personal experience D3: abstract opinion vs concrete description/summary D4: personal narration	123 -> 28
18	Pan	2018	Master's theses, PhD dissertations, and research articles by L1 & L2 speakers	yes	D1: Attitudinal vs descriptive D2: Immediate style vs reported style D3: Academic involvement vs information density D4: Clausal style vs phrasal style	150 ->33
19	Goulart	2021	BAWE & BrAWE (Brazilian uni students writing)		D1: Expression of personal opinion vs compressed procedural information D2: Expression of possibility vs account of completed actions D3: Informational density vs engaging presentation D4: Involved academic narrative vs	34

					elaborate description	
20	Issitt	2017	EAPCORP corpus consisting of 526 individual scripts and 263 matched pairs (pre and post course).		Six dimensions established in Biber (1988)	67
21	Biber	1986	500 written texts of 15 genres from the LOB corpus & 87 spoken British English texts from LL corpus	yes	D1: Interactive vs Editied Text D2: Abstract vs Situated Content D3: Reported vs Immediated Style	41
22	Doolan	2020	232 L1 university essay writing in Hispanic Serving Institution	yes	D1: Source-based concept density vs Prompt-based freewriting D2: Impersonal extension of source-based concepts D3: Source text deixis	
23	Gray, Cotos & Smith	2020	ISURA (the Iowa State University Research Article) corpus-published RAs follwing an IMRD/C structure from 30 disciplines	PCA	D1: Interpretation & expansion vs simple reportage D2: Abstraction / overt empiricism D3: Procedural narration D4: Interpreting results vs informational density	140+ ->45
24	Eric Friginal & Sara Weigle	2014	207 L2 student essays produced at three time points	yes	D1: invovled focus vs informational focus D2: addressee-focused description vs personal narrative D3:	140+ ->72

					simplified description vs elaborated description D4: personal opinion vs reflective/argument ative discourse	

Appendix 4-2: Pilot study 3: 34 linguistic features (with loadings larger than ± 0.35) on each dimension.

	Linguistic variables	D1	D2	D3	Examples (from PELIC data)
Clausal dependent types					
1	Adverbial clause		0.59		[After I came to the U.S], I had to manage to find a house by myself.
2	Adverbial modifier	(-0.33)			You have to practice and make mistakes [otherwise] you will never gain fluency.
3	Auxiliary verb		0.52		You [have to] practice and make mistakes otherwise you will never gain fluency.
4	Clausal complement		0.39		The most important point is [that I want to achieve my goal].
5	Clausal coordination	(-0.3)			I didn't take any English classes when I was studying university and later, [that's why it was regressed].
6	Clausal prepositional complement	0.46			~ according to [my observation on my near surroundings],*
7	Conjunction		-0.65		In the ancient society, we were starving and [finding something to eat].
8	Direct object		0.39		And getting [good grades] is the master key of that door.
9	Discourse marker	0.65			[Yes], it happens right now !!
10	Indirect object	0.61			Dad, Please give [me] some suggestions ...
11	Nominal subject		0.56		[Learning a foreign language][1)] is so difficult and required a lot of work~

12	Open clausal complement		0.41		Eventually, I decided to [move out] and had to find my house.
13	Agent	(0.39)		**<u>0.44</u>**	TV and IPOD are some of the electronic devices that are produced by some [components]~
14	Passive auxiliary verb			0.94	This means that some dogs [are] specially bred on food and pets [are] never eaten.
15	Passive clausal subject	0.97			[Living with someone][1)] may get upset of culture difference and life differences
16	Passive nominal subject			0.93	[This advantage] could be used almost anywhere.
17	Prepositional modifier: across	0.97			The last part of the movie "Far and away" flashed [across] my mind.
18	Prepositional modifier: after	0.49			That will decrease the time to do the HW and review the lessons [after] classes and let that are easy.
19	Prepositional modifier: against	0.85			The effect of insufficient representation is that people often protest [against] the government and demand a new election.
20	Prepositional modifier: among	0.84			This is very helpful factor of preventing this crime to spread [among] the young people as well as students.
21	Prepositional modifier: around	0.62			In Aid Al-fitir, all the Muslims [around] the world celebrate on the same day.
22	Prepositional modifier: between	(0.3)			In addition, you have to choose school that does not discriminate [between] students in dealing with them because that might affect your learning and thinking.
23	Prepositional modifier: in		(0.33)		Knowing a second language can give a person a chance to be [in] the leading.
24	Prepositional modifier: into	0.65			There are a lot of mistakes that new language learners make while they are learning a new language, these mistakes can be divided [into] three kinds: Fear, hard work and translation.
25	Prepositional modifier: like	0.47			But there are a lot of differences [like] here the family don't

					continue his life together [like] my home.
26	Prepositional modifier: of		-0.46		Due to this, a sergeant who is a soldier of the highest position[2)] should lead a team which consists [of] privates who are soldiers of lowest position[2)] and succeed works which the team must do.
27	Prepositional modifier: over	0.57			As I rushed down the stairs to eat a quick breakfast, I tripped [over] my cat and had to hurry up.
28	Prepositional modifier: through	0.68			On my way to work, I decided to take a shortcut [through] an old part of town.
29	Prepositional modifier: toward	0.92			In fact, it is very easy things to do [toward] a relationship with student and it will not cost anything.
30	Prepositional modifier: under	0.91			Some of them are being [under] pressure because of several difficulties.
31	Subordinating conjunction		0.5		I could not believe [that I did not have my house from that day].
Phrasal dependent types					
32	Determiners		-0.38		When you have [the] master key, you have many chances to open [any] door of your dream life.
33	Verbal modifiers	-0.35			When you organise your time you can find more time to [study] and [do] your task and assignment.
34	Conjunction: and		-0.68		~ that will increase your chance to choose your interesting college [and] dream life.

Appendix 4-3: A full report of standardised loadings (pattern matrix) based upon correlation matrix

linguistic variables	*PA1*	*PA2*	*PA4*	*PA3*	*h2*	*u2*	*com*
Word length	-0.3	0.58	-0.12	0.2	0.63	0.37	1.9
Be as main verb	-0.21	-0.09	0.82	-0.03	0.59	0.41	1.2
Causative adverbial subordinators	0.01	-0.16	0.19	0.53	0.38	0.62	1.5

Conditional adverbial subordinators	0.65	0.2	-0.03	-0.04	0.34	0.66	1.2
Contractions	0.39	-0.09	0.19	-0.04	0.28	0.72	1.6
First person pronoun	-0.03	-0.71	0.04	0.09	0.52	0.48	1
Attributive adjectives	-0.21	(0.31)	0	0.11	0.19	0.81	2.1
Total other nouns	-0.57	0.17	0	-0.07	0.44	0.56	1.2
Nominalisations	-0.1	0.35	-0.15	0.11	0.22	0.78	1.8
Predictive adjectives	(-0.33)	0.25	0.18	0.04	0.2	0.8	2.5
Predictive adjectives	-0.07	-0.04	0.74	0.04	0.52	0.48	1
Second person pronouns	0.7	0.24	-0.12	-0.2	0.38	0.62	1.5
infinitives	0.41	-0.05	-0.18	0.11	0.19	0.81	1.6
Past tense	-0.35	-0.85	-0.15	0.08	0.64	0.36	1.4
Present tense	0.25	0.41	0.47	-0.01	0.4	0.6	2.5
Analytic negation	0.37	-0.04	0.08	0.08	0.19	0.81	1.2
mental verbs	0.47	0	-0.06	0.1	0.22	0.78	1.1
causative verbs	0.08	0.16	0.03	0.82	0.7	0.3	1.1
other prepositions modifying clauses	0.06	0.06	-0.12	0.51	0.28	0.72	1.2
OF preposition modifying clauses	-0.39	0.17	-0.11	-0.08	0.31	0.69	1.7
all prepositions modifying phrases	-0.46	0.01	-0.17	-0.14	0.35	0.65	1.5

	PA1	PA2	PA4	PA3
SS loadings:	2.7	2.18	1.7	1.4
Proportion variance	0.13	0.1	0.08	0.07
Cumulative variance	0.13	0.23	0.31	0.38
Proportion explained	0.34	0.27	0.21	0.18
Cumulative proportion:	0.34	0.61	0.82	1
	Factor correlations			
	1.00	-0.39	0.40	0.15
		1.00	-0.18	-0.09
			1.00	0.02

				1.00

Appendix 6-1. Pilot Study: All collexemes of the that-clause constructions in the 326 PELIC texts

	verb	frequency	association strength
1	think	105	101.23
2	mean	40	38.14
3	realize	23	22.25
4	learn	23	20.72
5	believe	19	18.27
6	know	21	18.26
7	feel	16	12.93
8	find	12	9.76
9	see	12	9.73
10	hope	7	6.70
11	understand	6	5.22
12	expect	5	4.00
13	notice	4	3.59
14	assume	3	2.92
15	consider	3	2.35
16	conclude	2	1.92
17	discover	2	1.89
18	Imagine	2	1.76
19	guess	1	0.81
20	accept	1	0.62

Appendix 6-2. Final study: Verbs most strongly attracted to the That-complement clauses (based on the collostruction scores)

rank	verb	collostruction strength	V frequency in construction (total frequency of C: 1673)	overall freq of verb)
1	think	385.16	393	860

2	say	151.62	159	614
3	believe	132.39	135	213
4	know	92.40	104	909
5	mean	81.35	86	360
6	feel	63.77	72	627
7	learn	48.04	59	823
8	see	45.88	50	309
9	realize	41.12	42	66
10	find	35.68	43	543
11	hope	26.63	28	101
12	show	24.88	28	229
13	argue	18.47	19	39
14	remember	16.92	18	79
15	agree	15.48	17	111
16	write	14.58	17	176
17	consider	13.81	18	305
18	recognize	13.31	14	50
19	notice	12.08	13	67
20	suggest	11.28	12	52
21	understand	9.27	12	198
22	ensure	8.74	9	19
23	claim	8.50	9	36
24	decide	8.29	11	196
25	conclude	7.75	8	18
26	indicate	7.63	8	27
27	recommend	7.17	8	60
28	forget	7.13	8	63
29	report	7.09	8	66
30	imagine	7.00	8	72
31	deny	6.89	7	8
32	seem	6.87	8	82
33	expect	6.45	8	112
34	mention	6.27	7	53
35	guess	5.72	6	20
36	acknowledge	5.67	6	24

37	hear	5.08	8	211
38	prove	4.80	9	303
39	imply	4.71	5	21
40	accept	4.54	6	105
41	assume	3.86	4	10
42	demonstrate	3.79	4	15
43	discover	3.47	4	38
44	desire	3.21	4	57
45	promise	2.85	3	11
46	announce	2.83	3	12
47	perceive	2.81	3	14
48	predict	2.81	3	14
49	reveal	2.69	3	22
50	arrange	2.53	3	34
51	judge	2.14	3	62
52	complain	1.87	2	9
53	submit	1.85	2	11
54	admit	1.83	2	12
55	confirm	1.82	2	13
56	propose	1.81	2	14
57	comment	1.78	2	16
58	establish	1.39	2	44
59	appear	1.11	2	64
60	note	1.03	2	70
61	contend	0.99	1	1
62	presume	0.99	1	1
63	stipulate	0.99	1	1
64	assert	0.97	1	2
65	reply	0.94	1	4
66	repeat	0.90	1	7
67	infer	0.89	1	8
68	insist	0.78	1	16
69	estimate	0.76	1	17
70	warn	0.69	1	22
71	guarantee	0.65	1	25

72	maintain	0.53	1	34
73	object	0.53	1	34
74	grant	0.49	1	37
75	allow	0.08	2	138
76	order	0.08	3	210
77	dream	-0.02	1	73
78	check	-0.11	1	80
79	require	-0.20	2	158
80	ask	-1.33	3	311
81	prefer	-1.49	1	179
82	state	-2.06	3	363
83	add	-7.70	1	621
84	have	-23.15	30	3741
85	be	-236.52	70	18170