The Role of Prompt Characteristics in Shaping L2 Spanish Writing: Lexical Diversity and Pronoun Use in a Learner Corpus

By

PALOMA FERNÁNDEZ-MIRA

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Spanish

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Claudia Sanchez-Gutierrez, Chair

Agustina Carando

Álvaro Recio Diego

Committee in Charge

2024

# Acknowledgements

First and foremost, I would like to thank my advisor, Professor Claudia Sanchez-Gutierrez, for her exceptional mentorship, constant support, and dedication to my success over the past six years. I feel incredibly fortunate to have had the opportunity to work alongside you, learning from your outstanding work ethic, patience, passion, research skills, caring and collaborative spirit, among many other qualities. You have made a tremendous impact in my life both academically and personally, and I am beyond grateful for your guidance and encouragement week after week. Thank you for believing in me, for motivating me, and dedicating countless hours to help me grow into the teacher, researcher, and individual that I am today.

I would also like to express my deepest gratitude to the other members of my dissertation committee, Professor Alvaro Recio Diego and Professor Agustina Carando, for their insightful feedback and encouragement. Their expertise and dedication were invaluable throughout this journey, and I am sincerely thankful for their thoughtful reviews of my work. Additionally, I am grateful to my other professors, who were always helpful and kind.

A special thank you goes to my fellow graduate students, who have become not only my colleagues but also my close friends. Sharing this experience with you in research, teaching, and life has been a privilege. I will always cherish these years and the warmth, collaboration, and deep friendships we formed.

Lastly, I also want to thank my family and my friends, my chosen family, for their constant support. Your willingness to listen to my ups and downs and your encouragement have kept me going. I could not have reached this milestone without you.

# Dedication

*A Carlos, por brindarme un apoyo y amor incondicional a lo largo de todos nuestros años juntos. Sin ti, mi viaje académico y esta tesis habrían sido impensables.*

*Y a nuestros hijos, Logan y Cassandra, por iluminarme los días y enseñarme tantísimo.*

# Abstract

Learner corpus research has expanded from focusing primarily on English as a second language (L2) to include languages such as L2 Spanish, reflecting the growing importance of corpus linguistics in second language acquisition (SLA) research. In this context, and because prompts are the means by which learner corpora gather their texts, it has become necessary to consider the impact of prompt characteristics on text features. This dissertation examines how prompts, genre, and narrative voice affect the lexical and syntactic features of L2 Spanish learner writing, using the COWS-L2H corpus (Davidson et al., 2020; Yamada et al., 2020).

The research explores the influence of different prompts and narrative voices on the Measure of Textual Lexical Diversity (MTLD) and the rate of subject pronoun presence (SPP) errors across various proficiency levels. Mixed-effects regression models reveal that first-person texts, particularly self-descriptions, tend to have higher MTLD scores and fewer SPP errors compared to third-person descriptions of special or famous individuals. A similar pattern emerges in narrative texts, where first-person narratives exhibit greater MTLD scores than third-person narratives, demonstrating the impact of personal connection and emotional resonance in learner writing.

Contrary to expectations, no significant differences in MTLD are found between descriptions of a close person and a famous person, suggesting that emotional closeness alone does not drive lexical diversity in learner texts. Additionally, while descriptive texts show significant effects of narrative voice and course level on SPP errors, narrative texts do not, emphasizing the key role of genre in determining linguistic accuracy.

This dissertation contributes to the understanding of how corpus design choices impact the study of L2 language and offers insights into the broader implications for SLA research, including areas such as language testing and pedagogy. By examining the relationships between prompts, genre, and narrative voice, it offers practical guidance for corpus developers and researchers, aiming to enhance the validity and reliability of learner corpora and inform future research directions in SLA.

# Table of contents

**Chapter 1: Introduction** ..... **1**
**Chapter 2: Background** ..... **4**
2.1 Introduction ..... 4
2.2 Spanish learner corpora ..... 6
2.3 Task- and topic-related variables and their effect in writing ..... 14
2.3.1 Task complexity ..... 14
2.3.2 Genre ..... 20
2.3.3 Topic-related variables: Valence and closeness ..... 22
2.3.4 Subjective interpretation and within-prompt variation ..... 24
2.4 Lexical diversity in L2 development ..... 26
2.5 Subject pronoun presence in L2 development ..... 28
2.5.1 Characterization of subject pronoun presence and absence in Spanish ..... 29
2.5.2 Factors influencing SPP in L2 Spanish ..... 32
2.6 Research questions and hypotheses ..... 36
2.6.1 Study 1 ..... 36
2.6.2 Study 2 ..... 37
**Chapter 3: Methodology** ..... **42**
3.1 Introduction ..... 42
3.2 Corpus used ..... 42
3.2.1 Participants and context ..... 43
3.2.2 Prompts ..... 46
3.2.3 Annotation efforts ..... 48
3.3 Subcorpora used in this study ..... 51
3.3.1 Subcorpus 1: Famous-Special-Yourself ..... 51
3.3.2 Subcorpus 2: Beautiful-Vacation ..... 53
3.4 Dependent variables ..... 54
3.4.1 Study 1: MTLD ..... 54
3.4.2 Study 2: Subject pronoun presence error rate ..... 57
3.5 Study 1's analytical procedure ..... 58
3.5.1 Research question 1 ..... 58
3.5.2 Research question 2 ..... 59
3.6 Study 2's analytical procedure ..... 60

3.6.1 Research question 3 ..... 60
3.6.2 Research question 4 ..... 61
3.6.3 Research question 5 ..... 62
**Chapter 4: Study 1 results ..... 64**
4.1 Introduction ..... 64
4.2 Research question 1: Closeness effects in descriptive texts ..... 65
4.3 Research question 2: Person effects in narrative texts ..... 69
**Chapter 5: Study 2 results ..... 72**
5.1 Introduction ..... 72
5.2 Research question 3: Closeness effects on SPP errors in descriptive texts ..... 72
5.3 Research question 4: Narrative voice effects on SPP errors in descriptive texts ..... 76
5.4 Ad hoc analyses: Effect of level and narrative voice on average sentence length in descriptive texts ..... 78
5.5 Research question 5: Person effects on SPP errors in narrative texts ..... 84
**Chapter 6: Discussion ..... 86**
6.1 Discussion of Study 1: MTLD ..... 86
6.1.1 RQ 1: Descriptive subcorpus (Yourself-Famous-Special) ..... 86
6.1.2 RQ 2: Narrative subcorpus (Beautiful-Vacation) ..... 88
6.2 Discussion of Study 2: SPP errors ..... 90
6.2.1 RQ 3 and 4: Descriptive subcorpus (Yourself-Famous-Special) ..... 90
6.2.2 RQ 5: Narrative subcorpus (Beautiful-Vacation) ..... 92
**Chapter 7: Conclusion ..... 94**
7.1 Implications for research methods ..... 94
7.1.1 Methodological considerations ..... 94
7.1.2 Corpus design ..... 95
7.2 Implications for language teaching and learning ..... 96
7.3 Implications for language testing ..... 96
7.4 Limitations and future research ..... 97
**References ..... 99**
**Appendix I ..... 110**

# Chapter 1: Introduction

The field of learner corpus (LC, henceforth) research has seen significant growth in recent years, expanding from its initial focus on English as a second language (L2) to encompass a wide range of languages, including L2 Spanish. This expansion reflects the increasing recognition of corpus linguistics as a powerful methodological approach in second language acquisition (SLA) research. Unlike traditional SLA research methods, which often rely on controlled experiments and elicited data, corpus-based studies allow for the analysis of authentic language use in semi-naturalistic contexts. However, this methodological shift necessitates careful consideration of how LC are created and used, as each corpus comes with its own set of characteristics and limitations.

The design and utilization of LC are critical aspects that researchers must address to ensure the validity and reliability of their findings. Differences in corpus design result in unique affordances, influencing the types of insights that can be obtained about the linguistic development of language learners. Since written LC rely on prompts that can significantly shape the lexical and syntactic characteristics of the resulting data, the nature of these prompts becomes a crucial design choice. For example, previous literature has studied the relationship between task complexity and syntactic accuracy of the resulting text, and found that the type of complexity implied in the prompt either increases, reduces, or has no effect on learner accuracy (Ellis & Yuan, 2004; Ishikawa, 2006; Kuiken & Vedder, 2008; Cho, 2019; Rahimi & Zhang, 2019; Zhan, Sun, & Zhang, 2021). Textual genre (argumentative, narrative, descriptive, etc.) has been shown to influence lexical measures such as lexical variation and sophistication in L2 learners (Sadeghi & Dilmaghani, 2013; Castañeda-Jiménez & Jarvis, 2014; Heng, Pu, & Liu, 2023; Sánchez-Gutiérrez & Fernández-Mira, 2023). Other prompt characteristics such as topic

also impact vocabulary measures: for instance, Fernandez-Mira et al. (2021) found that positive prompts elicited texts with more varied vocabulary than negative prompts. Therefore, understanding the implications of prompt choices is essential for both corpus developers and researchers who rely on these resources.

This dissertation aims to contribute to this understanding by exploring the impact of prompt characteristics such as closeness, textual genre, and narrative voice (first- versus third-person narrator) on the lexical and syntactic characteristics of L2 Spanish learner writing. Using the *Corpus of Written Spanish of L2 and Heritage Speakers* (COWS-L2H; Davidson et al., 2020; Yamada, et al., 2020), the research investigates how different types of prompts and narrative voices influence linguistic features at various proficiency levels. Through this exploration, the dissertation seeks to provide practical insights into effective corpus design and to highlight their implications for research methods, language testing, and future research in SLA.

The structure of the dissertation is as follows: Chapter 2 presents an overview of the theoretical framework and reviews relevant literature in LC research and SLA. Chapter 3 details the methodology used in the studies, including the design of the COWS-L2H corpus and the specific prompts and statistical models employed. Chapter 4 presents the findings of the first study, focusing on the effect of closeness to the described individual, narrative voice (first- versus third-person narrator), course level, and textual genre on lexical diversity (LD, henceforth) as measured by the Measure of Textual Lexical Diversity (MTLD). Chapter 5 discusses the results of the second study, which examines the rate of subject pronoun presence (SPP, henceforth) errors in texts written with either a first- or third-person narrator, from descriptive or narrative genres, by learners at varying course levels. Chapter 6 synthesizes the findings and discusses their implications in the context of existing literature, fitting the results

within the broader picture of SLA research. Finally, Chapter 7 addresses the implications of the research for corpus design, pedagogy, and language testing, outlines the limitations of the studies, and suggests potential future research directions.

In conclusion, this dissertation positions itself as a crucial step in understanding the repercussions of corpus design choices in SLA research. By shedding light on the complex interplay between prompts, genre, and narrative voice in learner writing, it aims to guide both corpus developers and researchers towards more effective and insightful use of LC.

# Chapter 2: Background

## 2.1 Introduction

This dissertation examines the impact of closeness to a described individual, narrative voice (first- versus third-person narrator), genre (description versus narration), and learner course level on L2 Spanish texts from a corpus linguistics perspective. Specifically, it investigates how these variables influence two main aspects:

1. Lexical diversity (MTLD score): The dissertation explores whether closeness to the described individual (e.g., self versus a special person versus a famous person) affects the variety of vocabulary used in the texts. Previous research (Fernandez-Mira et al., 2021) suggests that learners use a wider variety of words when describing a famous person compared to a special person. This dissertation explores whether emotional closeness can account for this difference or if other factors are at play. For instance, a text about a famous person might be written as a biography, whereas a text about a special person could focus more on the relationship between the writer and the described individual. If learners approach these prompts differently, they are effectively completing distinct tasks despite the similar appearance of the prompts. Additionally, this study examines how narrative voice affects MTLD scores, hypothesizing that first-person narratives could lead to more varied vocabulary than third-person narratives. The role of genre is also examined to determine if descriptive and narrative texts behave similarly in regards to variations in topic closeness and narrative perspective. Finally, the study considers how course level influences these patterns.
2. Subject pronoun presence (SPP) errors: The research explores whether the use of subject pronouns varies with narrative voice and closeness to the described individual. Given that

Spanish is a pro-drop language, learners may use subject pronouns more frequently in contexts where the subject is less personally connected, such as when describing a famous person versus themselves. For example, when describing oneself, learners might rely less on explicit subject pronouns due to their sense of personal connection with the content, leading to fewer overuse errors. Conversely, when describing a third person, the need for explicit subject pronouns might vary depending on the level of involvement or personal connection to the subject. By examining these aspects, the dissertation seeks to clarify how closeness and narrative voice influence SPP errors and whether the effect of these variables differs across textual genres.

By analyzing these lexical and grammatical aspects, this dissertation aims to provide insights into how topic-related variables affect L2 Spanish learner writing. These findings have implications for corpus design, assessment practices, and L2 Spanish instruction.

In this chapter, I will conduct a review of LC and their prompts, followed by an exploration of existing literature on the dependent and independent variables under consideration. By doing so, the chapter sets the stage for understanding the significance and necessity of the studies conducted within this dissertation.

In the field of SLA research, analyzing large collections of learner language has become crucial for studying the complexities of language learning and proficiency assessment. Section 2.2 looks at how the largest Spanish LC are designed and their methodological decisions regarding prompts, or the topics learners are asked to write about. These decisions have implications when assessing writing samples and trying to discern proficiency from topic effects. Manual and automatic annotations included in these corpora are also described and discussed. Among these corpora, COWS-L2H stands out as it uniquely allows for the separation of the topic

variable from any other variables, such as proficiency or course level, and it includes error annotations. Hence, for the purposes of this dissertation, COWS-L2H is chosen as the Spanish LC of focus and will be thoroughly described in this chapter.

Moving forward, Section 2.3 examines how different writing tasks, genres, prompts, and topics affect learners' writing performance across lexical and grammatical metrics. This section sheds light on the crucial yet understudied aspects of closeness and narrative voice, which serve as focal points for the subsequent investigations. Questions regarding LD and grammatical accuracy in relation to varying topics are posed, setting the stage for the following sections and research questions.

Sections 2.4 and 2.5 review the existing literature pertaining to the dependent variables of Study 1 and Study 2: LD and SPP, respectively. They give an overview of past theories, research findings, and methodological choices that inform the later analyses, laying the groundwork for further discussions and conclusions. Lastly, Section 2.6 presents these dissertation's research questions and hypotheses.

## 2.2 Spanish learner corpora

The use of corpora has long been central to Spanish philological, linguistic, and lexicographic endeavors, dating back to the inception of the Real Academia Española (RAE). The RAE has compiled three large first-language (L1) Spanish corpora that cover the evolution of the language from its earliest traces to the 21st century: the Corpus Diacrónico del Español (CORDE; Real Academia Española, s.d.), the Corpus de Referencia del Español Actual (CREA; Real Academia Española, s.d.) and the Corpus del Español del Siglo XXI (CORPES XXI; Real Academia Española, s.d.). Additionally, the two-billion-word Corpus del Español, Web/Dialects (Davies

2016) offers a broader representation of Spanish varieties sourced from web pages across the Spanish-speaking world. These and other specialized Spanish corpora serve as resources for studying a wide variety of L1 linguistic topics, including morphosyntactic, semantic and pragmatic phenomena (for an overview of research studies, see Parodi, Cantos-Gómez, and Howe 2022). L1 corpora are also slowly emerging as pedagogical tools in the Spanish L2 classroom (for a comprehensive review and discussion, refer to the introductory article of Sánchez-Gutiérrez, De Cock & Tracy-Ventura, 2022).

In recent decades, there has been a heightened interest in the varieties of Spanish used by L2 learners, within corpus linguistics. This has been reflected in the compilation of LC that collect written or/and oral texts by this population of speakers at different proficiency levels. Traditionally, L2 learners' linguistic development has been examined through controlled data, such as grammaticality judgment tests, cloze tasks or vocabulary tests. Granger (2009) notes that the traditional preference for controlled data over more naturalistic and unconstrained production stems from the drawbacks associated with the latter, notably the lack of control over various variables and the difficulty of gathering enough data related to specific language features.

In this context, LC, which can contain substantial amounts of data from learners at diverse proficiency levels, have emerged as a viable alternative or complement to more controlled tasks to address SLA questions (for a comprehensive review, see Granger, Gilquin, & Meunier, 2015 and Tracy-Ventura & Paquot, 2021). These electronic collections of text also include metadata such as age, first language, and educational background, facilitating controlled and robust statistical analysis. Because of their large size, LC have the potential to provide enough occurrences of relatively infrequent linguistic items. Moreover, while it is true that different LC collect data that fall at numerous points on the scale of naturalness, LC data often

stem from more open production tasks, such as written compositions, and they can be combined with controlled experimental tasks to gain deeper insights into learners' language acquisition processes. Despite the primarily research-oriented use of LC, they also hold potential for informing pedagogical materials. For instance, analyses of LC can identify linguistic structures that learners tend to overuse or underuse, aiding in the design of tailored instructional materials (Miguel & Bonilla, 2022).

However, not all languages receive equal representation in LC research. Among the 137 LC cataloged by Granger et al. (2015) on the Learner Corpora Around the World website, 60% are of L2 English and only 8% focus on L2 Spanish. In their extensive search across directories and publications, Rojo et al. (2022) identified 37 L2 Spanish LC, significantly fewer than those available for L2 English. This disparity stands in contrast to the importance of L2 Spanish, with more than 8 million learners in the United States and 23 million worldwide (Instituto Cervantes, 2023).

Nonetheless, the development and distribution of new Spanish LC are steadily progressing. These corpora employ diverse data collection methods and, as a result, differ in variables such as type of participants, proficiency level classification, objectives, prompts used to elicit learner productions, programs in which the learners are enrolled, among others. Although LC can encompass both spoken and written data, written texts are more prevalent due to their easier data collection process.

When leveraging LC to better understand writing outcomes and L2 development, it is crucial to consider the prompts and the type of elicited response. This is because what participants write in an LC depends significantly on the prompts they are given. Given the central role of prompts in LC, it is valuable to study how different types of prompts impact

learner writing. Once we understand the impact of these prompts, we can establish best practices in corpus design concerning the creation and use of prompts. However, not all L2 Spanish corpora offer the possibility to control for the prompt variable equally, and therefore to study its impact comprehensively. What follows provides an overview of the different ways in which the main Spanish LC manage their data collection efforts in relation to the writing prompts they include. The six main available written LC in L2 Spanish are the *Corpus de Aprendices de Español* (CAES; Rojo & Palacios, 2016), the *Languages and Social Networks Abroad Project* (LANGSNAP; Mitchell, Tracy-Ventura, & McManus, 2017), the *Corpus Escrito del Español como L2* (CEDEL2; Lozano, 2009), the *Corpus of Written Spanish of L2 and Heritage Speakers* (COWS-L2H; Yamada, et al., 2020), the *Corpus para el análisis de errores de aprendices de E/LE* (CORANE; Cestero & Penadés, 2009), and the *Aprender a Escribir en Lovaina* corpus (Aprescrilov; Buyse & González Melón, 2012). Table 1, by specifying size, participant proficiency level or course, and specific prompts used (if known), also provides insights into the comparability of these written Spanish LC.

Table 1. Summary of main written Spanish LC

| **Name** | **Size** | **Participant level classification** | **Prompts used** |
|---|---|---|---|
| CAES | 2,544 participants<br>6,561 samples<br>1,045,097 words | A1, A2, B1, B2 and C1 (Common European Frame of Reference; CEFR) | *For A1:* email introducing themselves, email describing family, note to housemates.<br>*For A2:* postcard, biography of an admired person, hotel room reservation.<br>*For B1:* letter asking friend for favors, letter of complaint, funny story.<br>*For B2:* application for a university program, essay on importance of new |

| | | | |
|---|---|---|---|
| | | | technologies.<br>*For C1*: movie review, email with formal complaint. |
| LANGSNAP | 27 participants<br>162 samples<br>36,059 words | Intermediate<br>(Spanish Elicited Imitation Test; EIT by Ortega, 2000) | Gay marriage and adoption<br>Legalization of marijuana<br>Taxes on Junk Food<br>*(Three other prompts for oral production + an oral interview)* |
| CEDEL2 | 4,399 participants<br>4,399 samples<br>1,105,936 words | Lower beginner, upper beginner, lower intermediate, upper intermediate, lower advanced and upper advanced<br>(University of Wisconsin's placement test) | *(Participant chooses one)*<br>Region where you live<br>Famous person<br>Film<br>Last year holidays<br>Future plans<br>Recent trip<br>Experience<br>Terrorism<br>Anti-smoking law<br>Gay couples<br>Marijuana legalization<br>Immigration<br>Frog<br>Chaplin |
| COWS-L2H | 1,934 participants<br>5,383 samples<br>1,367,258 words | Introductory – SPA 1, 2, 3<br>Intermediate – SPA 21, 22<br>Composition – SPA 23, 24<br>Heritage – SPA 31, 32, 33<br>(Web-based Computer Placement Exam; WebCAPE 2.0)<br>Upper-division courses | *(All participants respond to the same prompt)*<br>A famous person<br>A perfect vacation<br>A special person in your life<br>A terrible story<br>A description about yourself<br>A beautiful story<br>A place you dislike<br>Chaplin |
| CORANE | 321 participants<br>1,091 samples | A2, B1, B2 and C1<br>(CEFR) | *Unknown* |
| Aprescrilov | 2,700 samples<br>c. 1 million words | A1, A2, B1 and B2<br>(CEFR) | *Unknown* |

Methodological differences between LC are influenced, among other factors, by the data collection pipeline, as some corpora are composed of data from real in-class coursework, whereas others include data that are independent of a particular language course and specifically collected for the purpose of the corpus. For example, among these six main Spanish LC, CAES, CEDEL2, COWS-L2H, and LANGSNAP fall in the second category and keep specific, identifiable, and quantifiable prompts, which do not correspond to any specific course assignment. This design offers clear advantages, allowing for varied analyses based on specific prompts. For instance, researchers may opt to focus on argumentative texts with a technological theme, filtering corpus results accordingly by selecting responses to prompts regarding the importance of new technologies.

In CAES, two or three prompts are assigned to each proficiency level, precluding direct comparisons of texts with the same topic across different proficiency levels. LANGSNAP is a longitudinal corpus with three prompts that are provided to study-abroad students, twice each, over the course of 20 months. In contrast, CEDEL2 and COWS-L2H maintain consistent prompts across levels of proficiency and experience with Spanish. However, CEDEL2 allows participants to select one of the 14 prompts offered, while COWS-L2H's prompts are pre-assigned in every data collection term. Giving participants a topic choice might cause similar effects to varying prompts by level, potentially attracting novice learners to different prompts than their intermediate or advanced counterparts–for instance, a beginner learner will avoid narrative prompts, which require past tense use, and will instead choose descriptions in the present tense.

Conversely, both CORANE and Aprescrilov constitute examples of *convenience sampling* (Alonso-Ramos, 2016), that is, teacher-researchers collect homework assignments,

in-class compositions, or tests from their students. As a result, these corpora have an undefined number of prompts, and the texts collected are heterogeneous in terms of genre and topic. Nonetheless, the advantage of such sampling methods is that it quickens data gathering efforts, as samples are already available without added collection efforts.

Another highly valuable feature of a LC are annotations. Annotations enable the analysis of specific language features within large amounts of text, offering the potential for drawing representative conclusions on the development of L2 Spanish. However, the process of annotation is resource-intensive, and few Spanish LC have undertaken such efforts. While Natural Language Processing libraries like Freeling (Padró, 2011) and spaCy (Honnibal & Montani, 2017) can automate tasks such as part-of-speech tagging or lemmatization, their performance with the Spanish language–especially the Spanish language used by learners–falls short, necessitating manual revision of the annotations.

Among the three freely-accessible written Spanish LC containing over one million words (CAES, CEDEL2, and COWS-L2H), both CAES and CEDEL2 have undergone automatic tagging followed by manual revision processes. Consequently, they incorporate a system for morphosyntactic coding of grammatical categories (part-of-speech tagging) as well as lemma identification. This lemmatization and morphosyntactic annotation of samples facilitate specific searches in their web application and enable automatic data retrieval. For example, in CEDEL2's website a researcher could search for instances of first person singular nominative personal pronouns in lower and upper beginner texts with the topic "Future plans" and download the concordances in a .csv file. Contrastingly, COWS-L2H does not offer a web application to complete queries on specific words or grammatical phenomena. Instead, it provides open access

to raw data and metadata, along with Python scripts for tasks like automatic part-of-speech tagging of texts via Freeling.

In addition and most importantly, COWS-L2H provides error-annotated texts with labels for common L2 Spanish mistakes, namely gender/number attribution or agreement, usage of "a personal", presence or absence of subject pronouns or articles, confusion of preposition usage or of the usage of the verbs *ser* and *estar*, and adjective placement errors. COWS-L2H also includes holistic corrections by graduate-level Spanish instructors, which have been used as training sets for automatic error-correction tool development (Davidson et al., 2020). The error annotations will be used in Study 2 of this dissertation and therefore will be more thoroughly detailed as part of the methodology in Chapter 3.

As will be further explored in the next sections, the methodological choices made by corpus designers may have an impact on the conclusions that can be drawn from the study of learners' writing samples. Indeed, several authors (Yu, 2010; Sadeghi & Dilmaghani, 2013; Gregori-Signes & Clavel-Arroitia, 2015) have demonstrated that varying tasks, genres, or prompts influence certain lexical measures such as LD, which will be the focus of Study 1. Study 2, instead of focusing on lexical characteristics, explores whether or not grammatical accuracy, concretely the variable linguistic feature of SPP, can also be affected by genre and topic-related variables in learner text. In the following sections, I will review the literature pertaining to 1) prompt-dependent factors that have an influence on Spanish writing outcomes across both grammatical and lexical dimensions (section 2.3) and 2) the two specific prompt-affected writing metrics that will be studied in this dissertation: LD (section 2.4) and accuracy in SPP (section 2.5).

## 2.3 Task- and topic-related variables and their effect in writing

L2 learners' writing outcomes, lexical development, and grammatical accuracy are influenced by numerous linguistic and extralinguistic factors. The present dissertation focuses on the effect of task- and topic-related variables on L2 Spanish texts' LD and SPP error rate. However, it is still important to acknowledge the number of learners' individual differences that can, and do, influence lexical and grammatical measures, such as L2 proficiency level (Castañeda-Jiménez & Jarvis, 2014), L1 writing proficiency (Schoonen et al., 2003), anxiety (Cheng, 2008), or motivation (Hashemian & Heidari, 2013). The research on learner-dependent variables highlights the usefulness of corpora, which normally provide a large dataset, as well as of statistical techniques that account for participant-level effects. Notably, proficiency level is a well-studied learner-dependent variable, yet there is still much to learn about its interaction with task-related variables. Therefore, course level is included as an independent variable in all analyses. In what follows, we turn our attention to task-level variables, which constitute the focus of this dissertation.

### 2.3.1 Task complexity

Task complexity, as defined by Robinson (2001), is the inherent difficulty level of a task, determined by factors such as the number of elements that need to be processed, the amount of planning required, and the cognitive load imposed on the learner. In L2 writing, different dimensions of task complexity have been studied, yielding contradicting effects on grammatical accuracy (Zhan, Sun, & Zhang, 2021).

For example, Ellis and Yuan (2004) found that, when reducing task complexity by allowing learners to plan and monitor their writing, grammatical complexity and accuracy

increased. However, Ishikawa (2006) observed that learners who narrated events in the past, a task requiring higher cognitive effort and presenting greater complexity, exhibited greater accuracy and syntactic complexity compared to those narrating events in the present, a less complex task. Kuiken and Vedder (2008) reported similar findings concerning the effect of increased task complexity, where increasing the number of elements required in a text (higher complexity) resulted in higher grammatical accuracy instead of lower. On the contrary, Cho (2019) found that narrations of past events resulted in decreased grammatical accuracy, contradicting Ishikawa (2006)'s findings. Similarly, Rahimi and Zhang (2019) found that an increased number of elements had a negative influence on accuracy, also opposing Kuiken and Vedder (2008)'s results. Zhan, Sun, and Zhang (2021) found no significant effect of task complexity on learners' writing accuracy. The lack of consensus regarding the impact of task complexity on overall grammatical accuracy stems from various factors, including the diversity of accuracy measurement methods as well as task- and topic-related variables.

Concretely, I propose that discrepancies in the findings of these studies could arise from the different prompts used to elicit the analyzed writing samples. For example, both Ishikawa (2006) and Cho (2019) examine differences between Here-and-Now and There-and-Then dimensions. However, Ishikawa's prompts elicit third-person narratives based on a strip cartoon, while Cho's prompts elicit first-person narratives about a past or imagined present success. In terms of grammatical accuracy, writing in the first person might engage learners more deeply with the content, causing them to prioritize narrative fluency over grammatical correctness. Conversely, third-person narratives may create a cognitive distance, allowing learners to focus more on form and structure, thus enhancing grammatical accuracy.

Additionally, Kuiken and Vedder (2008) and Rahimi and Zhang (2019) both control for number of elements required and elicit first-person narratives. However, Kuiken and Vedder's prompts involve writing a letter to a friend planning a vacation, a quite familiar and non-intimidating task. In contrast, Rahimi and Zhang's prompts require making a plan for allocating funds for multiple public projects with detailed justifications. This second, more overwhelming prompt might lead to cognitive overload and reduced grammatical accuracy. To support the argument that it is difficult to draw comparisons between the referenced studies due to the prompts, Table 2 presents the variety of prompts utilized, along with the conflicting conclusions.

Table 2. Summary of studies analyzing task complexity effects on grammatical accuracy

| **Study** | **Prompts** | **Aspect of task complexity under study** | **Results regarding grammatical accuracy** |
|---|---|---|---|
| Ellis and Yuan (2004) | Write a story based on a set of six pictures from Heaton (1975) | Opportunity for planning and monitoring | Higher task complexity, less grammatical accuracy |
| Ishikawa (2006) | 1. Write a story based on the strip cartoon "the supermarket" from Yule (1997) and the following prompt: "Today a woman goes to a supermarket. She enters the supermarket through a door. She is wearing a black shirt. She puts her bag in the shopping cart.<br>She is pushing the cart slowly. Maybe she is planning to buy many things for<br>Dinner."<br>2. Write a story based on the strip cartoon "the supermarket" | Here-and-Now versus There-and-Then | Higher task complexity, more grammatical accuracy |

| | | | |
|---|---|---|---|
| | from Yule (1997) and the following prompt: “Yesterday a woman went to a supermarket. She entered the supermarket through a door. She was wearing a black shirt. She put her bag in the shopping cart. She was pushing the cart slowly. Maybe she was planning to buy many things for dinner.” | | |
| Kuiken and Vedder (2008) | Write a letter to a friend selecting one holiday destination out of five options.<br>1. Take into account these six requirements to choose a Bed and Breakfast in Italy/France.<br>2. Take into account these three requirements to choose a holiday resort in a distant country (Curaçao, Isla Margarita, Madagascar, South-Africa, or Tunisia).<br>(Requirements included factors like the presence of a garden, a quiet location, and opportunities for physical exercise) | Number of elements required | Higher task complexity, more grammatical accuracy |
| Cho (2019) | 1. Think about your most memorable success in the past. You worked so hard to achieve your goals, and you had succeeded in achieving your goals. Now, write one narrative essay about your successful past. You can think of those questions when writing: “What are you thinking? What did you do? What do you see yourself doing? What kind of environment were you in? What types of people were around you?” Describe how you felt in your past successful story.<br>2. Imagine yourself in the future. Everything has gone as | Here-and-Now versus There-and-Then | Higher task complexity, less grammatical accuracy |

| | | | |
|---|---|---|---|
| | well as it possibly could have. You have worked hard and have succeeded in achieving your goals. Now, write one narrative essay about your successful future. You can think of those questions when writing: “What are you imagining? What do you see yourself doing? What kind of environment are you in? What types of people are around you?” Describe what you are imagining. In this imagined scene, how do you feel? What do you do on a typical day? | | |
| Rahimi and Zhang (2019) | 1. Imagine you are the government official in charge of allocating (giving) funds of $5,000,000 worth for public projects. You have received three competing projects for public causes: building a new school for the low-income families in the community, buying new buses to improve old and slow public bus transportation, and building a new dam to resolve water shortage.<br>You should allocate funds for all projects. Please prioritise (choose and rank) the projects and allocate the amount of the fund based on your own view of the importance of the projects for the local people. Please provide reasons and give examples when needed to make your choices as convincing as possible for the local people.<br><br>2. Imagine you are the government official in charge of allocating (giving) funds of $10,000,000 worth for public | Number of elements required | Higher task complexity, less grammatical accuracy |

| | | | |
|---|---|---|---|
| | projects. You have received six competing projects for public causes: reducing air pollution, creating jobs for the unemployed, building affordable accommodation for the low-income families, providing subsidised healthcare for the low-income families, providing free higher education for the high-achievers, and increasing school budgets.<br>You should allocate funds for all projects. Please prioritise (choose and rank) the projects and allocate the amount of the fund based on your own view of the importance of the projects for the local people. Please provide reasons and give examples when needed to make your choices as convincing as possible for the local people. | | |
| Zhan, Sun, & Zhang (2021) | 1. Think about your interest, expectation, and goals of your campus life. Write an essay about those.<br>2. A university is the dreamland for students' academic career, but students also go through challenges in campus life. What do you think are the benefits and challenges of campus life? Give your reasons to support your argument. | Degree of reasoning demands (narrative and argumentative) | No effect of task complexity on grammatical accuracy |

Given the potential impact of prompt differences beyond typical measures of task complexity on overall grammatical accuracy, I want to further explore how different textual genres and topic-related variables influence the accuracy of one specific linguistic element in L2 Spanish, namely, SPP errors. While my hypothesis is that task complexity may or may not affect

general grammatical accuracy, what remains unclear is whether it has an effect on the accuracy of specific grammatical forms, such as SPP errors. Understanding whether certain prompts (and their associated genres and topics) elicit texts more prone to specific grammatical errors will help clarify the role of task complexity on accuracy. This, in turn, can assist LC designers in making informed decisions when selecting writing prompts, as well as aid researchers in interpreting learner data when using LC. Regardless, one consistent result in the literature is that task complexity does not affect lexical aspects such as LD.

### 2.3.2 Genre

To build on the understanding of how task characteristics influence learner writing, it is essential to introduce other variables that do impact LD significantly. Research has shown that textual genre is one such variable. For example, Sadeghi and Dilmaghani (2013) observed that L2 English learners tended to use more complex and diverse lexical items in comparative than argumentative essays. Castañeda-Jiménez and Jarvis (2014) found that L2 Spanish texts written in response to an argumentative task displayed a more diverse lexicon than texts that were narrative. Heng, Pu, and Liu (2023) measured lexical density, variation, and sophistication in L2 English argumentative and expository compositions and saw increased lexical density in argumentative texts but greater lexical variation and sophistication in the expository ones. Lastly, in their analysis of L2 Spanish with COWS-L2H, Sánchez-Gutiérrez and Fernández-Mira (2023) also found genre to exert a significant main effect on LD, with descriptive texts (*A person you like* and *A person you dislike*) displaying greater LD than narrative ones (*A perfect vacation* and *A terrible story*).

The observed genre effect on lexical outcomes suggests that variability in lexical measures may be more closely tied to the content and overall structure of the text than to the external aspects that render a task more complex. This notion aligns with an unexpected finding in Tracy-Ventura, Mitchell, and McManus (2016). In their longitudinal corpus study, participants were tasked with writing argumentative texts about three topics: 1) *Gay Marriage and Adoption*, 2) *Legalization of Marijuana*, and 3) *Taxes on Junk Food*. Interestingly, the prompt about *Gay Marriage and Adoption* systematically elicited more varied vocabulary than the other two, irrespective of when the sample was collected. This thematic effect was also observed in L1 Spanish participants. This indicates that, even when the genre and task complexity levels are kept similar, certain attributes of the specific topic learners are tasked with writing about may influence the LD of their resulting texts.

Fernández-Mira et al. (2021) suggested that the LD of a text could be impacted by factors associated with the topic, potentially tied to emotions and personal experiences. One such measurable factor linked to emotions is valence, representing the inherent positive or negative emotional charge of a word. To gauge valence, native speakers are asked to rate the positivity or negativity of individual words, and the final score is determined by averaging these ratings. Fernández-Mira et al. (2021) assessed the valence of each topic in Tracy-Ventura et al. (2016) by summing up the average valence ratings of the words in the Spanish titles. They discovered that *Gay Marriage and Adoption* had the highest positive valence. As a result, they theorized that the texts discussing this topic would exhibit higher LD scores due to the positively charged nature of the subject, likely evoking stronger emotional responses. Fernández-Mira et al. (2021) then used COWS-L2H data to test the effect of topic-related variables such as valence and closeness on LD, and their results are discussed in the following section. In this dissertation I aim to refine the

concept of closeness by incorporating the variable of narrative voice (first- versus third-person narrator) and to explore its influence on the grammatical accuracy of SPP and LD across two genres: narrative and descriptive.

### 2.3.3 Topic-related variables: Valence and closeness

Previous studies have shown that words with extreme emotional charge, either negative or positive, are processed faster than neutral ones, especially by native speakers (Kousta, Vinson, & Vigliocco, 2009; Kuperman et al., 2014). Advanced L2 speakers seem to show similar modulations to those in monolinguals, but lower-proficiency L2 learners present with a positivity bias in which positive words are recognized and read faster than negative and neutral ones (Conrad, Recio, & Jacobs, 2011; Sánchez-Gutiérrez et al., 2022; Sheikh & Titone 2016). Pavlenko (2012) proposes that a reduced emotionality in the L2 may be due to what she coined as a *disembodied cognition*. She posits that, while children acquire their L1 in a context where words and the multi-sensorial experience of the realities they refer to are embedded in each other, L2 learners' learning conditions may not be as emotionally and sensorially rich, resulting in a lexicon that is partially depleted of such embodied layer of lexical representation. Given the results of the aforementioned studies, it seems like the L2 classroom does provide an emotionally rich environment for learners, and that such context is associated, overall, with positive experiences, flooded with more positive than negative words.

The question of valence effects has also been studied from a corpus-based approach by analyzing production of complete written texts instead of recognition and processing of individual words. Fernández-Mira et al. (2021) found that COWS-L2H learner texts responding to a prompt about a positive event (*A perfect vacation*) displayed higher LD than texts about a

negative event (*A terrible story)* across all proficiency levels, but the effects were stronger in the beginner Spanish courses. In a follow-up study, Sánchez-Gutiérrez and Fernández-Mira (2023) confirmed these findings by adding two other prompts of a descriptive genre: *A person you like* (positive) and *A person you dislike* (negative). Again, results showed that for low proficiency levels (the first three academic terms of Spanish instruction in a university-level language program), the topic valence effect on LD is quite strong, systematically displaying higher LD in positive texts than in negative ones, regardless of the genre. However, when comparing advanced L2 learners to heritage learners, although both groups did not significantly differ in terms of LD, topic valence and participant type did present a significant interaction: only heritage speakers, but not advanced L2 learners, had higher LD scores for the positive than the negative topic. This finding points to the need for more research on the interaction between proficiency and valence effects on both receptive and productive lexical skills.

Another topic-related variable related to emotionality is closeness. Following the same rationale posited earlier, if direct, lived experiences shape the emotional intensity associated with particular words, one may hypothesize that words used to describe a loved person that one interacts with on a daily basis might carry a heightened emotional weight compared to words used to describe a person with whom we lack personal interaction. It follows that such embodied and emotionally charged words would be more accessible and frequently employed compared to those representing more detached or emotionally neutral realities. To test this hypothesis, Fernández-Mira et al. (2021) examined COWS-L2H texts describing a famous person (-closeness) and a special person in the author's life (+closeness). Although the initial hypothesis was that descriptions of familiar people would elicit more varied vocabulary, the findings starkly contradicted this assumption, revealing higher LD scores in texts responding to the *A famous*

*person* prompt. Therefore, the need arises for more research on specific prompt characteristics that could be driving learners to display differing degrees of LD and that could affect grammatical measures as well.

Since LC aim to illustrate naturalistic language, they normally use open-ended prompts such as those listed in Table 1. Open-ended prompts are designed to elicit a wide range of responses from learners, allowing them to demonstrate their language proficiency in diverse and creative ways. This stands in contrast with production prompts used in formal assessments or hypothesis-testing studies such as those summarized in Table 2, which are carefully designed to provide a communicative context and sufficient scaffolding in the form of goals for what to accomplish in the task (Sasayama, Garcia Gomez & Norris, 2021). While highly-contextualized, scaffolded prompts are not ideal for corpus purposes, more open prompts can lead to a high degree of within-prompt variation, as different learners may interpret and respond to the same prompt in different ways. For example, COWS-L2H's prompt *A perfect vacation* elicits responses about past vacations, future vacation plans, or general ideas of an ideal vacation.

### 2.3.4 Subjective interpretation and within-prompt variation

As we have discussed in the previous sections, task- and prompt- related variables can have a great effect on several lexical and grammatical metrics, and it only follows that learners' interpretations of an open prompt would also affect writing outcomes. This is the specific issue that this dissertation aims to address. Since lexical and grammatical metrics are used to assess and describe writing quality, proficiency, and development in both LC research and production-based language assessment (Crossley, 2020), it is crucial for corpus researchers to

consider that the same benchmarks for proficiency cannot always be employed across topics and prompt interpretations.

The studies presented in section 2.3 shed light on the importance of considering prompt effects and topic-related variables such as valence and closeness when characterizing learners' grammatical and lexical development. However, the available research on topic emotionality and closeness leaves many questions unanswered, some of which will be explored in this dissertation and are outlined here:

1. Why do learners use a wider variety of words when writing about a famous person versus a special person in their lives? If closeness does not seem to explain the difference, what other factors could? *A famous person* and *A special person in your life* both elicit descriptions of a third person, but are learners interpreting and answering the prompts differently? For example, a text about a famous person can be written as a biography, while a text about a special person in the author's life can focus on describing the relationship between the two. If that were the case, learners are essentially completing two different tasks, despite both appearing initially as comparable prompts aimed at eliciting descriptions of people.
2. Is grammatical accuracy impacted by topics that differ in closeness? One grammatical aspect that could reasonably be influenced by this variable is the use of subject pronouns, since the degree to which the author wants to insert themselves in the text varies when writing about a special person versus a famous person. Although Spanish is a pro-drop language, meaning explicit subject pronouns are often unnecessary, learners may find varying needs to include them based on their perceived closeness to the subject matter. For example, when a learner is writing a description of themselves, they might already

feel reflected in the language used to describe their life and activities, and therefore not feel the need to explicitly state the subject pronoun as much, which results in less overuse errors than if they were writing about third person. When writing about a third person, subject pronoun errors might also differ depending on whether that person is their mother or Justin Bieber, because the level of author involvement in the description varies.

3. What is the role of textual genre in prompts that elicit texts differing in closeness? Since we know that genre affects both lexical and grammatical outcomes, any effect of closeness observed in descriptive texts might not be observed in narrations, or vice versa. The narrations analyzed here will be those responding to the prompts *A perfect vacation* and *A beautiful story,* which could be written either as a personal narrative or as a story with a third-person narrator.

Since there are reasons to believe that both LD and accuracy in SPP are affected by topics that differ in closeness and narrative voice, the following sections will define, explain and review the literature on those two concepts in L2 Spanish. I will start with a description of LD and its relevance in SLA studies and then proceed to provide a brief overview of SPP and its implications for L2 Spanish learning.

## 2.4 Lexical diversity in L2 development

LD, a measure of lexical richness that refers to the range and variety of words used in a given text, is a crucial aspect of task performance in L2 development. It has been identified as one of the most significant indicators of breadth of lexical knowledge and it has been consistently linked to proficiency, with researchers often employing LD alongside other measures to assess not only language learners' lexical development and outcomes, but also their overall writing

quality (Crossley & Skalicky, 2019; Daller & Xue, 2007; Meara, 1996; Milton, 2009; Salsbury, Crossley & McNamara, 2011). For example, Tack et al. (2017) concluded that LD measures found in short answers ranging from 30 to 200 words were the most informative predictors (compared with other measures, including syntactic, discursive, and readability features) of English L2 writing proficiency, and they were able to distinguish among A1, A2, B1, B2, and C levels of proficiency on the Common European Frame of Reference (CEFR).

In LC research, LD measures offer the added advantage of presenting a comprehensive snapshot of a learner's vocabulary repertoire without requiring manual annotation of the data or reliance on frequency lists. This stands in contrast to grammatical measures of accuracy and complexity, which necessitate error annotations and syntactic analyses, still not automatable for languages other than English. LD also differs from other measures of lexical richness, such as lexical density and lexical sophistication. Lexical density, denoting the ratio of content to function words, requires part-of-speech tagging for computation and reflects textual information rather than language competence (Johansson, 2008; Mavrou, 2016). Conversely, lexical sophistication, reflecting the presence of advanced words, correlates with proficiency and writing quality but entails lemmatization and comparison with frequency lists and lists containing imageability, concreteness, and familiarity ratings, or other word properties. (Daller & Xue, 2007; Meara, 1996; Milton, 2009; Saito et al., 2016; Salsbury, Crossley & McNamara, 2011). In L2 Spanish, where automation challenges persist and word frequency lists are not comprehensive yet, LD emerges as a straightforward and reliable measure for SLA research, particularly advantageous for non-English languages.

The concept of LD simply refers to how many unique words a text contains. Traditionally, LD was computed as a simple type-token ratio (TTR), namely, the number of

distinct words, types, divided by the number of total words in a text, tokens (Jarvis 2002). However, longer texts naturally repeat words more often, leading to artificially lower LD scores as text length increases. To account for this text-length bias, researchers have proposed several other methods to measure LD (Kyle 2020), which rely solely on a list of tokens (the complete text) and a mathematical formula, and do not use part-of-speech, syntactic, or error tags, lemmatization, frequency or word properties information. Currently, MTLD (McCarthy, 2005) is one of the most sophisticated and powerful indices available, as it is independent of text length, even for texts as short as 100 words (Koizumi & In'nami, 2012; Kyle, 2020; McCarthy & Jarvis, 2010). MTLD calculates the mean length of sequential word strings in a text while maintaining a given TTR value[1], providing a more robust assessment of LD independent of text length.

In the field of L2 Spanish, several authors have used MTLD to examine how LD correlates with course level or language proficiency. They have found that MTLD increased by level as expected and desirable. However, as mentioned in section 2.3, there were differences between genres and topics (Castañeda-Jiménez and Jarvis, 2014; Fernandez-Mira et al., 2021; Sánchez-Gutiérrez & Fernández-Mira, 2023) that needed further explanation. This line of inquiry is adopted in Study 1 of this dissertation, which aims to explore how the descriptive and narrative prompts differing in closeness and narrative voice (first- versus third-person narrator) presented in section 2.3.4 may influence MTLD scores in L2 Spanish written texts.

## 2.5 Subject pronoun presence in L2 development

As discussed in section 2.3.1, task and prompt-dependent factors affect overall grammatical accuracy in learner texts. However, a noticeable gap exists in the literature concerning the potential influence of prompts differing in narrative voice, as well as the learner's interpretation

[1] For more methodological details about MTLD's computation, see section 3.4.2 in chapter 3.

of these prompts, on the accuracy of a specific feature of the Spanish language: subject pronoun presence (SPP). As noted in section 2.3.4, overt subject pronoun use is one area of accuracy where we might anticipate an impact when topics differ in narrative voice. For instance, learners may overuse subject pronouns more when writing a story in the third person than in the first person. This section describes the concept of SPP in Spanish and presents the current state of research of this variable feature in contexts of L2 Spanish acquisition, thereby revealing a gap in the literature that has not considered topics as a potential source of variability.

### 2.5.1 Characterization of subject pronoun presence and absence in Spanish

Spanish exhibits pro-drop features, which means that subject pronouns can often be omitted. For instance, instead of saying (1), one could simply say (2), where the subject pronoun is inferred from the verb conjugation due to the one-to-one correspondence between a first person plural subject and the specific verb form.

(1) Nosotros vamos a la tienda
    We go.1pl to the store
    'We are going to the store'

(2) Vamos a la tienda
    go.1pl to the store
    'We are going to the store'

This differs from non-pro-drop languages like English, where subject pronouns are almost always required. In English, saying *Are going to the store* would be considered incorrect; the subject pronoun *we* is required to differentiate *We are going to the store* from *You are going to the store* or *They are going to the store*.

Although there is variability in the use or omission of the subject pronoun in Spanish, the issue of SPP is not random or indifferent. There are times when the appearance of explicit pronominal subjects is anomalous, while in other cases, their presence is possible or even necessary. For example, the subject pronoun becomes explicit for contrastive purposes or when it is the focus of the sentence, in which case it usually appears after the verb, like in (3). Additionally, subject pronouns are often specified to disambiguate between verb endings that coincide in certain tenses, as in (4), where *ella* clarifies the subject due to the ambiguity of verb endings in the imperfect past. Similarly, subject pronouns with gender variation may be used to explicitly denote the gender of the referent, as in (5) (Real Academia Española y Asociación de Academias de la Lengua Española, s.d.; sample sentences 3-5 extracted from CREA, Real Academia Española, s.d.).

(3) Si no lo hace él, lo hago yo

If not it do.3s he, it do.1ps I

'If he does not do it, I do it'

(4) Me pasmaba la naturalidad con la que ella podía convivir con el horror

Me amazed the naturalness with which she could.3s.imp live with the horror

'I was amazed by the naturalness with which she could live with the horror'

(5) Todos (ellas y nosotros) tenemos un espíritu libre

Everyone (they.f and we.m) have.1p a spirit free

'Everyone (them women and us men) has a free spirit'

These examples highlight the nuanced nature of SPP in Spanish, where considerations of syntax, context, pragmatics, and stylistics significantly influence subject pronoun inclusion or omission in discourse. For an L2 learner, mastering this aspect of Spanish grammar requires a delicate

balance to avoid both overuse and underuse, which might otherwise result in misinterpretation or awkwardness in communication. Beyond merely understanding rules, proficiency entails developing a sensitivity to the contextual factors that influence SPP.

The complexity of SPP can be further understood by considering theoretical grammatical perspectives. According to the *Nueva gramática de la lengua española* by the Real Academia Española and Asociación de Academias de la Lengua Española, the alternation between expressed and tacit subjects is influenced by multiple internal syntactic factors. For instance, the tacit subject, lacking phonetic features, is generally non-contrastive, while explicit pronouns can serve a contrastive function (Real Academia Española y Asociación de Academias de la Lengua Española, §33.5a). This implies that a sentence like *Nosotros no vamos ahora a preocuparnos de esa cuestión* (*We are not going to worry about this matter now*) suggests that others might do so, compared to *No vamos ahora a preocuparnos de esa cuestión*, which lacks this contrastive nuance.

Additionally, the use of tacit subjects cannot serve as focal points in a sentence, unlike explicit pronouns, which can be interpreted as such (Real Academia Española y Asociación de Academias de la Lengua Española, §33.5b). For example, in response to the question *¿Llamó Javier? (Did Javier call?)*, one could say *No, llamé yo* (*No, it was me who called)* or *No, YO llamé* (*No, I called*) to emphasize the speaker. However, the response *No, llamé* would be insufficient as the tacit subject cannot receive a contrastive interpretation.

Moreover, third person tacit subjects are common in contexts where the discourse topic has been previously introduced and clearly refers to the same entity (Real Academia Española y Asociación de Academias de la Lengua Española, §33.5c). For instance, instead of *Ellos fueron a*

*la cocina (They went to the kitchen)*, which would imply a contrast, *Fueron a la cocina* maintains coherence with the previously mentioned subject without introducing ambiguity.

These theoretical insights highlight the inherent complexity of SPP in Spanish, suggesting that both syntactic and discursive elements significantly influence the presence or absence of subject pronouns. This complexity is reflected in the external factors discussed in the following section, affecting how L2 learners acquire and use subject pronouns in Spanish.

### 2.5.2 Factors influencing SPP in L2 Spanish

Research on SPP in L2 Spanish acquisition has received attention across various contexts and learner groups (Geeslin et al., 2013, 2015; Linford, 2016; Bessett, 2017; Torres Cacoullos & Travis, 2015). Some studies have examined linguistic factors like proficiency levels and language transfer effects (Geeslin, Linford, & Fafulas, 2015; Fernandez Florez, 2019), while others have focused on socio-pragmatic aspects such as discourse functions and communicative strategies (Bayley and Pease-Alvarez, 1997; Martínez, 2007) . The following section reviews the available literature on which factors seem to influence SPP in L2 Spanish.

Studies examining the acquisition of SPP among L2 Spanish speakers have revealed initial tendencies for overuse in beginner-level learners, which has been attributed to transfer effects from a non-pro-drop L1 to a pro-drop L2. Over time and with increasing proficiency, learners tend to approximate native-like patterns (Bayley & Pease-Alvarez, 1997; Geeslin at al., 2015; Otheguy, Zentella, & Livert, 2007).

A critical variable influencing SPP is switch reference, wherein speakers transition between different referents within discourse by changing the subject from one clause to the next. In native speakers, overt subject pronouns are used at a higher rate in switch reference contexts.

For L2 learners, the significance of switch reference increases across levels of proficiency. This suggests that as learners become more proficient, they become more adept at adjusting their use of subject pronouns based on changes in the subject (Geeslin et al. 2015).

Furthermore, tense, mood, and aspect (TMA) play a role in SPP, primarily for first-year learners, as highlighted by Geeslin et al. (2015). Bayley and Pease-Alvarez (1997) observed a correlation between TMA ambiguity and the use of overt subject pronouns, indicating a tendency for increased pronoun usage under conditions of linguistic uncertainty.

Discourse connectivity or connectedness also influences SPP patterns. The variable of discourse connectivity integrates the two previous elements, switch reference and TMA, by assessing the continuity of subjects and verb tenses across adjacent clauses. Martínez (2007) and Bayley and Pease-Alvarez (1997) noted that overt pronoun usage tends to increase as discourse connectedness weakens across clauses. For example, in a narrative where characters are introduced and their actions are described in separate clauses without explicit reference to previously mentioned subjects, L2 learners may opt for more overt subject pronouns to ensure clarity and coherence in their storytelling.

Regarding the grammatical person of the subject, Bayley and Pease-Alvarez (1997) found that the first and third person trigger more overt subject pronouns than the second person. Additionally, in their study of L2 Spanish learners, Geeslin et al. (2015) found that grammatical person was a significant predictor of SPP in third and fourth-year students. Interestingly, the direction of this effect ran counter to native speaker tendencies: while L2 learners exhibited higher rates of selecting overt subject pronouns with the third person singular compared to the first person singular, native speakers favored more overt subject pronouns in the first person singular over the third. This divergence suggests that learners at this stage might be influenced

by linguistic factors other than those typically observed in native speakers. One such factor, as mentioned by Geeslin et al. (2015), could be the lack of sensitivity to priming effects among L2 learners. Native speakers are highly attuned to priming, where previous linguistic context influences subsequent language production. However, in the study, priming effects were intentionally controlled for, making them irrelevant to the learners' decision-making process. As a result, L2 learners might rely on different cues or linguistic features when selecting subject pronouns, leading to the observed discrepancy in pronoun usage compared to native speakers. Another factor, not discussed by the authors, could be the character of the third person singular as an unmarked form, which might lead L2 learners to compensate by using more explicit subjects. While the first person singular has a clear morphological marker (e.g., the morpheme *-o* in *como*), the third person singular only has a root (e.g., *com-*) and a thematic vowel (*-e*), making it less salient. Consequently, L2 learners might use explicit subjects to address what they perceive as an ambiguity, unlike native speakers, who can interpret the third person singular naturally and implicitly without needing a pronoun.

Additionally, the context of communication, whether formal/high-stakes or informal/low-stakes, has been shown to influence SPP usage among first-year heritage learners of Spanish. Martínez (2007) discovered that anomalous overt subject pronouns were more prevalent in a graded formal writing assignment, whereas a relaxed free writing task, presented as an ungraded brainstorming activity, resulted in fewer SPP errors. The observed disparity was attributed to the pressure on learners, when writing in a formal high-stakes context, to adhere to a rigid template of writing and rhetorical traditions transferred from their schooling in English. However, in such contexts, the transfer of skills from English to Spanish leads to the overuse of subject pronouns due to conflicting grammatical rules and pragmatic conventions governing

subject expression in both languages. Although this study focuses solely on heritage learners, given the absence of research on this aspect concerning L2 learners, it is plausible to consider that L2 learners whose dominant language is English might exhibit similar patterns.

Regarding experience in Spanish-speaking countries, Geeslin et al. (2015) found that the lowest proficiency group demonstrated rates of overt subject selection closer to the highest learner level and native speaker groups, suggesting a potential impact of immersion experience on SPP patterns. However, more research is necessary to fully understand the role of experience in Spanish-speaking countries on SPP variability among learners.

This synthesis reveals the intricate nature of SPP acquisition in L2 Spanish contexts and emphasizes the need for further research into the nuanced factors shaping learners' subject pronoun usage. Despite the efforts reviewed in this section, there is a lack of systematic investigation into how topic variation affects SPP accuracy in written L2 texts. This represents a significant gap in the literature that this dissertation aims to address. The importance of examining topic variation lies in its potential to influence the cognitive and linguistic processes involved in writing. Different topics may elicit varying levels of engagement, emotional response, and familiarity, all of which can impact grammatical accuracy, specifically SPP error rates, in the resulting texts. Study 2 focuses on this issue. In Study 1, the same topic-related variables are considered to analyze the LD scores of learner texts. Understanding if learners use their full vocabulary repertoire or make more grammatical mistakes under certain topic conditions is essential to ensure fair comparisons between texts with different prompts. By studying how topic variation affects both LD and SPP accuracy, this dissertation will enhance our understanding of L2 writing development and inform best practices in LC design and prompt creation, as well as SLA research.

## 2.6 Research questions and hypotheses

This dissertation consists of two studies using the same dataset of L2 Spanish written texts. Study 1 aims to better understand how several topic-related variables impact a lexical measure: how many different words are used, that is, the LD of a text. Study 2 is concerned with the effect of the same topic-related variables on a syntactic characteristic: the number of SPP errors. Another common interest of both studies is the evolution (from a cross-sectional perspective) of these lexical and syntactic features as learners advance throughout the course levels in the Spanish language curriculum.

The three topic-related variables that are considered are:

1. Closeness: whether or not the person described in the text is close to the author in their everyday life. There are three levels of closeness in this dataset: the focus of the description can be 1) the author themselves, 2) a special person in the author's life, like their mother or partner, or 3) a famous person.
2. Narrative voice: whether the text is mainly written in the first or third person.
3. Genre: whether the text is a description or a narration. However, genre is not directly comparable across the datasets due to the fact that Subcorpus 1 (descriptive texts) and Subcorpus 2 (narrative texts) were collected at different times in the academic term.

### 2.6.1 Study 1

In Study 1, LD, concretely MTLD, is the dependent variable under study. The specific research questions (RQs) and hypotheses are the following:

RQ 1: How is the MTLD score affected by course level and closeness to the person described in the text, and consequently by the narrative voice, in descriptive texts?

As explained in Chapter 2, Fernandez-Mira et al. (2021) tested, also with COWS-L2H data, whether everyday-life closeness with the person described impacts MTLD. According to the disembodied cognition hypothesis (Pavlenko, 2012), learners, who do not have as much of a direct embodied experience in their L2, may not make different lexical choices or use more diverse vocabulary when writing about more embodied or emotionally charged topics such as a special person in their life. In Fernandez-Mira et al. (2021), the *Famous* and *Special* texts were compared to each other. The results indicated that *Famous* texts had higher MTLD than *Special* texts, so the proposed hypothesis was contradicted. In this dissertation, RQ 1 is the same question posited in Fernandez-Mira et al. (2021)'s study, but the dataset includes another level of "closeness" by adding the *Yourself* topic to the *Famous* and *Special* topics already analyzed.

RQ 2: How is the MTLD score affected by course level and narrative voice (first- versus third-person narrator) in narrative texts?

Following the same rationale as in RQ 1, it is plausible that writing about personal and more embodied experiences like a family vacation or a beautiful story written in the first person results in higher LD scores than writing about less personal stories written with a third-person narrator. For L2 learners, we might find no difference because both personal and non-personal topics are perceived similarly in terms of emotionality. Either way, it is worth exploring the disembodied cognition hypothesis in narrations too.

As for the effect of level on MTLD, for both RQs 1 and 2, it is hypothesized that LD scores will be higher as course level progresses and learners increase their vocabulary repertoire.

### 2.6.2 Study 2

Study 2 examines subject pronoun expression in relationship with proficiency level and topic-related variables. Concretely, the RQs and corresponding hypotheses are the following:

RQ 3: How is the rate of SPP errors affected by course level and closeness to the person described in the text?

The hypothesis I posit–the "closeness hypothesis"–is that, when writing about more personal topics, that is, when the author describes a person that is close to them, the author will have the subject more present or "embodied" in their mind throughout their writing and, in turn, they will feel less of a need to make it explicit with a pronoun. Thus, the prediction is that learners will make the fewest redundancy mistakes when describing themselves, and then fewer when writing about a special person than when writing about a famous person.

RQ 4: How is the rate of SPP errors affected by course level and narrative voice in descriptive texts?

In relation to RQ 4, two hypotheses, one based on linguistic factors and the other based on extralinguistic factors, are made. Both hypotheses predict that texts written in the third person will contain more SPP errors than texts written in the first person.

At the linguistic level, first (and second) person pronouns are deictic, that is, they refer to an entity that must be identified according to the context of utterance. For example, the referent of *yo* is the speaker, whoever that might be within a given communicative context. Conversely, third person pronouns in these texts tend to be anaphoric, meaning that they refer back to constituents in the sentence or discourse domain. For example, the referent of *ella* in the sentence *Ana es muy cariñosa y me gusta estar con ella* is *Ana*, a person that appears only in the text and is not a participant of the context of utterance. For learners, deictic pronouns might be easier to leave out than anaphoric pronouns because they are tied to the "real world" context (more concrete) as opposed to the textual context (more abstract). Anaphoric pronouns need to be mentioned more (unnecessarily) because they refer to an entity that is not present in the right

then and there, but only in the mind of the writer. Additionally, when writing in the third grammatical person in Spanish one has three options: 1) using a proper noun or noun phrase (*Ana*/*mi amiga*), 2) using a subject pronoun, or 3) omitting any kind of explicit subject. For a first person subject, option 1 does not exist. The more choices a learner has, the more likely it is that they make a mistake.

Another linguistic factor that might make a difference is the fact that first person singular verb forms in the present indicative are morphologically more salient or transparent, almost always ending in *-o*, than third person singular verb forms, which can end in *-a* or *-e* depending on the type of verb. Additionally, the third person singular often lacks specific morphemes, presenting a zero morpheme after the thematic vowel, which corresponds to the verbal conjugation (e.g., *cant-a-Ø*, where *-a* is the thematic vowel for the first conjugation (verbs ending in *-ar*) and a zero morpheme Ø indicates third person singular). In contrast, the first person singular ending *-o* does not correspond to the thematic vowel, and its analysis is more complex. Learners might be less prone to redundantly use the first person subject pronoun because they notice the corresponding verb ending more, as it is distinct from the thematic vowel, and see it as sufficient clue to the reader of the subject of the sentence.

On the extralinguistic front, given that subject pronoun expression is usually taught in the beginner levels and these levels are more focused on practicing language in the "personal/individual sphere", it is likely that teachers have more opportunities to correct first person pronoun overuse as opposed to third person. In other words, beginner learners will mainly be answering questions about themselves, their hobbies, their families, their daily routine, etc., so they will have listened to and produced more examples with an omitted *yo* subject, in turn making less mistakes of this type.

RQ 5: How is the rate of SPP errors affected by course level and narrative voice in narrative texts?

The narrative texts analyzed to address RQ 5 are different from the descriptive texts needed for RQ 3 and RQ 4 in several ways. First of all, learners write them at different points in the academic term: descriptive texts are assigned to them in week 4 and narrative texts in week 8. Because of this, they can't be compared directly to each other. However, it is reasonable to think that learners might make SPP errors at a different rate when describing events or actions (narrations) than when describing people. In a description of a person, the main verbs used will be copulas or a small set of common descriptive verbs: *Soy María, tengo 20 años, soy alta y simpática, estoy contenta y me gusta* bailar. In narrations, or descriptions of actions, a wider set of verbs will likely be used to complete the task. If verbs are more commonly repeated when describing people, perhaps learners will notice the unnecessary presence of subject pronouns more and, as a result, make less mistakes in texts of this genre.

Additionally, descriptive texts predominantly feature stative verbs, while narrative texts are characterized by action verbs. Action verbs are considered prototypical, being more cognitively accessible and easier to retrieve (Tomé Conejo & Recio Diego, 2022). This prototypical nature and higher cognitive availability of action verbs may result in fewer errors compared to stative verbs. Therefore, the difference in verb types between descriptive and narrative texts could also contribute to variations in the frequency of SPP errors. For these reasons, I find it interesting to explore the effect of the same variables as RQ 4 (level and person) in a dataset of narrations.

For RQs 3 through 5, it is hypothesized that there will be less SPP errors as course level, or proficiency level, increases and learners produce more target language output.

This chapter has reviewed the literature on the main variables under analysis in Studies 1 and 2 of this dissertation, presented the identified gaps in the literature, and defined all research questions and hypotheses. Looking ahead, the next chapter will detail the methodologies employed in these studies: Study 1, which investigates MTLD scores in Spanish learner texts varying in closeness, narrative voice, and genre; and Study 2, which focuses on SPP errors within the same subcorpora of texts.

# Chapter 3: Methodology

## 3.1 Introduction

This chapter details the methodology of two studies, Study 1 and Study 2. Section 3.2 describes the corpus used in both Study 1 and 2, COWS-L2H, along with its participants and context. I focus on the methodological decisions that make it possible to address my research questions: the course level distribution and equivalencies (Section 3.2.1) and the careful selection of prompts, with different topic-related variables and genres (Section 3.2.2). The general annotation efforts carried out in COWS-L2H are also described in Section 3.2.3. In Section 3.3, I introduce two specific subcorpora of COWS-L2H that make up the data used in both Study 1 and 2. Section 3.4 explains the methodological aspects involved in the dependent variables of Study 1 (LD; Section 3.4.1) and Study 2 (SPP errors; Section 3.4.2). Section 3.5 details the analytical procedure of Study 1 by defining the models used to answer each research question, and Section 3.6 does the same for Study 2.

## 3.2 Corpus used

Both studies in this dissertation use data from the same corpus, COWS-L2H. In this section I outline the rationale behind selecting COWS-L2H among all other L2 Spanish corpora to answer my research questions. As shown in Table 1 in Chapter 2, COWS-L2H is one of the LC with the highest number of participants, samples, and tokens in the current LC research landscape in Spanish. Its large-scale collection of data is made feasible through the use of the Canvas Learning Management System (LMS), an online classroom platform that is used at the university where the data is collected. Students who participate in the research project enroll in an online

Canvas site where they sign a consent form, read all the necessary instructions, complete a linguistic background questionnaire, and then electronically submit two typed writing samples–one during week 4 and the other during week 8 of the academic term. Each of these compositions needs to adhere to a minimum of 250 –or 150 for participants enrolled in Spanish 1, the course for absolute beginners– and a maximum of 500 words. Participants are given a window of one week to compose and electronically submit each text, at a time and place of their choosing. Every term of data collection, two different writing prompts (for week 4 and 8) are chosen and kept consistent for all participants. The LMS organizes the participants' submissions into a spreadsheet database accessible to the research team. After anonymization, COWS-L2H is made freely available in TXT format to all researchers under a Creative Commons license, via a GitHub repository[2] from which researchers can download the data.

### 3.2.1 Participants and context

COWS-L2H participants are recruited from the Spanish courses they are enrolled in during a given academic term in their university. The university follows a quarter system, so recruitment and data collection efforts take place three times per academic year. Recruitment efforts are carried out by one or more members of the research team by making in-person or virtual announcements at the beginning of each quarter in all of the Spanish classes. These announcements share the importance of students' writing samples to deepen our understanding of Spanish learners' writing development and improve the course curriculum. Students are also offered class extra credit in exchange for their participation in the corpus. However, they are assured that their instructor will not have access to their identified samples and that the texts are not graded. Since this process is repeated every quarter, many students are invited to participate

[2] https://github.com/ucdaviscl/cowsl2h

several times as they advance through the Spanish program. Repeated participation in COWS-L2H is encouraged as long as participants write original texts each quarter. In fact, longitudinal data is a very valuable feature of COWS-L2H.

The language program at the institution where COWS-L2H's data is collected is based on two itineraries that run in parallel: one for L2 Spanish learners and one for Spanish as a Heritage Language learners. After completing the language component, students may enroll in upper division courses on various topics in Hispanic Literature and Linguistics. All of these types of courses (L2, SHL and upper-division) are represented in COWS-L2H, but in this dissertation I will only use the texts written by L2 Spanish students enrolled in lower-division language courses. The L2 Spanish program consists of seven courses organized in three series: Introductory, Intermediate, and Composition. The Introductory series comprises the first year (three quarters) of language study: Spanish 1, 2 and 3. By completing Spanish 3, students can satisfy the foreign language requirement at this university and, as a result, enrollment in the Introductory series is much higher than in the Intermediate (Spanish 21 and 22) and Composition (Spanish 23 and 24) series, which is reflected in COWS-L2H's distribution of the data by course (see Table 3).

Table 3. Summary of COWS-L2H data by L2 course

| **L2 course** | **Participants** | **Texts** | **Tokens** |
|---|---|---|---|
| Spanish 1 | 438 | 785 | 133,493 |
| Spanish 2 | 560 | 1029 | 265,234 |
| Spanish 3 | 647 | 1197 | 313,100 |
| Spanish 21 | 170 | 307 | 80,972 |
| Spanish 22 | 141 | 259 | 70,375 |
| Spanish 23 | 213 | 392 | 110,508 |

| | | | |
|---|---|---|---|
| Spanish 24 | 261 | 499 | 140,962 |
| **Total** | – [3] | 4,468 | 1,114,644 |

As for the proficiency level that can be roughly attributed to these participants based on the course they are taking, we need to rely on scores from the Web-based Computer Placement Exam[4] (WebCAPE 2.0), the placement test that is used at this institution to assign learners to a given course level. Table 4 shows the raw WebCAPE scores necessary to be placed into the corresponding language courses, along with the equivalencies established by the creators of the test between WebCAPE scores and the levels defined by the Common European Framework of Reference (CEFR) (Emmersion, 2021).

Table 4. Correspondence between WebCAPE scores, course placement, and CEFR levels

| **WebCAPE score** | **Course placement** | **CEFR level** |
|---|---|---|
| 0-259 | Spanish 1 | A1 |
| 260-314 | Spanish 2 | A1+/A2 |
| 315-373 | Spanish 3 | A2/A2+ |
| 374-423 | Spanish 21 | A2+/B1 |
| 424-463 | Spanish 22 | B1 |
| 464 and above | Spanish 23 | B1/B1+ |

The linguistic background questionnaire that participants complete before writing their texts provides some more demographic information about them. The sample of L2 participants ranges in age from 17 to 64 (M=20.09, SD=2.79) and 74% identifies as female, 24% as male and 2% as other. As for their language profiles, the majority of the learners (69%) consists of

[3] The total number of individual L2 participants (1,620) is not the sum of the above distribution by course, since some learners participated more than once, while they were enrolled in different courses. The table only shows how many L2 learners participated at each course level, but not how many of them participated again at subsequent course levels.

[4] https://perpetualworks.com/

monolingual L1 English speakers, but there is a considerable proportion of participants whose L1 is Mandarin (10%), or who speak two or more languages (21%).

### 3.2.2 Prompts

The prompts in COWS-L2H were carefully selected in different ways. First, as mentioned in Section 2.2 (Chapter 2), a conscious decision was made to provide all participants with the same writing prompt in a given week and quarter, regardless of their proficiency level or course enrollment. Since by week 4 Spanish 1 students have only learned descriptive verbs (*ser* and *estar*) and very basic vocabulary about introductions and people, this limits the range of possible genres and topics chosen by the researchers. Accordingly, and as a second characteristic of the prompt methodology, the prompt that is assigned in week 4 is descriptive in nature and the one assigned in week 8 is narrative. This genre consistency is maintained across the complete data collection period. Third, the prompts are meant to be open-ended enough such that participants can interpret them in their own way and write in a more unrestricted fashion. The deliberate open-endedness of the prompts is due to proficiency-related constraints but also due to the belief that corpus language should be as naturalistic and representative as possible, and learners should be able to use their full repertoire of linguistic and pragmatic resources. Most of COWS-L2H's prompts allow learners freedom when they approach writing their text, as opposed to more controlled, topic-restricted, closed tasks, which might limit the grammatical and lexical choices made by the writer. Fourth, the chosen topics are general as opposed to specialized in an effort to, again, elicit texts that are more revealing of real learner language.

Table 5. Writing prompts provided per week and quarter of data collection

| **Quarters** | **Week** | **Prompt** |
|---|---|---|
| Spring 2017, Fall 2017, Winter 2018, Spring 2018 | Week 4 | *A famous person (Famous)* |
| | Week 8 | *A perfect vacation (Vacation)* |
| Fall 2018, Winter 2019, Spring 2019, Fall 2019 | Week 4 | *A special person in your life (Special)* |
| | Week 8 | *A terrible story (Terrible)* |
| Winter 2020, Spring 2020, Fall 2020, Winter 2021 | Week 4 | *A description about yourself (Yourself)* |
| | Week 8 | *A beautiful story (Beautiful)* |
| Spring 2021, Fall 2021, Winter 2022 | Week 4 | *Describe a place you dislike (Place)* |
| | Week 8 | *Watch the following clip (4 minutes) and summarize the story (Chaplin)* |

Table 5 compiles a list of the prompts assigned to all COWS-LH participants each quarter and week. A shorter name for the prompt, which will be used throughout this dissertation, is in parenthesis. As the table shows, every three or four academic quarters, the pair of prompts is changed to allow for variation in specific topic-related features:

- Closeness: Writing about oneself (*Yourself*), someone close to the author on a day-to-day basis (*Special*), or someone the author does not personally know and is not involved in the author's life (*Famous*).
- Valence: Writing about positive events (*Vacation*, *Beautiful*) or negative topics (*Terrible, Place*).
- Person: Writing about a personal experience with a first-person narrator (*Yourself, Special, Vacation, Place*) or about a non-personal experience in a third-person voice (*Chaplin*, potentially *Beautiful* and *Terrible*).

- Room for creativity: Telling a story based on a video (*Chaplin*) or based on the author's imagination or memories (*Terrible, Beautiful, Vacation*).

### 3.2.3 Annotation efforts

In late 2019, a team of Spanish native experts from the University of Salamanca in Spain joined the COWS-L2H project for the purpose of annotating and correcting learner texts over the course of several years. As a result, a subset of COWS-L2H has been manually annotated for specific errors that are commonly found in L2 Spanish development, belonging to five categories: gender and number attribution and agreement, presence/absence, exchange and placement errors. What follows details the logistics, the tagging protocol in place, and the error categories.

First, a personal Dropbox folder for each annotator was set up with the original TXT texts that were assigned to them. Every sample was assigned to two annotators so that an inter-annotator agreement analysis was made possible. Then, the annotators performed the tagging, saved the annotated text with the name format "ParticipantID.Quarter_Topic.annotated.txt" (for example: 157048.S17_Vacation.annotated.txt) and uploaded the resulting TXT document to their individual folder. When every annotator finished their batch, a manager compiled all the annotated texts in the same folder.

As for the tagging, the five types of errors are marked with the following bracket format, in the following order:

1. [error]: The word that contains the error is placed inside square brackets as it was written by the participant.
2. {correction}: Curly brackets contain the correct form, as provided by the annotator.
3. <tags>: The specific error tag or tags are typed inside angle brackets.

There are no white spaces between the closing square bracket and the opening curly bracket (]{) nor between the closing curly bracket and the opening angle bracket (}<). However, if two errors follow each other, there is a single space between them:

[error1]{correction}<tags> [error2]{correction}<tags>

If the error is an omission, the square brackets are left empty, without any symbol or white space inside. If the error is an unnecessary word, the curly brackets are left empty.

As mentioned above, only five types of common L2 Spanish errors were the focus of the tagging efforts:

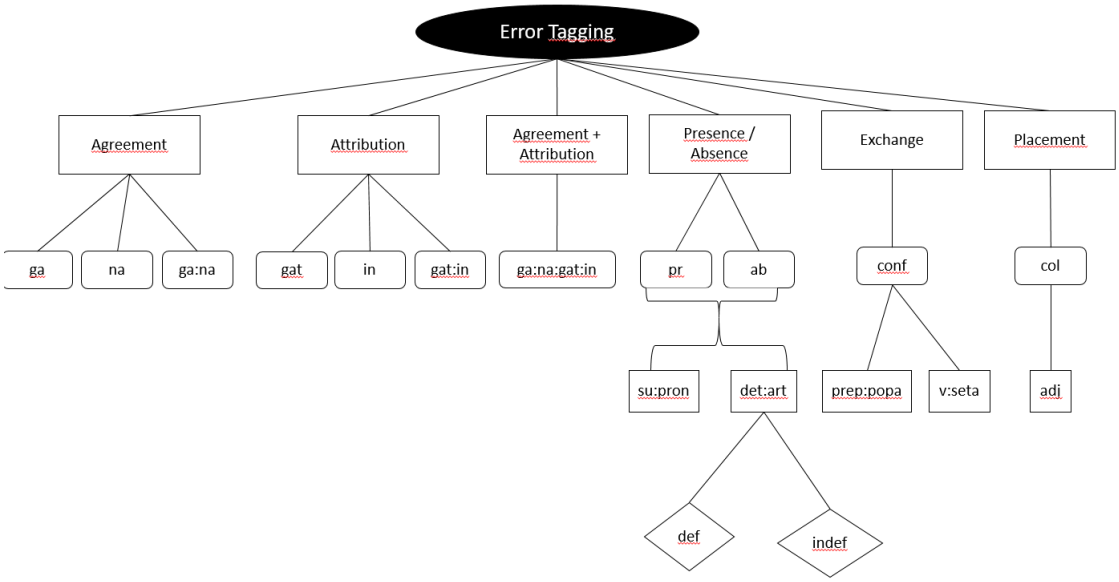
- Attribution errors are made when the learner assigns an incorrect grammatical gender (tagged with the code "gat") or number (tagged with "in") to a noun. The code for the word type of the affected word is added after a colon. For example: *[el]{la}<gat:det> [persono]{persona}<gat:noun>, [La]{Las}<in:det> [vacación]{vacaciones}<in:noun> [perfecta]{perfectas}<in:adj>*.
- Agreement errors were tagged as gender agreement ("ga"), number agreement ("na"), or gender and number agreement ("ga:na") errors followed by a colon and the word type of the affected word: "det" for determiner, "adj" for adjective, "noun" for noun, "pron" for pronoun, or "adv" for adverb. For example: *El chico es [una]{un}<ga:det> [buena]{buen}<ga:adj> [jugadora]{jugador}<ga:noun>*.
- The next category is presence/absence errors. The first part of the tag notes whether there is an extra word (incorrect presence, or "pr") or an omitted word (incorrect absence, or "ab"). The second part specifies the word type, either a subject pronoun ("su:pron") or a determiner article ("det:art"). In the latter case, the determiner is also marked as definite ("det") or indefinite ("indef"). For example: Johnny Depp es un actor famoso.

*[Él]{}<pr:su:pron> nació en 1963 en Kentucky., []{El}<ab:det:art:def> Señor Sandler es muy feliz.* In this dissertation, the errors marked with the tag <pr:su:pron>, that is, the incorrect presence of a subject pronoun, were extracted and analyzed.

- Exchange errors, or errors where the learner got confused between two options, were tagged with the code "conf". Two of the most common errors of this type are the exchange of the verbs *ser* and *estar* and the prepositions *por* and *para*, so these were marked in the second part of the tag with "v:seta" or "prep:popa" respectively. For example: *Te doy gracias [para]{por}<conf:prep:popa> todo, .[Soy]{Estoy}<conf:v:seta> cansado de estudiar.*
- Placement errors (tagged with "col") refer to the ones where the adjective ("adj") is placed in the incorrect order with respect to the noun. For example: *Es mi [favorito papel]{papel favorito}<col:adj>*.

Figure 1 summarizes the different categories and error tags used in the COWS-L2H texts that have been annotated so far.

Figure 1. Summary of COWS-L2H error categories and tags



## 3.3 Subcorpora used in this study

Two subcorpora from COWS-L2H will be used in this dissertation. Subcorpus 1 consists of texts describing a person with varying levels of closeness to the writer, which allows me to examine how this variable affects LD (RQ 1 from Study 1) and SPP errors (RQ 3 from Study 2).Because of the nature of the topics, it also allows for the analysis of how narrative voice (first- or third-person narrator) impacts SPP errors (RQ 4). Subcorpus 2 is made of narrative texts written predominantly in either the first or third person, which will be used to assess the effect of narrative voice on LD (RQ 2 from Study 1) and SPP errors (RQ 5 from Study 2). Due to the different times of collection for Subcorpus 1 and 2—descriptive texts being collected in one part of the term and narrative texts in another—the two subcorpora reflect different proficiency levels and genres. Consequently, while genre will be discussed in the context of its potential influence on the results, direct statistical comparisons between genres are not feasible.

### 3.3.1 Subcorpus 1: *Famous-Special-Yourself*

To carry out the analyses addressing RQ 1 from Study 1 and RQ 3 and RQ 4 from Study 2, I selected the subcorpus composed of annotated texts in all seven L2 Spanish course levels (i.e., Spanish 1, 2, 3, 21, 22, 23 and 24) and written to the *Famous*, *Special*, and *Yourself* prompts. I will call it "Subcorpus 1" or "*Famous-Special-Yourself*". As mentioned in previous sections, Subcorpus 1 texts were all written during the fourth week of the term and represent descriptions of people with different degrees of closeness to the writer. Table 6 contains a descriptive summary of this subset, including the number of texts and tokens organized by course level and by prompt, along with the total number of texts and tokens.

Table 6. Descriptive summary of Subcorpus 1

| **Course level** | **Prompt** | **Texts** | **Tokens** |
|---|---|---|---|
| Spanish 1 | *Famous* | 80 | 16,365 |
| | *Special* | 46 | 7,287 |
| | *Yourself* | 99 | 15,013 |
| Spanish 2 | *Famous* | 100 | 24,455 |
| | *Special* | 55 | 13,880 |
| | *Yourself* | 123 | 30,767 |
| Spanish 3 | *Famous* | 110 | 28,684 |
| | *Special* | 60 | 15,674 |
| | *Yourself* | 131 | 33,388 |
| Spanish 21 | *Famous* | 8 | 2,174 |
| | *Special* | 20 | 4,957 |
| | *Yourself* | 23 | 5,736 |
| Spanish 22 | *Special* | 15 | 4,002 |
| | *Yourself* | 26 | 6,638 |
| Spanish 23 | *Famous* | 9 | 2,347 |
| | *Special* | 11 | 2,797 |
| | *Yourself* | 77 | 21,075 |
| Spanish 24 | *Famous* | 15 | 3,994 |
| | *Special* | 14 | 3,586 |
| | *Yourself* | 69 | 18,440 |
| **Total** | | **1,091** | **261,259** |

### 3.3.2 Subcorpus 2: *Beautiful-Vacation*

The analyses that address RQ 2 from Study 1 and RQ 5 from Study 2 were conducted on the subcorpus made of annotated texts in the three beginner course levels (i.e., Spanish 1, 2 and 3) and written to the prompts *Beautiful* and *Vacation*. The subcorpus is limited to Spanish 1-3 because *Vacation* texts in the subsequent Spanish 21-24 course levels have not been error annotated yet. I will call this subset "Subcorpus 2" or "*Beautiful-Vacation*". Both prompts were provided during the eighth week of the term and elicit positive narrations. Further, to specifically address the RQs in both studies, I had to individually and manually annotate Subcorpus 2 texts on whether they were written primarily in the first person singular (for example, a story about a beautiful day in the writer's life or a recounting of a vacation the writer took with their family) or the third person singular (for example, a fictional story with a third-person narrator). Table 7 shows the number of texts and tokens in each course level, prompt, and narrative voice.

Table 7. Descriptive summary of Subcorpus 2

| **Course level** | **Prompt** | **Narrative voice** | **Texts** | **Tokens** |
|---|---|---|---|---|
| Spanish 1 | *Beautiful* | 1st person | 40 | 7,079 |
| | | 3rd person | 39 | 7,043 |
| | *Vacation* | 1st person | 38 | 8,837 |
| Spanish 2 | *Beautiful* | 1st person | 43 | 11,748 |
| | | 3rd person | 56 | 16,269 |
| | *Vacation* | 1st person | 66 | 16,566 |
| Spanish 3 | *Beautiful* | 1st person | 50 | 13,336 |
| | | 3rd person | 67 | 18,490 |
| | *Vacation* | 1st person | 53 | 13,488 |
| | | 3rd person | 1 | 260 |

| | | | | |
|---|---|---|---|---|
| **Total** | | | **453** | **113,116** |

3.4 Dependent variables

The two dependent variables that are the focus of this dissertation are MTLD (Study 1) and SPP error rate (Study 2). It is interesting to examine these linguistic characteristics in the context of topic-related variables because they represent both lexical and grammatical aspects of texts that might be influenced by learners being assigned one topic as opposed to another. Given that the field of LC research is concerned with learners' language at different points of their development, it seems relevant to learn more about how a given prompt influences how many mistakes are made or how many different words are used.

3.4.1 Study 1: MTLD

In this dissertation, MTLD is the focus of Study 1 and two related RQs:

*RQ 1: How is the MTLD score affected by course level and closeness to the person described in the text, and consequently by the narrative voice, in descriptive texts?*

*RQ 2: How is the MTLD score affected by course level and narrative voice (first- versus third-person narrator) in narrative texts?*

MTLD is defined as the calculation of "the mean length of sequential word strings in a text that maintain a given type-token ratio (TTR) value" (McCarthy & Jarvis, 2010: 384). After an extensive analysis of English texts with different genres, McCarthy and Jarvis (2010) established this predetermined TTR value, called the stabilization point, to be 0.72. This value is therefore considered the point at which TTR tends to become stable, that is, not dependent on text length. To calculate MTLD, the number of word strings that fall below this 0.72 TTR threshold is counted. Each word string with the stabilization point of TTR is called a factor. Then, the total

number of tokens (*N*) in the text is divided by the total number of factors (*F*), and the result is the final MTLD score (see Equation 1). Thus, the longer the segment before TTR values fall below 0.72, the higher the MTLD score and the more lexically diverse a text is assumed to be (Kyle, 2020). For illustrative purposes, Tables 8 and 9 provide the calculation of MTLD and TTR for two sample texts from COWS-L2H's Subcorpus 2.

Equation 1. *MTLD calculation*

$$\text{MTLD} = N \div F$$

*N*: number of tokens

*F*: number of factors

Both texts have an almost identical TTR but, as we read them, in Sample text 1 (see Table 8) we notice a repetition of lexical items from one sentence to the other. For example, the learner finishes sentence 5 writing about getting to the hotel room and starts the next sentence with the words "in my big hotel room". This happens numerous times throughout the text and affects its LD score because of MTLD's calculation, which takes into account how well the words are distributed within the segments. Conversely, Sample text 2 (see Table 9) receives a higher MTLD score because it does not show the same kind of repetition.

Table 8. Sample text 1: Number of tokens and types, TTR calculation, and MTLD score

| | |
|---|---|
| **Sample text 1** | En mi vacación perfecto, viajaría en Las Bahamas. Viajaría a Las Bahamas porque me gusta las playas, el mar y el sol mucho. Cuando llego allí, cambio en mi traje de baño. Siguiente, correría en la playa al mar. Nadaría mucho en el agua hasta cansado, entonces yo caminaría a mi habitación de hotel. En mi grande habitación de hotel, guardaría mi ropa y ducha. Luego, dormiría en mi grande, fresco cama. El próximo mañana, comería desayunar en una café cerca de el hotel. En el café, como huevos, patatas, pan, y tocino. Después de comer, veía el periódico y tómate mi tiempo. Durante la mañana, caminaría por la ciudad. Pararía y comprar en las tiendas pequeñas, compra regalos para mi familia y mis amigos. Para mi mama, visitaría la librería porque ella le gustan los libros. Para mi papa, compro tres ganchos de pesca para cuando pescamos. Para mi hermana y mis amigas, compraría las camisetas y cosas en las tiendas de souvenirs. Para mí, compraría un |

| | |
|---|---|
| | libro de historia de Las Bahamas y el Caribe. Me gusta la historia mucho, y las gentes y sus culturas. Durante la tarde, yo regreso a mi habitación del hotel, y dormir una siesta. Después de mi siesta, leería mi libro nuevo por dos horas. En la noche, visitaría un restaurante agradable, y beber limonada. Entonces, ordenaría un bistec y las patatas fritas. Me gusta mi bistec poco hecho. Después, ordenaría el helado vainilla con chocolate salsa. Me gusta relajarme y las cosas simples en mi vacaciones. |
| **Tokens** | 255 |
| **Types** | 132 |
| **TTR** | 132 / 255 = 0.52 |
| **MTLD** | 53.25 |

Table 9. Sample text 2: Number of tokens and types, TTR calculation, and MTLD score

| | |
|---|---|
| **Sample text 2** | Mis vacaciones perfectas es ir a Hawaii. me gustaría ir a todas las islas, especialmente *PLACE*. Que es donde nací. Me gustaría comprar una habitación de Hotel en la playa que tiene vistas a la playa. Me volvería a alojar aquí por una semana. Me encantaría caminata manoa falls trail. He oído la vista es hermosa y la caminata no es agotadora. Yo también soy un fan de la actividad física, por lo que esta sería una gran oportunidad para descubrir Hawaii. También me gustaría ir a bucear en la bahía de Hanauma. Nunca he ido a bucear antes. para la comida, me gustaría probar sushi de Hawai porque debe ser fresco. También me gustaría comprar sus malasadas, que es una rosquilla hawaiana. He tenido estos antes y son deliciosas y caliente. Después, quiero comprar en sus centros comerciales. Tal vez incluso voy a conseguir un tatuaje aquí. También me gustaría ir a la playa con mi familia y tan. Finalmente, me gustaría ir a un auténtico luau con mi familia y disfrutar de música hawaiana y comida. Quiero aprender a surfear también porque siempre me pareció que era divertido. Cuando tengo veinte uno, me gustaría ir a un club con mis amigos más cercanos. También me encantaría tener una luna de miel aquí cuando soy mucho mayor. Quiero comer comida que no se puede tener en California. Cuando yo era un bebé, mi madre me dijo que me encantaba comer Poke. Sería bueno comer Poke hawaiano una vez más. ¡ Estas son mis vacaciones perfectas! |
| **Tokens** | 256 |
| **Types** | 132 |
| **TTR** | 132/256 = 0.52 |
| **MTLD** | 70.44 |

As it was mentioned before, MTLD has become one of the standard measures when quantifying LD. As studied in the literature and briefly demonstrated here, in addition to how many different types a text has, it also captures their distribution. Therefore, MTLD was the chosen LD index in this dissertation. The specific Python implementation that was used was that by Frens (2017).

### 3.4.2 Study 2: Subject pronoun presence error rate

Study 2 aims to address the following three RQs, all with SPP error rate as a dependent variable:

*RQ 3: How is the rate of SPP errors affected by course level and closeness to the person described in the text?*

*RQ 4: How is the rate of SPP errors affected by course level and narrative voice in descriptive texts?*

*RQ 5: How is the rate of SPP errors affected by course level and narrative voice in narrative texts?*

As explained in Chapter 2, in Spanish, when the subject of a sentence is a personal pronoun, it only reiterates some of the information provided by the verb ending–concretely, the person and number information. Therefore, the subject personal pronoun is not realized in oral or written production unless it increases the informative content of the sentence. This occurs in three cases: 1) when the subject is a piece of new information being highlighted in a focus position, 2) when there is an expression of contrast, and 3) when the referent needs to be clarified because the verb ending may correspond to more than one grammatical person. The following are examples of these three situations where the subject pronoun is necessary, extracted from COWS-L2H texts:

1) *Sus clientes se quejaron de ella y le echaron la culpa al color de su piel.* ***Yo*** *no entendí por qué... ¿por qué?* (*Beautiful*, participant 228875)
2) *Comparto una habitación con Anh, mi compañera de cuarto.* ***Yo*** *soy limpia y* ***ella*** *es un poco sucia.* (*Yourself*, participant 176665)
3) *Anthony era muy macho, y* ***yo*** *no sabía que* ***él*** *era gay, pero en todos los días que trabajé me di cuenta.* (*Beautiful*, participant 205678)

For the annotation process of SPP errors, native Spanish speakers were trained on the linguistic phenomenon and three exceptions described above. They were told to be conservative and only mark the obvious cases of redundancy, where the subject pronoun clearly did not add any information and could not fit in any of the three categories. As mentioned in Section 3.2.3, clear errors were marked with the tag <pr:su:pron>.

## 3.5 Study 1's analytical procedure

### 3.5.1 Research question 1

Two mixed-effects regression models were run using the lme4 package in R to describe the effect of both course level and topic, as well as their interaction, on MTLD scores in Subcorpus 1 (*Famous-Special-Yourself*). The two models differ only in their reference levels, allowing for comparisons across the three topic categories. These analyses address RQ 1: *How is the MTLD score affected by course level and closeness to the person described in the text, and consequently by the narrative voice, in descriptive texts?* As defined in Section 2.6.1, the prediction is that there is no difference between *Special* and *Yourself* texts in terms of LD, but that *Famous* texts are outliers and will display higher LD, based on Fernandez-Mira et al. (2021)'s results.

The models' parameters were set as follows:

- Dependent or outcome variable: LEXICAL_DIVERSITY, as MTLD scores.
- Predictor 1, fixed effect: LEVEL, coded as a discrete variable with successive-difference contrast. This variable has seven ordered course levels or categories: Spanish 1, 2, 3, 21, 22, 23, 24.
- Predictor 2, fixed effect: TOPIC, a nominal categorical variable. Because there were three topics or categories in this subcorpus, the first model used *Famous* as the reference level and compared *Special* and *Yourself* to it. The second model had *Special* as the reference category and thus addressed the remaining comparison, *Yourself* to *Special*.
- Predictor 3, random effect: PARTICIPANT. A random intercept for PARTICIPANT was included to account for individual differences and the fact that each participant may idiosyncratically have different LD scores. The inclusion of random slopes for level and topic was not possible because not enough participants participated in different levels or wrote in response to different topics (i.e. the number of observations was too low).

### 3.5.2 Research question 2

As the second part of Study 1, I carried out an analysis to address RQ 2: *How is the MTLD score affected by course level and narrative voice (first- versus third-person narrator) in narrative texts?* It is run on Subcorpus 2 (*Beautiful-Vacatio*n) and describes the effect of both level and person, along with their interaction, on MTLD scores. The parameters are LEXICAL_DIVERSITY as the dependent variable, LEVEL (only Spanish 1, 2, and 3) and TOPIC (*Beautiful* as 0, *Vacation* as 1) as the two fixed effects, and PARTICIPANT as the random variable. With this model, the disembodied cognition hypothesis is further tested with narrative texts, and we would expect to see no difference in MTLD scores due to a personal versus non-personal topic, especially within the beginner levels represented in the dataset.

3.5.3 Summary of Study 1's models

Table 10. Summary of analytical procedures in Study 1

| | | | |
|---|---|---|---|
| Subcorpus 1: *Famous-Special-Yourself* | RQ 1 | lexical_diversity ~ level * topic + (1 \| participant) | Reference level: *Famous* |
| | | lexical_diversity ~ level * topic + (1 \| participant) | Reference level: *Special* |
| Subcorpus 2: *Beautiful-Vacation* | RQ 2 | lexical_diversity ~ level * person + (1 \| participant) | |

Table 10 compiles a short outline of the three models described in Sections 3.5.1 and 3.5.2. Additionally, in order to understand whether the course level as a whole plays a role in explaining LD, I compared the models presented above to models that only included TOPIC or PERSON as a fixed effect–i.e. removing the LEVEL parameter. These nested model comparisons were performed with a likelihood ratio test and the results are reported in Chapter 4.

3.6 Study 2's analytical procedure

3.6.1 Research question 3

Two mixed-effects regression models were used to address RQ 3: *How is the rate of SPP errors affected by course level and closeness to the person described in the text?* The models describe the effect of both course level and topic, along with their interaction, on the number of SPP errors per 100 words in Subcorpus 1 (*Famous-Special-Yourself*). As explained in Section 2.6.2, the "closeness hypothesis" predicts that the closer the person described is to the author (the learner), the less mistakes of this type they will make. Accordingly, the models' parameters were set as follows:

- Dependent or outcome variable: SPP_ERRORS, or the ratio of SPP errors per 100 words in each text.

- Predictor 1, fixed effect: LEVEL, coded as a discrete variable with successive-difference contrast. There are seven categories corresponding to Spanish 1, 2, 3, 21, 22, 23, and 24.
- Predictor 2, fixed effect: TOPIC, a nominal categorical variable. Because there were three topics or categories in this subcorpus, the first model used *Famous* as the reference level and compared *Special* and *Yourself* to it. The second model had *Special* as the reference category and thus addressed the remaining comparison, *Yourself* to *Special*.
- Predictor 3, random effect: PARTICIPANT. A random intercept for PARTICIPANT was included to account for individual differences and the fact that each participant may idiosyncratically make more or less SPP errors. The inclusion of random slopes for level and topic was not possible because not enough participants participated in different levels or wrote in response to different topics (i.e. the number of observations was too low).

### 3.6.2 Research question 4

RQ 4 is concerned with whether texts written in the first person (with the *Yourself* topic) are different in terms of subject pronoun error rate from texts written mainly in the third person (with the *Special* or *Famous* topics): *How is the rate of SPP errors affected by course level and narrative voice in descriptive texts?* The prediction is that learners will make fewer mistakes of this type when writing in the first person because of the difference between deictic and anaphoric pronouns, the saliency and transparency of first person singular verb forms, and the structure of the course curriculum. To test this idea, another mixed-effects regression model was run on Subcorpus 1 to describe the effect of both course level and narrative voice, as well as their interaction, on SPP errors. The new parameter, PERSON, which replaced the TOPIC variable in the models described in Section 3.6.1, was coded as a binary categorical variable: 0 for first person and 1 for third person.

### 3.6.3 Research question 5

The last model addresses RQ 5: *How is the rate of SPP errors affected by course level and narrative voice in narrative texts?* In this case, Subcorpus 2, *Beautiful-Vacation*, was the dataset used. The hypothesis is the same as in RQ 4. The inclusion of the PERSON variable was possible because Subcorpus 2 was annotated for narrative voice, since some *Beautiful* texts were written with a first-person narrator and some with a third-person narrator. This model was run with the following parameters:

- Dependent or outcome variable: SPP_ERRORS.
- Predictor 1, fixed effect: LEVEL, coded as a discrete variable with successive-difference contrast. Because Subcorpus 2 only contains the first three course levels–Spanish 1, 2, and 3–this parameter is a three-category ordered variable.
- Predictor 2, fixed effect: PERSON, a binary categorical variable for narrative voice. It was coded as 0 for first-person narrator (reference level) and 1 for third-person narrator.
- Predictor 3, random effect: PARTICIPANT.

### 3.6.4 Summary of Study 2's models

Table 11. Summary of analytical procedure in Study 2

| | | | |
|---|---|---|---|
| Subcorpus 1: *Famous-Special-Yourself* | RQ 3 | SPP_errors ~ level * topic + (1 \| participant) | Reference level: *Famous* |
| | | SPP_errors ~ level * topic + (1 \| participant) | Reference level: *Special* |
| | RQ 4 | SPP_errors ~ level * person + (1 \| participant) | |
| Subcorpus 2: *Beautiful-Vacation* | RQ 5 | SPP_errors ~ level * person + (1 \| participant) | |

Table 11 compiles the four models described in Sections 3.6.1-3.6.3. Like in Study 1, for RQs 3, 4 and 5 the effect of the LEVEL variable as a whole was inferred through nested model comparisons and the results are reported in Chapter 5.

# Chapter 4: Study 1 results

## 4.1 Introduction

This chapter presents the results of Study 1. In Study 1, the focus is on investigating how factors such as closeness, narrative voice, course level, and genre influence LD (MTLD scores) among Spanish L2 learners. Two primary research questions and corresponding hypotheses guide the inquiry:

RQ 1: How is the MTLD score affected by course level and closeness to the person described in the text, and consequently by the narrative voice, in descriptive texts?

This question is explored using Subcorpus 1, consisting of texts responding to prompts about a famous person, a special person in the writer's life, and a self-description. Guided by the "closeness hypothesis," I anticipate that learners will demonstrate higher LD when writing about themselves, followed by the topic *Special* and lastly *Famous*. Section 4.2 presents the outcomes of two mixed-effects models addressing RQ 1.

RQ 2: How is the MTLD score affected by course level and narrative voice (first- versus third-person narrator) in narrative texts?

This question is addressed using Subcorpus 2, containing narrations of beautiful stories and vacations. I expect that narratives recounting personal experiences using a first-person voice may exhibit higher LD compared to more detached narratives primarily written in the third person. The mixed-effects regression model results for RQ 2 are described in Section 4.3.

Throughout Sections 4.2 and 4.3, the impact of course level is also evaluated to determine whether observed trends persist across varying levels of Spanish proficiency. In both cases, I hypothesize that, as learners progress through their Spanish language courses, their LD

will expand. This hypothesis aligns with the established notion that language proficiency correlates positively with vocabulary richness.

4.2 Research question 1: Closeness effects in descriptive texts

Tables 12 and 13 display the random and fixed effects of the mixed-effects models that estimate the effect of course level and closeness on MTLD scores in the descriptive subcorpus (RQ 1).

Table 12. Parameter estimates and derived values from the mixed-effects model lexical_diversity ~ level * topic + (1 | participant) – Subcorpus 1 (*Famous-Special-Yourself*) – Reference level: *Famous*

| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 115.3 | 10.74 |
| | Residual | | 287.6 | 16.96 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 67.0482 | 2.4194 | $< 2 \times 10^{-16}$ *** |
| | Level SPA 1-2 | 10.6015 | 2.9872 | 0.000404 *** |
| | Level SPA 2-3 | 1.6186 | 2.6672 | 0.544092 |
| | Level SPA 3-21 | 12.1006 | 7.2345 | 0.094702 |
| | Level SPA 21-22 | -6.2181 | 10.6023 | 0.557677 |
| | Level SPA 22-23 | 1.2282 | 4.3366 | 0.777091 |
| | Level SPA 23-24 | 6.1629 | 8.4377 | 0.465311 |
| | *Special* Topic | -4.9874 | 2.9519 | 0.091409 |
| | *Yourself* Topic | 4.4763 | 2.5784 | 0.082848 |
| | SPA 1-2 × *Special* | -3.3899 | 4.8259 | 0.482597 |
| | SPA 2-3 × *Special* | 10.9010 | 4.3547 | 0.012547 * |

| | | | | |
|---|---|---|---|---|
| | SPA 3-21 × *Special* | -2.0013 | 8.8028 | 0.820204 |
| | SPA 21-22 × *Special* | -0.1712 | 12.5388 | 0.989112 |
| | SPA 22-23 × *Special* | -0.8161 | 8.8411 | 0.926474 |
| | SPA 23-24 × *Special* | -6.6655 | 11.5421 | 0.563735 |
| | SPA 1-2 × *Yourself* | -8.6824 | 3.9985 | 0.030128 * |
| | SPA 2-3 × *Yourself* | 5.0836 | 3.6027 | 0.158586 |
| | SPA 3-21 × *Yourself* | -11.1556 | 8.4440 | 0.186760 |
| | SPA 21-22 × *Yourself* | 9.2445 | 10.7181 | 0.388602 |
| | SPA 23-24 × *Yourself* | -4.8955 | 8.9779 | 0.585669 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.1606956 | | 0.4008196 | |

Table 13. Parameter estimates and derived values from the mixed-effects model lexical_diversity ~ level * topic + (1 | participant) – Subcorpus 1 (*Famous-Special-Yourself*) – Reference level: *Special*

| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 115.3 | 10.74 |
| | Residual | | 287.6 | 16.96 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 62.0607 | 1.7075 | $< 2 \times 10^{-16}$ *** |
| | Level SPA 1-2 | 7.2116 | 3.7842 | 0.057087 |
| | Level SPA 2-3 | 12.5196 | 3.4325 | 0.000292 *** |
| | Level SPA 3-21 | 10.0994 | 5.0147 | 0.044309 * |
| | Level SPA 21-22 | -6.3892 | 6.5871 | 0.332342 |
| | Level SPA 22-23 | 0.4121 | 7.6913 | 0.957286 |

| | | | | |
|---|---|---|---|---|
| | Level SPA 23-24 | -0.5026 | 7.8757 | 0.949133 |
| | *Famous* Topic | 5.1040 | 3.1557 | 0.106094 |
| | *Yourself* Topic | 9.4638 | 2.0039 | $2.64 \times 10^{-6}$ *** |
| | SPA 1-2 × *Famous* | 3.3899 | 4.8259 | 0.482597 |
| | SPA 2-3 × *Famous* | -10.9010 | 4.3547 | 0.012547 * |
| | SPA 3-21 × *Famous* | 2.0013 | 8.8028 | 0.820204 |
| | SPA 21-22 × *Famous* | 0.9873 | 12.1557 | 0.935283 |
| | SPA 22-23 × *Famous* | 6.6655 | 11.5421 | 0.563735 |
| | SPA 23-24 × *Famous* | -5.2925 | 4.6355 | 0.253895 |
| | SPA 1-2 × *Yourself* | -5.8174 | 4.2162 | 0.168151 |
| | SPA 2-3 × *Yourself* | -9.1544 | 6.6438 | 0.168600 |
| | SPA 3-21 × *Yourself* | 9.4157 | 8.6114 | 0.274532 |
| | SPA 21-22 × *Yourself* | 0.8161 | 8.8411 | 0.926474 |
| | SPA 23-24 × *Yourself* | 1.7699 | 8.4975 | 0.835053 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.1606956 | | 0.4008196 | |

The analysis revealed a main effect of TOPIC, specifically between the *Yourself* and *Special* texts. Self-descriptions (*Yourself* topic) exhibited significantly higher MTLD scores compared to descriptions of a special person in the writer's life (*Special* topic). However, no statistically significant differences in MTLD were observed between the *Famous* and *Yourself* topics nor between the *Famous* and *Special* topics. This finding contradicts the results of Fernandez-Mira et al. (2021), who identified higher MTLD scores in essays with the *Famous* topic compared to the *Special* topic. In the current study, a different subcorpus consisting solely of annotated texts was used, and the Yourself topic was included in the analyses. The reduced sample as well as the

inclusion of a third level of comparison might explain why the difference between *Famous* and *Special* did not reach significance.

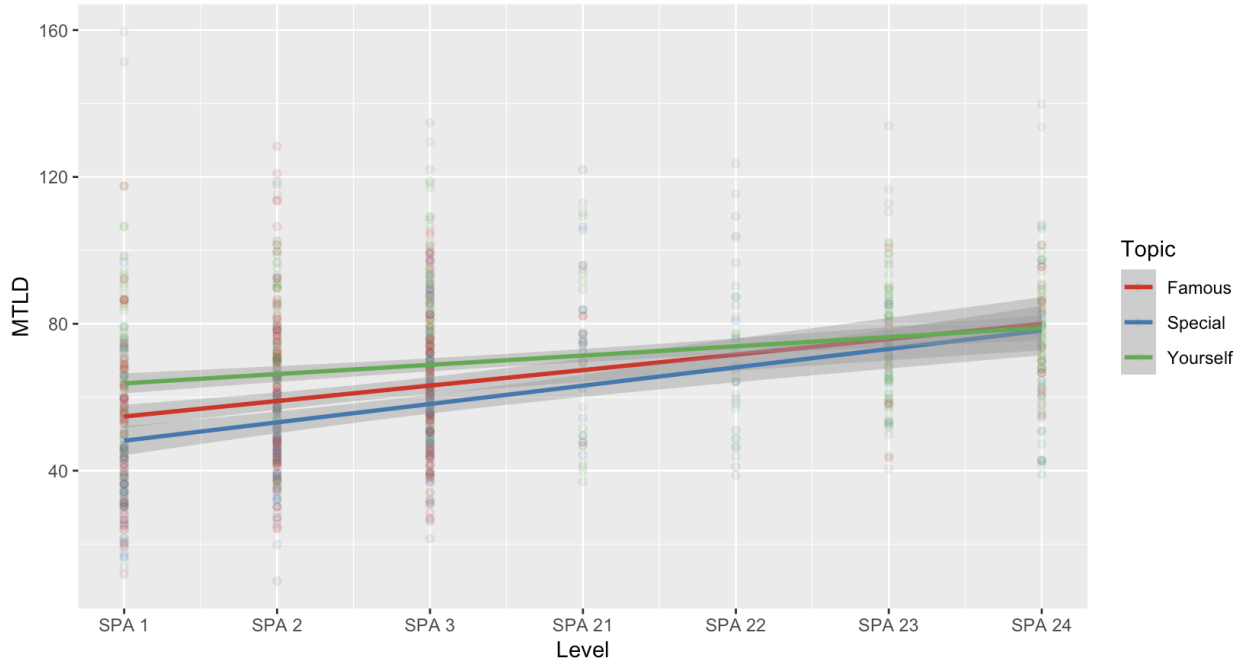
Examining the TOPIC effect, it is apparent that texts written in the first person (*Yourself* topic) displayed higher MTLD scores than those written in the third person (*Special* and *Famous* topics).

Regarding the effect of LEVEL, texts exhibited significantly more varied vocabulary in all consecutive course levels from SPA 1 to SPA 21 (considering both when the reference level is *Famous* and when it is *Special*). To assess the effect of LEVEL as a whole, a nested model comparison was conducted via an ANOVA test on two models: one with and one without the LEVEL variable. The result of the ANOVA confirmed the overall significance of course level in predicting MTLD scores: learners are indeed using more varied vocabulary as they progress through the curriculum.

As for the interactions of LEVEL and TOPIC, two reached significance: 1) MTLD scores increased at a faster rate from SPA 2 to 3 when writing to the *Special* prompt compared to *Famous* and *Yourself*, and 2) MTLD scores did not increase as rapidly from SPA 1 to 2 when writing to the *Yourself* prompt compared to *Famous* and *Special*. Figure 2 provides a visual representation of the significant main effects and interactions in these analyses.

In these models, the fixed effects account for 16.07% of the variance in MTLD scores, while the fixed and random effects together explain 40.08% of the variance. These $R^2$ values indicate a moderate level of explanatory power for the mixed-effects models, suggesting that while the included predictors contribute significantly to explaining the variability in MTLD scores, there are other factors not accounted for that need to be further investigated.

Figure 2. Effect of topic and level on MTLD




4.3 Research question 2: Person effects in narrative texts

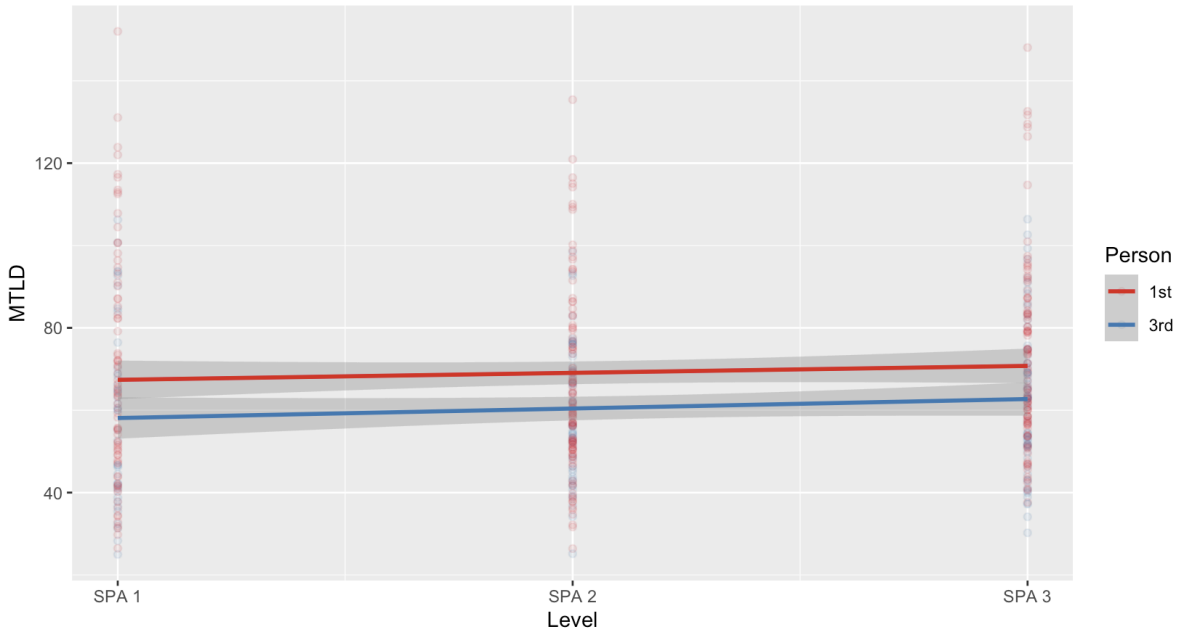
Table 14 presents the estimates from the mixed-effects model examining the influence of course level and narrative (first- or third-person narrator) on MTLD scores within Subcorpus 2 (RQ 2). The same results are depicted visually in Figure 3. Subcorpus 2 comprises narrative texts with the topics *Beautiful* and *Vacation* written by students enrolled in the three beginner Spanish courses (SPA 1, 2 and 3).

Table 14. Parameter estimates and derived values from the mixed-effects model lexical_diversity ~ level * person + (1 | participant) – Subcorpus 2 (*Beautiful-Vacation*)

| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 183.1 | 13.53 |
| | Residual | | 296.1 | 17.21 |
| **Fixed** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |

| | | | | |
|---|---|---|---|---|
| **effects** | (Intercept) | 68.983 | 1.312 | $< 2 \times 10^{-16}$ *** |
| | Level SPA 1-2 | -2.004 | 3.179 | 0.528879 |
| | Level SPA 2-3 | 6.008 | 2.956 | 0.042756 * |
| | 3rd Person | -8.243 | 2.215 | 0.000224 *** |
| | SPA 1-2 × 3rd Person | 3.000 | 5.433 | 0.581182 |
| | SPA 2-3 × 3rd Person | -2.246 | 4.772 | 0.638276 |
| ***$R^2$* values** | ***$R^2$* marginal** | | ***$R^2$* conditional** | |
| | 0.04031885 | | 0.4069672 | |

Figure 3. Effect of person and level on MTLD in Subcorpus 2 (*Beautiful-Vacation*)




The analysis revealed a significant main effect of PERSON, indicating that texts recounting a beautiful story or a vacation tended to exhibit lower MTLD scores when predominantly written in the third person, in contrast to those written in the first person. In other words, learners used more diverse vocabulary when writing narratives from a first-person perspective. Notably, the size of the effect of PERSON was smaller in Subcorpus 2 (narrative texts) than in Subcorpus 1

(descriptive texts). Regarding LEVEL, a statistically significant increase in MTLD scores was found from SPA 2 to 3, but not from SPA 1 to 2. When doing a nested model comparison to assess the effect of LEVEL as a whole, the result was not significant. The $R^2$ values of the model indicate that fixed effects explain approximately 4% of the variance in MTLD scores, while both fixed and random effects together account for around 40.7% of the variance. This suggests a considerable portion of the variability in MTLD scores is attributed to participant-level factors.

# Chapter 5: Study 2 results

## 5.1 Introduction

This chapter presents the results of Study 2, which investigates the impact of closeness, narrative voice, and genre on SPP errors among Spanish L2 learners. Subcorpus 1 contains texts written to three prompts about a person with differing degrees of closeness to the author: *A famous person*, *A special in your life* and *A description about yourself*. According to the "closeness hypothesis", I predict that learners will make less overuse of the subject pronouns the closer the person described is to them. Section 5.2 presents the results of two mixed-effects models addressing this question (RQ 3). In section 5.3, I address RQ 4 by performing the same analysis as in section 5.2 but with narrative voice (first or third) instead of topic as one of the predictors. The hypothesis is that texts written mainly in the first person will contain less mistakes because first person verb forms are grammatically, pragmatically, and pedagogically more salient and thus less prone to require, in a learner's mind, clarification through an explicit subject pronoun. Section 5.4 takes what is learned from the previous results and includes an ad hoc analysis considering the role of a new variable, average sentence length, on these kinds of mistakes. Finally, since Subcorpus 1 only contains descriptions, section 5.5 aims to do the same analysis of person, level and pronoun overuse on narrative texts answering the prompts *A beautiful story* and *A perfect vacation* (RQ 5, Subcorpus 2).

## 5.2 Research question 3: Closeness effects on SPP errors in descriptive texts

Tables 15 and 16 display the random and fixed effects of the mixed-effects models that estimate the effect of course level and closeness on SPP errors (RQ 3).

Table 15. Parameter estimates and derived values from the mixed-effects model SPP_errors ~ level * topic + (1 | participant) – Subcorpus 1 (*Famous-Special-Yourself*) – Reference level: *Famous*

| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 3.148 | 1.774 |
| | Residual | | 3.099 | 1.760 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 2.383 | 0.2963 | $2.33 \times 10^{-15}$ *** |
| | Level SPA 1-2 | -2.107 | 0.3557 | $4.31 \times 10^{-9}$ *** |
| | Level SPA 2-3 | -0.4574 | 0.3130 | 0.144349 |
| | Level SPA 3-21 | -1.686 | 0.8704 | 0.052964 |
| | Level SPA 21-22 | 1.262 | 1.278 | 0.323481 |
| | Level SPA 22-23 | -0.7889 | 0.4942 | 0.111047 |
| | Level SPA 23-24 | -0.1689 | 1.042 | 0.871221 |
| | *Special* Topic | 1.358 | 0.3594 | 0.000167 *** |
| | *Yourself* Topic | -0.7263 | 0.3176 | 0.022420 * |
| | SPA 1-2 × *Special* | 1.714 | 0.5604 | 0.002308 ** |
| | SPA 2-3 × *Special* | -0.7117 | 0.4941 | 0.150345 |
| | SPA 3-21 × *Special* | 0.6830 | 1.049 | 0.515082 |
| | SPA 21-22 × *Special* | -1.881 | 1.500 | 0.210029 |
| | SPA 22-23 × *Special* | 1.650 | 1.021 | 0.106620 |
| | SPA 23-24 × *Special* | 0.1473 | 1.377 | 0.914811 |
| | SPA 1-2 × *Yourself* | 1.696 | 0.4718 | 0.000340 *** |
| | SPA 2-3 × *Yourself* | -0.005059 | 0.4193 | 0.990377 |
| | SPA 3-21 × *Yourself* | 1.698 | 1.004 | 0.091048 |

| | | | | |
|---|---|---|---|---|
| | SPA 21-22 × *Yourself* | -0.6321 | 1.294 | 0.625383 |
| | SPA 23-24 × *Yourself* | -0.4200 | 1.097 | 0.701882 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.2334196 | | 0.6197091 | |

Table 16. Parameter estimates and derived values from the mixed-effects model SPP_errors ~ level * topic + (1 | participant) – Subcorpus 1 (*Famous-Special-Yourself*) – Reference level: *Special*

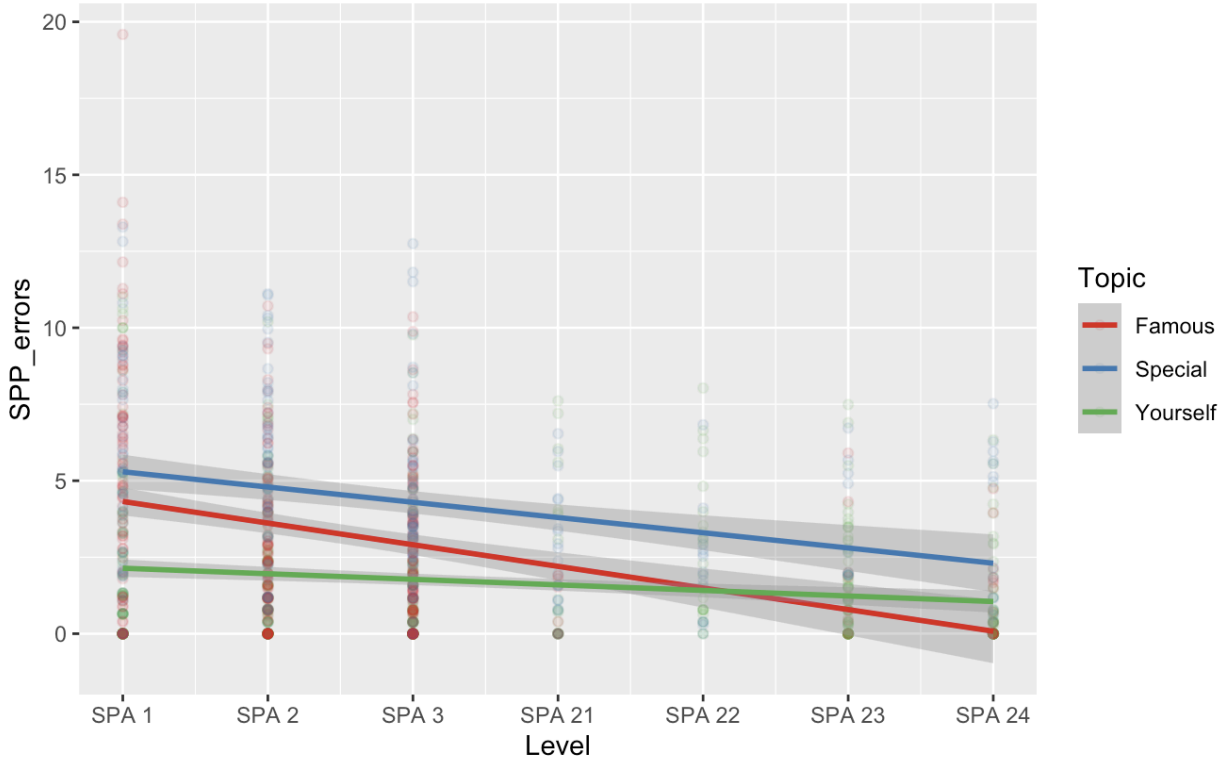
| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 3.148 | 1.774 |
| | Residual | | 3.099 | 1.760 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 3.74102 | 0.20744 | $< 2 \times 10^{-16}$ *** |
| | Level SPA 1-2 | -0.39323 | 0.43168 | 0.36273 |
| | Level SPA 2-3 | -1.16914 | 0.38015 | 0.00226 ** |
| | Level SPA 3-21 | -1.00347 | 0.58493 | 0.08668 |
| | Level SPA 21-22 | -0.61888 | 0.76243 | 0.41725 |
| | Level SPA 22-23 | 0.86153 | 0.89249 | 0.33474 |
| | Level SPA 23-24 | -0.02161 | 0.90022 | 0.98085 |
| | *Famous* Topic | -1.59355 | 0.38298 | $3.43 \times 10^{-5}$ *** |
| | *Yourself* Topic | -2.08404 | 0.24149 | $< 2 \times 10^{-16}$ *** |
| | SPA 1-2 × *Famous* | -1.71359 | 0.56036 | 0.00231 ** |
| | SPA 2-3 × *Famous* | 0.71173 | 0.49410 | 0.15034 |
| | SPA 3-21 × *Famous* | -0.68298 | 1.04881 | 0.51508 |
| | SPA 21-22 × *Famous* | 0.23077 | 1.46104 | 0.87453 |

| | | | | |
|---|---|---|---|---|
| | SPA 22-23 × *Famous* | -0.14733 | 1.37692 | 0.91481 |
| | SPA 23-24 × *Famous* | -0.01728 | 0.53369 | 0.97417 |
| | SPA 1-2 × *Yourself* | 0.70667 | 0.47451 | 0.13708 |
| | SPA 2-3 × *Yourself* | 1.01493 | 0.76868 | 0.18719 |
| | SPA 3-21 × *Yourself* | 1.24916 | 0.99428 | 0.20947 |
| | SPA 21-22 × *Yourself* | -1.65045 | 1.02136 | 0.10662 |
| | SPA 23-24 × *Yourself* | -0.56736 | 0.97283 | 0.55995 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.2334196 | | 0.6197091 | |

A main effect of TOPIC was observed, in which texts about a special person had the most SPP errors, followed by texts about a famous person, and lastly descriptions of the writer (*Yourself* topic), which had the least amount of errors of this type. As for the effect of LEVEL, there are significantly less errors in SPA 2 when compared to SPA 1 when the reference level is *Famous* and also in SPA 3 when compared to SPA 2 when the reference level is *Special*. To assess the effect of LEVEL as a whole, a nested model comparison was conducted by performing an ANOVA test on two models: one with and one without the LEVEL variable. The result of the ANOVA indicated that learners do make less SPP mistakes as they advance through the curriculum. As for the interactions, only one reaches significance: the errors drop at a faster rate from SPA 1 to SPA 2 when writing to the *Famous* prompt as compared to *Special* and *Yourself*. Figure 4 visually illustrates the significant main effects and interactions.

In these models, when only the fixed effects are taken into account, the $R^2$ marginal is small (i.e., 0.23) but when random effects are included, the $R^2$ conditional shows a strong correlation (i.e., 0.62). This indicates that participant idiosyncrasies play an important role on SPP errors and thus participant variables need to be further investigated.

Figure 4. Effect of topic and level on SPP errors




5.3 Research question 4: Narrative voice effects on SPP errors in descriptive texts

Table 17 displays the estimates of the mixed-effects model that analyzes the effect of course level and narrative voice (first- or third-person narrator) on SPP errors in the descriptive subcorpus (RQ 4).

Table 17. Parameter estimates and derived values from the mixed-effects model SPP_errors ~ level * person + (1 | participant) – Subcorpus 1 (*Famous-Special-Yourself*)

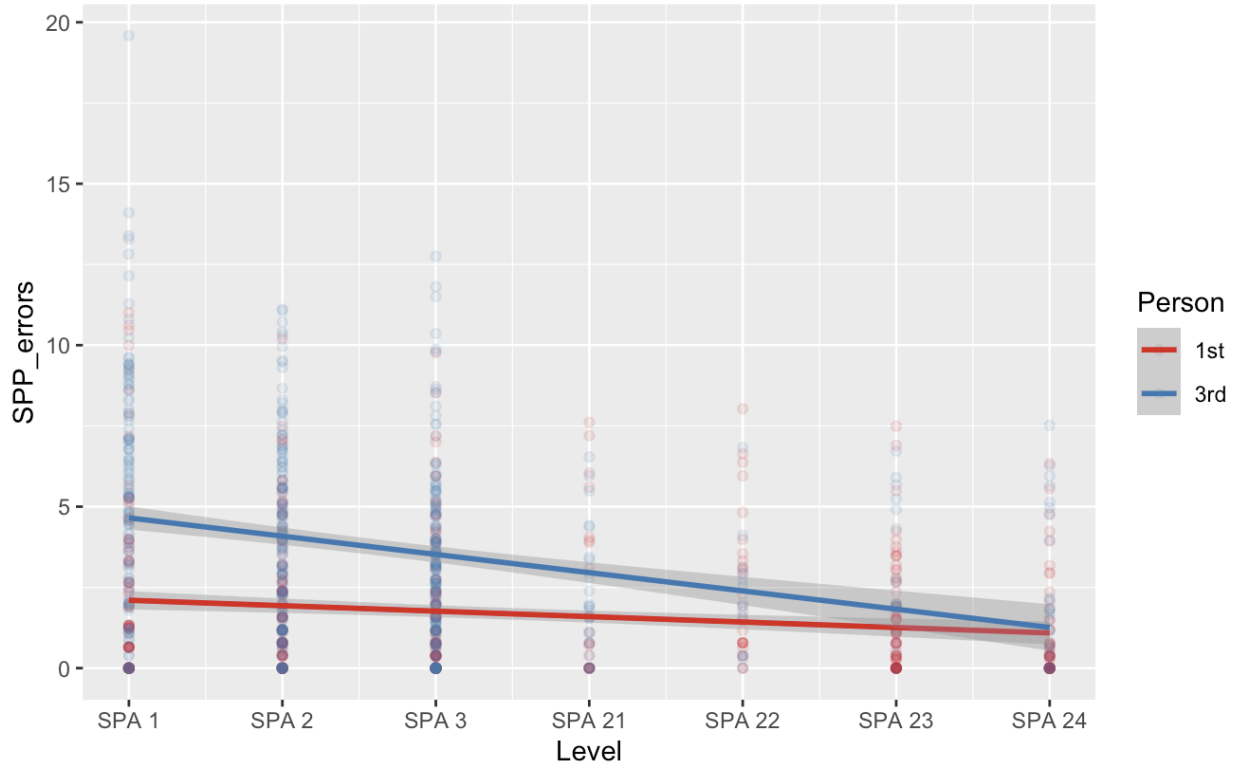
| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 3.386 | 1.840 |
| | Residual | | 3.110 | 1.763 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 1.6184 | 0.1392 | $< 2 \times 10^{-16}$ *** |

| | | | | |
|---|---|---|---|---|
| | Level SPA 1-2 | -0.4414 | 0.3151 | 0.1617 |
| | Level SPA 2-3 | -0.4872 | 0.2826 | 0.0851 |
| | Level SPA 3-21 | -0.1012 | 0.5078 | 0.8421 |
| | Level SPA 21-22 | 0.6692 | 0.6234 | 0.2836 |
| | Level SPA 22-23 | -0.6988 | 0.4993 | 0.1623 |
| | Level SPA 23-24 | -0.5956 | 0.3477 | 0.0874 |
| | 3rd Person | 1.4377 | 0.2100 | $1.27 \times 10^{-11}$ *** |
| | SPA 1-2 × 3rd Person | -1.0025 | 0.4216 | 0.0176 * |
| | SPA 2-3 × 3rd Person | -0.2511 | 0.3746 | 0.5029 |
| | SPA 3-21 × 3rd Person | -0.6160 | 0.6951 | 0.3758 |
| | SPA 21-22 × 3rd Person | -0.7453 | 0.9687 | 0.4420 |
| | SPA 22-23 × 3rd Person | 0.8382 | 0.9371 | 0.3714 |
| | SPA 23-24 × 3rd Person | 0.3914 | 0.7798 | 0.6158 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.2003439 | | 0.6171746 | |

The PERSON variable has a significant effect in the positive direction, which means that these descriptive texts contain more SPP errors when they are written with a third-person narrator versus a first-person narrator (the reference level). This only confirms the results from the previous models (RQ 3), given that texts with the *Special* and *Famous* topics are written primarily in the third person and the *Yourself* topic generates texts in the first person. As for the effect of LEVEL, this analysis did not find any significant effect in the comparison of any two successive course levels. However, when a nested model comparison was calculated, a main effect of LEVEL as a whole was observed. Lastly, only one interaction reaches significance: the

errors drop at a faster rate from SPA 1 to SPA 2 when writing in the third person as compared to the first person. Figure 5 visually illustrates the significant main effects and interactions.

Figure 5. Effect of person and level on SPP errors in descriptive texts




This model has very similar $R^2$ values to the models in section 5.2. Again, the $R^2$ marginal is small (i.e., 0.2) and the $R^2$ conditional is large (i.e., 0.62), which means that the random effect of PARTICIPANT adds meaningful strength to the statistical analysis.

## 5.4 Ad hoc analyses: Effect of level and narrative voice on average sentence length in descriptive texts

After obtaining the results outlined in sections 5.2 and 5.3, one more question arose that could explain the fact that there is a higher incidence of SPP errors when descriptive texts are about a third person (*Famous* or *Special*) than when they are written mainly using the first person (*Yourself*). Is it possible for the narrative voice to affect sentence length? Longer sentences could

mean more SPP mistakes because learners may want to explicitly restate the subject after a long break. If learners were writing longer sentences systematically for certain topics, those topics would in turn produce this overuse of pronouns as observed in section 5.3.

To explore this hypothesis, another mixed-effects model was run with the dependent variable being average sentence length (AVG_SENT_LENGTH) and the independent variables being LEVEL and PERSON, along with a random effect to account for participant variability. The analysis was carried out on Subcorpus 1 (*Famous-Special-Yourself*), the descriptive texts, because it is only in this genre where the difference in error rates was observed. The results are presented in Table 18.

Table 18. Parameter estimates and derived values from the mixed-effects model avg_sent_length ~ level * person + (1 | participant) – Subcorpus 1 (*Famous-Special-Yourself*)

| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 6.401 | 2.530 |
| | Residual | | 3.483 | 1.866 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 11.4436 | 0.1690 | $< 2 \times 10^{-16}$ *** |
| | Level SPA 1-2 | 1.0967 | 0.3673 | 0.002913 ** |
| | Level SPA 2-3 | 1.4496 | 0.3234 | $8.91 \times 10^{-6}$ *** |
| | Level SPA 3-21 | 0.5868 | 0.5760 | 0.308769 |
| | Level SPA 21-22 | 0.4941 | 0.6982 | 0.479555 |
| | Level SPA 22-23 | 2.0579 | 0.5604 | 0.000271 *** |
| | Level SPA 23-24 | 0.3656 | 0.3864 | 0.344697 |
| | 3rd Person | 1.2011 | 0.2502 | $1.82 \times 10^{-6}$ *** |
| | SPA 1-2 × 3rd Person | 0.6863 | 0.4908 | 0.162422 |

| | | | | |
|---|---|---|---|---|
| | SPA 2-3 × 3rd Person | -0.7366 | 0.4274 | 0.085350 |
| | SPA 3-21 × 3rd Person | 1.2612 | 0.7946 | 0.113006 |
| | SPA 21-22 × 3rd Person | 0.7224 | 1.1006 | 0.511915 |
| | SPA 22-23 × 3rd Person | -2.3795 | 1.0715 | 0.026766 * |
| | SPA 23-24 × 3rd Person | 1.3248 | 0.9081 | 0.145012 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.2853858 | | 0.7481664 | |

First, as expected, the impact of LEVEL on sentence length is substantial, with significant increases in average sentence lengths from SPA 1 to 2, SPA 2 to 3, and SPA 22 to 23. Additionally, the result of nested model comparisons reveals that LEVEL as a whole is significant, with learners writing longer sentences as they increase their proficiency.

Notably, the third person element stands out as a significant contributor to longer sentences. When learners are answering to the *Famous* and *Special* prompts, they tend to write longer sentences than when they answer, using the first person, to the *Yourself* prompt. For a qualitative view of this phenomenon, see Table 19, which contains three writing samples from learners at the same proficiency level answering first and third person prompts with simpler and more complex sentences respectively. These results align with the hypothesis that the prevalence of SPP errors within third person descriptive topics might be influenced by these prompts generating responses with longer, and thus more complex, sentences.

Furthermore, the examples in Table 19 shed light on why texts responding to the *Special* prompt might exhibit more SPP errors compared to those responding to the *Famous* prompt. A closer examination reveals less switch reference, and more consistent subjects within and across sentences, in the *Famous* texts, whereas the Special texts exhibit higher switch reference rates,

with subjects changing more regularly. To illustrate these switch references, third-person sentences are highlighted in pink, while first-person sentences are marked in blue. For instance, in response to the Famous prompt, the learner predominantly uses third-person sentences (pink) to describe the celebrity Carli Lloyd, with only occasional references to themselves (blue). In contrast, when responding to the Special prompt, the learner describes their friend, including details about their relationship, shared memories, and common activities, employing both third-person (pink) and first-person (blue) subjects. Despite similar sentence lengths in both types of texts, the increased grammatical complexity inherent in *Special* texts contributes to a higher rate of SPP mistakes.

Table 19. Sample texts from three same-level participants responding to *Yourself*, *Famous* and *Special* prompts

| ***Yourself* (first person)** | ***Famous* (third person)** | ***Special* (third person)** |
|---|---|---|
| Soy china. [Yo]{}<pr:su:pron> tengo el pelo negro y largo. Soy del sur de California. Me gusta comer la ensalada, el pollo, y las verduras. La universidad es difícil. [Yo]{}<pr:su:pron> estudio y trabajar mucho. Estoy muy estresado. Quiero mantenerme en forma, pero no tengo tiempo. Me gusta hacer caminatas por la mañana y jugar al tenis. Soy un poco atlético y encantar hacer ejercicio. Tengo una madre y un hermano mayor y [yo]{}<pr:su:pron> encantar mi familia mucho. Mis amigos dicen que soy graciosa y simpática. Me tomo una ducha cada noche. Me gusta dormir pero no tengo tiempo. Estoy muy cansado y triste. Quiero estar más emocionado y equilibrado. No creo que voy a pasar una de mis clases. Quiero tiempo para estudiar. La vida es muy difícil. Compre un coche ayer y la gasolina es caro. El coche es muy viejo. No soy un buen conductor pero practico en el parque. Hablo con mi mama todos los días por teléfono. Ella me motiva. Nunca estoy en casa. Siempre estoy en la biblioteca. Necesito un descanso de | Carli Lloyd es un jugador de fútbol femenino para el equipo nacional de los Estados Unidos. [Ella]{}<pr:su:pron> juega defensa y ofensiva y es un jugador muy talentoso. [Ella]{}<pr:su:pron> anota muchos goles y es líder en el equipo. Yo también soy un jugador de fútbol y he mirado hacia ella desde que era pequeña. Carli es una jugadora increíble por su desempeño en momentos cruciales. [Ella]{}<pr:su:pron> también es un jugador de equipo y con frecuencia ayuda en muchos goles. Carli ha ganado muchos premios diferentes por jugar al fútbol. [Ella]{}<pr:su:pron> es un olímpico, un ganador de la copa del mundo y ha ganado el jugador más valioso muchas veces. Carli está muy bien decorada, pero [ella]{}<pr:su:pron> sigue siendo humilde y trabaja duro para convertirse en un mejor jugador de fútbol. Carli no siempre fue el mejor jugador que trabajó duro y poner en mucho tiempo para convertirse en el jugador hábil que [ella]{}<pr:su:pron> es hoy. Siempre me ha gustado la forma en que Carli golpea la pelota con autoridad y poder. [Ella]{}<pr:su:pron> es un jugador talentoso, pero debido a lo duro que trabaja, se ha convertido en grande. Carli es de Nueva Jersey y creció jugando al fútbol desde que era pequeña. Era una niña muy activa y le | Una persona especial en mi vida es mi amiga, *FIRST_NAME*. [Ella]{}<pr:su:pron> viva en *CITY* y estudia arte en *UNIVERSITY*. [Ella]{}<pr:su:pron> es vegana y está un razón que yo soy vegana (y que mi gemela es vegana). *FIRST_NAME* y yo fuimos a la escuela mismo en cuarto grado (más o menos). Después graduación, *FIRST_NAME*, mi hermana, nos amiga *FIRST_NAME*, y yo fuimos a *CITY* por nos "senior trip". [Nosotros]{}<pr:su:pron> visitamos muchos museos, muchos restaurantes, y muchos librarias. [Nosotros]{}<pr:su:pron> caminamos cerca de *CITY* por seis días. [Nosotros]{}<pr:su:pron> usamos el metro o tomamos prestado el carro de la tía de *FIRST_NAME*. Cada mañana [nosotros]{}<pr:su:pron> nos despertamos temprano y su tía condujo al estación de tren (la tía de *FIRST_NAME* vive en *STATE*, cerca de *CITY* porque [ella]{}<pr:su:pron> es una bibliotecario en la ciudad). *FIRST_NAME* hizo un itinerario con museos y eventos, y *FIRST_NAME* hizo una lista de |

| | | |
|---|---|---|
| la escuela. Sueno con ser feliz. Después de estudiar me voy a trabajar. Trabajo hasta las 12 de la mañana. Trabajo treinta horas a la semana. El trabajo es divertido pero estoy cansado. Programación de la tarea me lleva 18 horas a la semana. Llevo una camisa azul para trabajar. El dinero para la habitación es caro y no puedo pedir a mi familia para la ayuda. Esta bien. Sueno con una vida mejor después de la universidad.<br><br>*Participant 180855, SPA 2, Winter 2020, SPP error rate 1.1811024* | encantaba jugar al fútbol. [Ella]{}<pr:su:pron> era muy competitiva y se convirtió en un jugador duro jugando con los niños. [Ella]{}<pr:su:pron> atento a la universidad en Nueva Jersey y se convirtió en la goleadora líder de sus estudiantes de primer año y sophmore años. Fueron sus habilidades en el balón, así como su capacidad para distribuir la pelota que la diferenciaba de otros jugadores de fútbol de su edad.<br><br>*Participant 140406, SPA 2, Spring 2017, SPP error rate 2.8985507* | restaurantes vegana en la ciudad. Durante de la día [nosotros]{}<pr:su:pron> andamos (o tomamos el tren), seguimos el itinerario, comimos desayuno y almorzar cuando nosotros tenemos hambre. Cada noche, [nosotros]{}<pr:su:pron> tomamos el tren a la tía de *FIRST_NAME* o ella encontro nos en la ciudad y condujo nos a su casa. [Nosotros]{}<pr:su:pron> hicimos cena (vegana, por supuesto) y leemos los libros nosotros compramos, o [nosotros]{}<pr:su:pron> miramos el nuevo Game of Thrones (esta fue en 2016). *FIRST_NAME* cocina muy bien, me encanta cocinar con ella. [Nosotros]{}<pr:su:pron> tomamos el autobús de *CITY*, *STATE* a *CITY* (y vuelta), por seis o más horas cada tiempo. Cada verano, más o menos, [yo]{}<pr:su:pron> visito *FIRST_NAME* en *STATE* y [nosotros]{}<pr:su:pron> cocinar comida vegana y mirar películas. [Ella]{}<pr:su:pron> es una mi mejores amigas.<br><br>*Participant 152099, SPA 2, Spring 2019, SPP error rate 5.576208* |

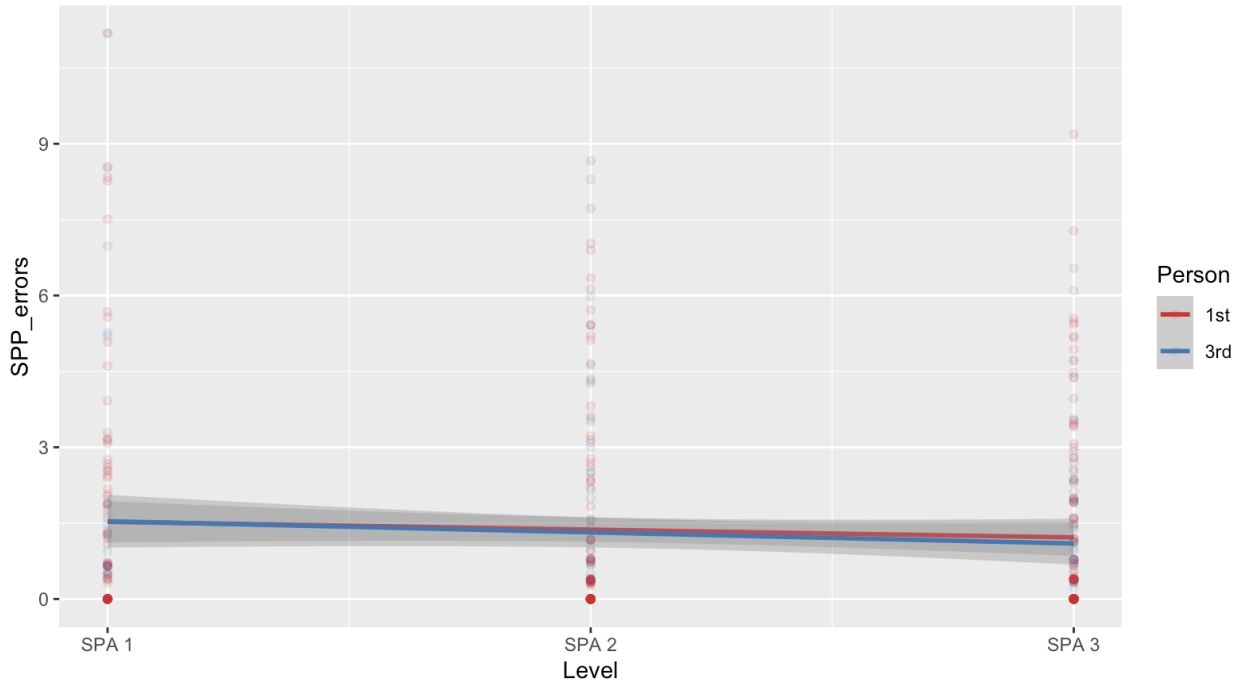
5.5 Research question 5: Person effects on SPP errors in narrative texts

Table 20 displays the results of the mixed-effects model that estimates how SPP errors in Subcorpus 2 are affected by course level and narrative voice (RQ 5). Figure 6 visually illustrates the same results. Subcorpus 2 contains narrative texts with the topics *Beautiful* and *Vacation* written by students enrolled in the three beginner courses (SPA 1, 2 and 3).

Table 20. Parameter estimates and derived values from the mixed-effects model SPP_errors ~ level * person + (1 | participant) – Subcorpus 2 (*Beautiful-Vacation*)

| **Random effects** | **Groups** | **Name** | **Variance** | **Standard deviation** |
|---|---|---|---|---|
| | Participant | (Intercept) | 0.594 | 0.7707 |
| | Residual | | 3.446 | 1.8564 |
| **Fixed effects** | **Parameter** | **Estimate** | **Standard error** | ***p*-value** |
| | (Intercept) | 1.38468 | 0.11988 | $< 2 \times 10^{-16}$ *** |
| | Level SPA 1-2 | -0.37511 | 0.29634 | 0.206 |
| | Level SPA 2-3 | 0.05551 | 0.27490 | 0.840 |
| | 3rd Person | -0.06442 | 0.20242 | 0.750 |
| | SPA 1-2 × 3rd Person | 0.13607 | 0.51045 | 0.790 |
| | SPA 2-3 × 3rd Person | -0.27202 | 0.45130 | 0.547 |
| **$R^2$ values** | **$R^2$ marginal** | | **$R^2$ conditional** | |
| | 0.007076366 | | 0.1530456 | |

Figure 6. Effect of person and level on SPP errors in narrative texts




This model does not result in any significant main effects or interactions. The SPP error rate is statistically the same when writing in first and third person in these narrative texts, and there are no differences between the three beginner levels either. When doing a nested model comparison to assess the effect of LEVEL as a whole, the result was also not significant. As for the $R^2$ values, the $R^2$ marginal is exceedingly small (i.e., 0.007) and the $R^2$ conditional is small (i.e., 0.15), which means that, although participant individualities have some weight, the model as a whole has very poor statistical strength and is not able to explain the variance found in the data.

# Chapter 6: Discussion

This dissertation comprises two studies that utilize COWS-L2H, a Spanish LC that maintains consistent prompts across course levels. Subcorpus 1 consists of descriptive texts covering three topics: *A description of yourself*, *A special person in your life* and *A famous person*; while Subcorpus 2 contains narrations centered around *A beautiful story* and *A perfect vacation*. Study 1 and 2 investigate the impact of topic, closeness to the described person, and narrative voice on the lexical and syntactic characteristics of the texts across different proficiency levels. Specifically, Study 1 concentrates on a lexical measure known as MTLD, while Study 2 focuses on the rate of SPP errors. In this chapter I will examine how the results of these two studies correspond (or not) to the initial hypotheses and existing literature, addressing each research question individually. Additionally, I will explore the methodological and pedagogical implications of these findings and suggest avenues for future research.

## 6.1 Discussion of Study 1: MTLD

### 6.1.1 RQ 1: Descriptive subcorpus (*Yourself-Famous-Special*)

RQ 1 set out to investigate the relationship between MTLD and the perceived closeness of the person described, as well as the narrative voice used, in L2 Spanish learners' written texts. Two competing hypotheses were tested: the “closeness hypothesis” (Fernández-Mira et al., 2021), positing that greater emotional attachment to the person described would lead to higher LD, and the “disembodied cognition hypothesis” (Pavlenko, 2012), suggesting no difference in LD across different levels of closeness due to the lack of emotional context in the classroom environment.

The findings provided partial support for the closeness hypothesis. Specifically, descriptive texts written in the first person (*Yourself* topic) exhibited significantly higher MTLD scores compared to those written in the third person (*Special* and *Famous* topics). This suggests that learners are more competent using a broader range of vocabulary when describing themselves, their interests, and daily routines. These findings align with previous research on the emotional weight native speakers attach to personal experiences and the greater ease of verbal expression in such contexts (Kousta, Vinson, & Vigliocco, 2009; Kuperman et al., 2014; Sánchez-Gutiérrez et al., 2022). Thus, it appears that learners do indeed leverage emotional connections to achieve greater LD, particularly when describing their own experiences.

Contrary to expectation per the closeness hypothesis, no significant differences in MTLD were observed between descriptions of a person close to the participant (*Special*), and a celebrity (*Famous*). This finding challenges the notion that closeness alone drives LD in L2 Spanish learners' written production. The observed lack of differentiation suggests that the emotional richness of the classroom environment may not be sufficient to elicit distinct lexical choices based solely on the perceived closeness of the individual described. Instead, it appears that narrative voice—whether the text is written in the first or third person—may play a more significant role in influencing LD. To further understand this, comparing these findings with native speaker texts could help determine if the effect observed is a feature of narrative voice or if it reflects a broader pattern of self-centered versus other-centered writing.

In addition, Study 1 failed to replicate the findings of Fernández-Mira et al. (2021), which revealed higher MTLD in *Famous* texts than in *Special* texts. This highlights the importance of considering contextual factors and sample characteristics in research. Although both studies utilized similar methodologies, there were two discrepancies within subcorpora: one

study compared only *Famous* and *Special* topics while the other included *Yourself* in the analysis, and one study included all texts while the other only included error-annotated texts. These distinctions may have impacted the observed outcomes.

### 6.1.2 RQ 2: Narrative subcorpus (*Beautiful-Vacation*)

RQ 2 explores the influence of narrative voice (whether the story is told through and first- or third-person narrator) on MTLD scores in narrative texts within the *Beautiful* and *Vacation* topics. Building upon the framework and hypotheses established in RQ 1, this analysis provides insight on whether narratives of personal experiences, such as family vacations or captivating stories about one's life, written in the first person, elicit higher LD scores compared to narratives featuring a third-person perspective. While RQ 1 focused on descriptions of people, the analysis pertaining to RQ 2 extends into the narration of events, exploring whether emotional resonance, as posited by the closeness hypothesis, manifests similarly across diverse textual genres.

Examining lexical measures across different genres is essential, as previous research has demonstrated that textual genre significantly affects LD (Sadeghi & Dilmaghani, 2013; Castañeda-Jiménez & Jarvis, 2014; Heng, Pu, & Liu (2023). Sánchez-Gutiérrez and Fernández-Mira (2023) specifically compared descriptive to narrative texts and showed that genre significantly impacts MTLD, with descriptive texts eliciting higher LD scores than narrative texts. In section 6.1.1, focusing on the descriptive subcorpus, we observed higher MTLD in first-person than third-person text. Given that genre can influence LD, it is important to determine whether the results of RQ 2, using a narrative subcorpus, will follow the same trend as in RQ 1.

The findings of RQ 2 support the closeness hypothesis, which predicts that texts written in the first person exhibit greater LD than those in the third person due to the emotional richness inherent in personal narratives. First-person narratives had higher MTLD scores compared to third-person narratives, which indicates a richer vocabulary repertoire utilized by learners when recounting personal experiences.

The observed increase in MTLD scores in first-person narratives can be attributed to the heightened emotional resonance and embodiment associated with personal storytelling. When narrating from a first-person standpoint, learners draw upon their emotional experiences, enabling a more vivid and nuanced expression of their narratives. Emotionally-charged words are more readily accessible and, because they are more used in personal narratives, first-person prompts produce texts with higher LD.

Conversely, narratives adopting a third-person viewpoint tend to distance the author from the events, potentially decreasing the emotional intensity in the storytelling process. This contributes to a more restrained use of emotionally charged language, resulting in comparatively lower MTLD scores. In other words, the detachment inherent to third-person narratives limits the intensity of emotional engagement, thereby constraining the variety of vocabulary displayed by learners.

In summary, the disparity in MTLD scores between first-person and third-person narratives illustrates the crucial role of narrative perspective in shaping LD within L2 learner texts. Importantly, this trend of higher MTLD in first-person texts is observed in both descriptive and narrative genres. The fact that the effect of narrative voice can be consistently observed across different genres is powerful, considering that genres themselves significantly influence LD. By clarifying the influence of narrative voice on the lexical characteristic of LD, these

analyses contribute to a nuanced understanding of the interplay between narrative context, emotional resonance, and LD in L2 Spanish learner production across two textual genres.

## 6.2 Discussion of Study 2: SPP errors

### 6.2.1 RQ 3 and 4: Descriptive subcorpus (*Yourself-Famous-Special*)

In Study 2, RQ 3 sought to determine how the rate of SPP errors is influenced by both the course level of the learner and the closeness to the person described in the text. RQ 4, while similar, focused on how the rate of SPP errors is affected by the course level and the narrative voice of the descriptive texts. Specifically, RQ 4 examined the differences in SPP errors between texts written in the first person (descriptions of oneself) and texts written in the third person (descriptions of a special or famous person).

The study hypothesized that grammatical accuracy in SPP would be influenced by the everyday closeness of the person described. Specifically, it was expected that descriptions of oneself would have fewer SPP errors due to a more natural and reflective use of language, whereas descriptions of third persons, particularly those not personally known (like a famous person), would show more SPP errors due to varying levels of personal involvement in the description.

In terms of grammatical complexity, it was anticipated that longer sentences and higher instances of switch reference (changing subjects between clauses) would correlate with increased SPP errors, reflecting a trade-off between complexity and accuracy. Additionally, mirroring prior research, it was expected that higher proficiency levels would coincide with fewer SPP errors, indicative of a progression towards more native-like pronoun usage patterns (Bayley & Pease-Alvarez, 1997; Geeslin, Linford, & Fafulas, 2015; Otheguy, Zentella, & Livert, 2007).

Results revealed that texts about a special person had the highest rate of SPP errors, followed by texts about a famous person, and finally descriptions of the writer themselves, which had the lowest rate of SPP errors. This finding partially supports the hypothesis concerning the effect of closeness, as descriptions of oneself showed the least overuse of overt subject pronouns. Additionally, texts written with a third-person narrator (*Special* and *Famous*) contained more SPP errors compared to those written with a first-person narrator (*Yourself*). Interestingly, when comparing *Special* and *Famous*, the results contradicted the closeness hypothesis, with texts about a person closer to the author exhibiting more SPP errors than those about someone less personally relevant.

An ad hoc analysis highlighted that texts responding to the *Famous* and *Special* prompts tended to have longer sentences, contributing to the higher rate of SPP errors. Furthermore, a qualitative examination indicated that texts about a special person exhibited more switch reference, potentially explaining the higher rate of SPP errors observed compared to texts about a famous person. This is because when writing about a special person, learners often discuss their relationship with that person, leading to more switch reference, whereas writing about a famous person typically focuses solely on that individual. The data confirmed that higher grammatical complexity, indicated by longer sentences and more switch reference, led to increased SPP errors. Additionally, the trend towards fewer SPP errors with higher proficiency levels was consistent with previous research, indicating that learners increasingly approximate native-like patterns as they progress.

These results align with the findings of Geeslin et al. (2015), which highlighted the significant influence of narrative voice on SPP errors among third and fourth-year L2 students. Notably, while native speakers favored overt subject pronouns in the first person singular over

the third, L2 learners exhibited the opposite pattern, possibly due to lesser sensitivity to priming effects or the unmarked character of the third-person verb form. This study's findings suggest that L2 learners rely on task complexity cues, such as sentence length and switch reference, and that they overuse third person subject pronouns when these increase as required by the prompt.

In the broader context of previous research on the impact of task complexity on grammatical accuracy, these findings contribute to the ongoing discourse. The literature presents mixed results: some studies (Ellis & Yuan, 2004; Cho, 2019; Rahimi & Zhang, 2019) found that higher task complexity reduced grammatical accuracy, while others (Ishikawa, 2006; Kuiken & Vedder, 2008) found the opposite. A recent study (Zhan, Sun, & Zhang, 2021) reported no effect. This dissertation's findings align with those suggesting that higher task complexity, particularly in the form of longer sentences and increased switch reference, correlates with reduced grammatical accuracy in this specific type of error, SPP errors. However, this might (or might not) apply to learners' grammatical accuracy in other structures and with other measures of task complexity.

### 6.2.2 RQ 5: Narrative subcorpus (*Beautiful-Vacation*)

RQ 5 aimed to explore how the rate of SPP errors is influenced by both the course level of the learner and the narrative voice in narrative texts. This investigation extends our understanding of the impact of narrative voice observed in RQs 3 and 4, which focused on descriptive texts. Understanding these dynamics in narrative contexts, such as recounting beautiful stories or vacations, provides valuable insights into the role of genre on SPP errors.

Previous research suggests that genre can significantly affect grammatical accuracy of SPP errors due to specific characteristics inherent in each genre, such as tense, mood, aspect

(TMA), or discourse connectivity (Bayley and Pease-Alvarez, 1997; Geeslin et al., 2015; Martínez, 2007). However, in alignment with findings from RQs 3 and 4, I hypothesized that narratives primarily written in the first person would exhibit fewer SPP errors compared to those written in the third person. Additionally, I expected to observe a progression in course level tied with fewer SPP errors, consistent with prior research (Bayley & Pease-Alvarez, 1997; Geeslin, Linford, & Fafulas, 2015; Otheguy, Zentella, & Livert, 2007).

Contrary to expectations, the analysis did not reveal any significant main effects or interactions. The rate of SPP errors remained statistically the same regardless of whether narratives were written in the first or third person. Similarly, there were no discernible differences in SPP errors among the three beginner levels. These results demonstrate the significant impact of genre on how narrative voice influences the rate of SPP errors in learner texts. Unlike the findings of RQs 3 and 4, where both narrative voice and course level had a significant impact on SPP errors in descriptive texts, no such effects were observed in the narrative subcorpus. This discrepancy emphasizes the importance of considering genre when studying SPP accuracy. The lack of influence of narrative voice and course level on SPP errors in narratives suggests that different linguistic and cognitive processes may be at play in narrative writing compared to descriptive writing in this respect.

These findings contribute to the existing literature by highlighting the nuanced relationship between narrative voice, genre, and SPP errors. They also point to the need for future research to explore how linguistic and contextual factors interact to shape SPP patterns across different genres and writing tasks.

# Chapter 7: Conclusion

This dissertation explored the influence of prompts, genre, and narrative voice on lexical and syntactic characteristics in L2 Spanish learner writing, through two studies using the COWS-L2H corpus. Study 1 examined LD using MTLD, while Study 2 focused on the rate of SPP errors. The findings have important implications for research methods, corpus design, language teaching, and language testing, as well as highlighting limitations and suggesting future research directions. These aspects will be discussed in this chapter.

## 7.1 Implications for research methods

### 7.1.1 Methodological considerations

The findings from these two studies underscore the importance of considering prompts, topics, genre, and narrative voice when analyzing learner writing. The choice of prompt can significantly influence the linguistic features observed in written texts. Researchers need to make better design decisions regarding prompts and make sure to filter texts based on prompt characteristics. Additionally, when comparing different groups of texts, researchers must account for prompt and text differences to avoid making misleading statements about learners' proficiency levels based solely on surface-level features. For instance, assuming that narrative prompts about a beautiful story uniformly trigger similar MTLD scores can lead to erroneous interpretations. This is evidenced by the difference in MTLD scores that arises depending on whether the story is narrated in the first or third person.

### 7.1.2 Corpus design

LC play a crucial role in empirical research on SLA and language teaching. However, the variability introduced by different prompts and topics necessitates careful consideration in the design phase to ensure the corpus can support nuanced and accurate analyses. One critical aspect of corpus design is the ability to filter texts based on prompt characteristics. Different prompts can elicit varying linguistic features, which can impact the overall analysis if not properly accounted for. For instance, prompts designed to elicit descriptive texts might lead to different lexical and syntactic patterns compared to narrative prompts. Providing researchers with tools to filter and categorize texts based on these prompt characteristics can help isolate specific linguistic phenomena and yield more reliable findings.

Moreover, LC developers should consider incorporating prompts that cover a wide range of topics and writing tasks to capture the diverse linguistic abilities of language learners. A diverse LC allows researchers to examine how different contexts and communicative goals influence language use. For example, including prompts that require all learners to describe a person, detail a relationship, or narrate a past event can reveal how the same learners handle different grammatical structures and vocabulary across contexts.

To maximize the utility of LC, developers should include metadata about prompts and writing tasks, enabling researchers to select appropriate subsets of data for their specific research questions. Developers should also engage with the research community to gather feedback and ensure the corpus meets the evolving needs of SLA researchers.

## 7.2 Implications for language teaching and learning

Understanding how prompts, topics, genre and narrative person impact learner writing informs pedagogical practices. Educators can leverage insights from corpus-based research to design more effective writing prompts and tasks that target specific linguistic skills. By tailoring writing activities to address challenges identified through corpus analysis, instructors can better support language development. For example, prompts that naturally elicit a range of grammatical persons or encourage the use of more varied vocabulary can help learners practice and develop these features in meaningful contexts.

## 7.3 Implications for language testing

The findings from this dissertation have significant implications for language proficiency assessment and the automatic rating of texts. Automated and AI-driven scoring in L2 testing is increasingly prevalent, already in use for L2 English tests like TOEFL and the Duolingo English Test. It is plausible that L2 Spanish proficiency tests such as DELE and SIELE will adopt similar technologies in the coming decades. Automated scoring systems often rely on statistical measures of linguistic sophistication, such as the frequency and complexity of vocabulary and grammatical structures. However, the variability in language use observed across and within different prompts highlights the need for these systems to account for the context and communicative intent behind student responses. This variability suggests that current automated scoring mechanisms may not fully capture the nuanced language abilities of learners, potentially leading to inaccurate assessments of proficiency.

Language assessors and automated scoring systems should exercise caution when interpreting the presence or absence of specific linguistic features, recognizing that these features

may be influenced by the particular prompts and tasks presented to learners, as well as their own individual interpretation, rather than reflecting their overall language proficiency. For example, a prompt that elicits a narrative in the first person might produce different lexical and syntactic patterns compared to a descriptive task, even if completed by the same learner. As a result, scoring systems need to be sophisticated enough to consider these contextual factors to avoid misjudging a learner's true abilities. Incorporating a more holistic approach that evaluates language use within its communicative context can lead to more accurate and fair assessments, ultimately benefiting learners by providing a clearer picture of their language proficiency. This approach could involve integrating more advanced AI techniques that analyze not just the surface-level features of the text but also the underlying communicative purposes and situational variables.

### 7.4 Limitations and future research

These two studies have several limitations that warrant acknowledgment. Firstly, the focus on MTLD and SPP errors means that other important lexical and grammatical features were not examined. While MTLD provides insights into LD and SPP errors highlight specific syntactic challenges, this narrow focus limits our understanding of the broader spectrum of linguistic features, such as verb tense usage, morphological errors, and discourse markers, which are also critical in assessing L2 proficiency.

Secondly, the studies were conducted using a specific population—university students enrolled in Spanish courses at a public West Coast university. This demographic may not be representative of all L2 Spanish learners, potentially limiting the generalizability of the findings. Factors such as the learners' native languages, cultural backgrounds, prior exposure to Spanish,

and instructional contexts can significantly influence language acquisition and may differ markedly in other learner populations. Additionally, the exclusive use of the COWS-L2H corpus confined the results to the contexts and conditions under which this corpus was collected. Combining COWS-L2H data with a subset of the CEDEL2 corpus, which contains similar prompts, could have broadened the scope of the study and enhanced the generalizability of the results. However, such an approach also introduces potential challenges due to differences in populations and methodologies, potentially leading to less reliable significant effects.

Finally, the reliance on primarily statistical measures to interpret linguistic data has its constraints. Quantitative analyses provide valuable insights but should be complemented by qualitative approaches to fully capture the complexities of language use and development. Future research could benefit from a more comprehensive approach that includes multiple corpora and a wider array of linguistic features, thereby providing a more holistic understanding of L2 Spanish acquisition. Incorporating longitudinal studies tracking individual learner progress over time would offer deeper insights into the developmental trajectories of language learning, overcoming some of the static limitations of the current study design.

In conclusion, this dissertation underscores the importance of considering prompts, genre, and narrative voice in empirical research on learner writing. By accounting for these factors, researchers can conduct more robust analyses and provide deeper insights into the linguistic abilities and developmental trajectories of language learners. The methodological and pedagogical implications highlighted here can guide future research and practice in L2 Spanish writing, ultimately enhancing our understanding and support of language learners.

# References


Alonso-Ramos, M. (2016). *Spanish Learner Corpus Research: Current trends and future perspectives*. John Benjamins Publishing Company.

Bayley, R. & Pease-Álvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change, 9*(3), 349–371.

Bessett, R. M. (2017). Testing English influence on first person singular 'yo' subject pronoun expression in Sonoran Spanish. *Contemporary trends in Hispanic and Lusophone linguistics, 15*, 355–372.

Buyse, K., & González Melón, E. (2012). El corpus de aprendices Aprescrilov y su utilidad para la didáctica de ELE en la Bélgica multilingüe. *Multilingüismo y enseñanza de ELE en contextos multiculturales*, 247–261.

Castañeda-Jiménez, G., & Jarvis, S. (2014). Exploring lexical diversity in second language Spanish. In K. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 498–513). Wiley & Sons.

Cestero, A., & Penadés, I. (2009). *Corpus de textos escritos para el análisis de errores de* aprendices de E/LE (CORANE) [CD-ROM]. Universidad de Alcalá.

Cheng, Y.-S. (2008). Factors associated with foreign language writing anxiety. *Foreign Language Annals, 35*(5), 647–656. https://doi.org/10.1111/j.1944-9720.2002.tb01903.x

Cho, M. (2019). The effects of prompts on L2 writing performance and engagement. *Foreign Language Annals, 52*(3), 576–594. https://doi.org/10.1111/flan.12411

Conrad, M., Recio, G., & Jacobs, A. (2011). The time course of emotion effects in first and second language processing: A cross cultural ERP study with German–Spanish bilinguals. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00351

Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of Salsbury, Crossley & McNamara (2011). *Language Teaching*, *52*(3), 385–405. https://doi.org/10.1017/S0261444817000362

Daller, H. & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton, & J. Treffers‐Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 150–164). Cambridge University Press.

Davidson, S., Yamada, A., Fernández-Mira, P., Carando, A., Sánchez-Gutiérrez, C. H., and Sagae, K.. 2020, May. Developing NLP tools with a New Corpus of Learner Spanish. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 7238–7243).

Davies, M. 2016. *Corpus del Español: Two Billion Words, 21 countries*. Retrieved March 22, 2024, from http://www.corpusdelespanol.org

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59–84. https://doi.org/10.1017/S0272263104026130

Emmersion. (2021). *Understanding WebCAPE/CEFR Equivalencies.* Emmersionlearning. Retrieved March 22, 2024, from

http://support.emmersion.ai/support/solutions/articles/60000676566-webcape-cefr-equivalencies

Fernandez Florez, C. (2019). *Written variation among speakers of Spanish as their first, second and heritage language: The case of the subject pronoun expression.* [Unpublished doctoral dissertation]. University of Arizona.

Fernández-Mira, P., Morgan, E., Davidson, S., Yamada, A., Carando, A., Sagae, K., & Sánchez-Gutiérrez, C. H. (2021). Lexical diversity in an L2 Spanish learner corpus: The effect of topic-related variables. *International Journal of Learner Corpus Research*, *7*(2), 230-258. https://doi.org/10.1075/ijlcr.20017.fer

Frens, J. (2017). *lexical_diversity*. GitHub. Retrieved March 22, 2024, from https://github.com/jennafrens/lexical_diversity

Geeslin, K., Linford, B., Fafulas, S., Long, A., & Díaz-Campos, M. (2013). The group vs. the individual: subject expression in L2 Spanish. In J. Aaron, J. Cabrelli Amaro, G. Lord, & A. de Prada Pérez (Eds.), *Selected Proceedings of the 16th Hispanic Linguistics Symposium*. Cascadilla Proceedings.

Geeslin, K., Linford, B., & Fafulas, S. (2015). Variable subject expression in second language Spanish: Uncovering the developmental sequence and predictive linguistic factors. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective* (pp. 191-209). Georgetown University Press.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. *Studies in Corpus Linguistics, 33*, 13–32. https://doi.org/10.1075/scl.33.04gra

Granger, S., Gilquin, G., & Meunier, F. (2015). Introduction: Learner corpus research– past, present and future. In *The Cambridge Handbook of Learner Corpus Research* (pp. 1–5). Cambridge University Press.

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia – Social and Behavioral Sciences, 198*, 546–556. https://doi.org/10.1016/j.sbspro.2015.07.477

Hashemian, M., & Heidari, A. (2013). The relationship between L2 learners' motivation/attitude and success in L2 writing. Procedia – *Social and Behavioral Sciences, 70*, 476–489. https://doi.org/10.1016/j.sbspro.2013.01.085

Heaton, J. (1975). *Beginning composition through pictures*. Longman.

Heng, R., Pu, L., & Liu, X. (2023). The effects of genre on the lexical richness of argumentative and expository writing by Chinese EFL learners. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1082228

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.

Instituto Cervantes (2023). *El español: una lengua viva. Informe 2023.* Madrid: Instituto Cervantes. Retrieved June 11, 2024, from https://cvc.cervantes.es/lengua/anuario/anuario_23/informes_ic/p01.htm

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *Journal of Asia TEFL, 3*(4), 193–225.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, *19*(1), 57–84. https://doi.org/10.1191/0265532202lt220oa

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers, 53*, 61–79.

Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*(4), 554–564. https://doi.org/10.1016/j.system.2012.10.012

Kousta, S. T., Vinson, D.P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*(1), 48–60. https://doi.org/10.1016/j.jslw.2007.08.003

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143*(3), 1065–1081. https://doi.org/10.1037/a0035669

Kyle, K. (2020). Measuring lexical richness. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 454–476). Routledge.

Linford, B. (2016). *The second-language development of dialect-specific morpho-syntactic variation in Spanish during study abroad,* ProQuest Dissertations and Theses.

Lozano, C. (2009). CEDEL2: Corpus Escrito del Español como L2. In Bretones José, C. M., Fernández Sánchez, F., Ibáñez Ibáñez, J. R., García Sánchez, M. E., Cortés de los Ríos, M. E., Salaberri Ramiro, S., Cruz Martínez, M. S., Perdú Honeyman, N., & Cantizano Márquez, B. (Eds.), *Applied Linguistics Now: Understanding Language and Mind/La*

*Lingüística Aplicada Actual: Comprendiendo el Lenguaje y la Mente* (pp. 197–212). Universidad de Almería.

Martinez, G. (2007). Writing back and forth: The interplay of form and situation in heritage language composition. *Language Teaching Research, 11*(1), 31-41. https://doi.org/10.1177/1362168806072454

Mavrou, I. (2016). Complejidad, precisión, fluidez y léxico: Una revisión. *Moderna Språk, 110*(1), 49-69.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* [Doctoral dissertation]. Retrieved from ProQuest Dissertations & Theses A&I. (305349212).

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/brm.42.2.381

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge University Press.

Miguel, N. M., & Bonilla, C. L. (2022). Utilización de corpus de español L1 en el aula de LE/L2/LH: ¿por qué y para qué? *Journal of Spanish Language Teaching*, *9*(2), 148–160. https://doi.org/10.1080/23247797.2022.2157083

Milton, J. (2009). *Measuring second language vocabulary acquisition (Vol. 45)*. Multilingual Matters.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. Routledge.

Otheguy, R., Zentella A., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language, 83*(4). 770-802. https://doi.org/10.1353/lan.2008.0019

Padró, L. (2011). Analizadores Multilingües en FreeLing. *Linguamática, 3*(2), 13–20. https://doi.org/10.21814/lm.3.2.115

Parodi, G., Cantos-Gómez P., & Howe, C. (2022). *Lingüística de corpus en español / The Routledge Handbook of Spanish Corpus Linguistics*. Routledge.

Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners* [Unpublished doctoral dissertation]. University of Hawaii.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*(6), 405–428. https://doi.org/10.1080/00207594.2012.743665

Rahimi, M., & Zhang, L. J. (2019). Writing task complexity, students' motivational beliefs, anxiety and their writing production in English as a second language. *Reading and Writing*, *32*(3), 761–786. https://doi.org/10.1007/s11145-018-9887-9

Real Academia Española y Asociación de Academias de la Lengua Española. (n.d.). *Diccionario panhispánico de dudas (DPD)*, 2.ª edición (versión provisional). Retrieved March 22, 2024, from https://www.rae.es/dpd/

Real Academia Española y Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española.* Espasa.

Real Academia Española. Banco de datos (CORPES XXI) *Corpus del Español del Siglo XXI*. Retrieved March 22, 2024, from https://www.rae.es

Real Academia Española. Banco de datos (CREA). *Corpus de referencia del español actual*. Retrieved April 11, 2024, from https://www.rae.es

Real Academia Española. Banco de datos (CORDE). *Corpus diacrónico del español actual*. Retrieved March 22, 2024, from https://www.rae.es

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Rojo, G., & Palacios, M. I. M. (2016). Learner Spanish on computer: The CAES Corpus de Aprendices de Español project. In *Spanish learner corpus research* (pp. 55–87). John Benjamins.

Rojo, G., Palacios, I., Sampedro Mella, M., & Marisily, A. (2022) Los corpus de aprendices de español LE/L2: panorama actual y perspectivas futuras. J*ournal of Spanish Language Teaching, 9*(2), 174-189. https://doi.org/10.1080/23247797.2022.2157085

Sadeghi, K., & Dilmaghani, S. (2013). The relationship between lexical diversity and genre in Iranian EFL Learners' writings. *Journal of Language Teaching and Research, 4*(2), 328–334. https://doi.org/10.4304/jltr.4.2.328-334

Saito, K., Webb, S., Trofimovich, S. P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. S*tudies in Second Language Acquisition 38*(4), 677–701. https://doi.org/10.1017/s0272263115000297

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research, 27*(3), 343–360. https://doi.org/10.1177/0267658310395851

Sánchez-Gutiérrez, C., De Cock, B., & Tracy-Ventura, N. (2022) Spanish corpora and their pedagogical uses: challenges and opportunities. *Journal of Spanish Language Teaching*, *9*(2), 105-115. https://doi.org/10.1080/23247797.2022.2157078

Sánchez-Gutiérrez, C., & Fernández-Mira, P. (2023). Writing about positive and negative topics in Spanish as a heritage or second language. In A. Blanco Canales & S. Martín Leralta (Eds.), *Emotion and identity in second language learning.* Peter Lang.

Sánchez-Gutiérrez, C., Hoyos-Álvarez, C., Guerra, K., & Fernández-Mira, P. (2022). Emotional processing of Spanish words in L2 learners and heritage speakers: Interactions between cognateness and emotional valence. In I. Mavrou, M. Pérez Serrano, & J. M. Dewaele (Eds.), *Recent advances in second language emotion research*. Civitas Aranzadi Thomson Reuters.

Sasayama, S., Garcia Gomez, P., & Norris, J. M. (2021). Designing Efficient L2 Writing Assessment Tasks for Low‐Proficiency Learners of English. *ETS Research Report Series*, *2021*, 1–31. https://doi.org/10.1002/ets2.12341

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning, 53*(1), 165–202. https://doi.org/10.1111/1467‐9922.00213

Sheikh, N. A., & Titone, D. (2016). The embodiment of emotional words in a second language: An eye-movement study. *Cognition and Emotion, 30*(3), 488–500. https://doi.org/10.1080/02699931.2015.1018144

Tack, A., Francois, T., Roekhaut, S., & Fairon, C. (2017). Human and automated CEFR-based grading of short answers. In J. Tetreault, J. Burstein, C. Leacock, & H. Yannakoudaki

(Eds.), *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 169–179). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-5018

Tomé Cornejo, C. & Recio Diego, Á. (2022). De la disponibilidad léxica a la disponibilidad gramatical. Desarrollos metodológicos y aplicación didáctica. *Tejuelo*, *35*(3), 299–336. https://doi.org/10.17398/1988-8430.35.3.299

Torres Cacoullos, R. & Travis, C. E. (2015) Assessing inter-linguistic (dis) similarity via intra-linguistic variability for subject expression. In Carvalho, A. M., Orozco, R., & Shin, N. L. (Eds.). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Georgetown University Press.

Tracy-Ventura, N., Mitchell, R., & McManus, K. (2016). The LANGSNAP longitudinal learner corpus: Design and use. In M. Alonso Ramos (Ed.), *Spanish learner corpus research: State of the art* (pp. 117–142). John Benjamins.

Tracy-Ventura, N. & Paquot, M. (2021). *The Routledge Handbook of Second Language Acquisition and Corpora*. Routledge.

Yamada, A., Davidson, S., Fernández-Mira, P., Carando, A., Sagae, K., & Sánchez-Gutiérrez, C. (2020). COWS-L2H: A corpus for measuring learner Spanish writing development. *Research in Corpus Linguistics, 8*(1), 17–32. https://doi.org/10.32714/ricl.08.01.02

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236–259. https://doi.org/10.1093/applin/amp024

Yule, G. (1997). *Referential communication tasks*. Erlbaum.

Zhan, J., Sun, Q., & Zhang, L.J. (2021). Effects of manipulating writing task complexity on learners' performance in completing vocabulary and syntactic tasks. *Language Teaching Research*. https://doi.org/10.1177/13621688211024360

# Appendix I

Translations to English from sample texts in Table 19

*Yourself*:

I am Chinese. I have black and long hair. I am from Southern California. I like to eat salad, chicken, and vegetables. University is difficult. I study and work a lot. I am very stressed. I want to stay in shape, but I don't have time. I like to go for walks in the morning and play tennis. I am a bit athletic and love to exercise. I have a mother and an older brother, and I love my family very much. My friends say that I am funny and nice. I take a shower every night. I like to sleep but don't have time. I am very tired and sad. I want to be more excited and balanced. I don't think I'll pass one of my classes. I want time to study. Life is very difficult. I bought a car yesterday and the gas is expensive. The car is very old. I am not a good driver but I practice in the park. I talk to my mom every day on the phone. She motivates me. I am never at home. I am always at the library. I need a break from school. I dream of being happy. After studying, I go to work. I work until midnight. I work thirty hours a week. Work is fun but I am tired. Homework programming takes me 18 hours a week. I wear a blue shirt to work. The rent is expensive and I can't ask my family for help. It's okay. I dream of a better life after college.

*Famous*:

Carli Lloyd is a female soccer player for the United States national team. She plays defense and offense and is a very talented player. She scores many goals and is a leader on the team. I am also a soccer player and have looked up to her since I was little. Carli is an amazing player for her performance in crucial moments. She is also a team player and often helps with many goals.

Carli has won many different awards for playing soccer. She is an Olympian, a World Cup winner, and has won the MVP award many times. Carli is very decorated, but she remains humble and works hard to become a better soccer player. Carli wasn't always the best player; she worked hard and put in a lot of time to become the skilled player she is today. I have always liked the way Carli strikes the ball with authority and power. She is a talented player, but due to how hard she works, she has become great. Carli is from New Jersey and grew up playing soccer since she was little. She was a very active child and loved playing soccer. She was very competitive and became a tough player playing with boys. She attended college in New Jersey and became the leading scorer during her freshman and sophomore years. It was her ball skills, as well as her ability to distribute the ball, that set her apart from other soccer players her age.

*Special*:

A special person in my life is my friend, *FIRST_NAME*. She lives in *CITY* and studies art at *UNIVERSITY*. She is vegan and is one reason why I am vegan (and why my twin is vegan). *FIRST_NAME* and I went to the same school in fourth grade (more or less). After graduation, *FIRST_NAME*, my sister, our friend *FIRST_NAME*, and I went to *CITY* for our "senior trip". We visited many museums, many restaurants, and many bookstores. We walked around *CITY* for six days. We used the subway or borrowed *FIRST_NAME*'s aunt's car. Every morning we woke up early and her aunt drove us to the train station (the aunt lives in *STATE*, near *CITY* because she is a librarian in the city). *FIRST_NAME* made an itinerary with museums and events, and *FIRST_NAME* made a list of vegan restaurants in the city. During the day, we walked (or took the train), followed the itinerary, had breakfast and lunch when we were hungry. Every night, we took the train to *FIRST_NAME*'s aunt's place, or she met us in the city and drove us to her home.

We had dinner (vegan, of course) and read the books we bought, or we watched the new Game of Thrones (this was in 2016). *FIRST_NAME* cooks very well, and I love cooking with her. We took the bus from *CITY*, *STATE* to *CITY* (and back), for six or more hours each time. Every summer, more or less, I visit *FIRST_NAME* in *STATE* and we cook vegan food and watch movies. She is one of my best friends.