UCLouvain
Faculté de philosophie, arts et lettres – FIAL

Faculté de philosophie, arts et lettres

Diachronic Evolution of Subordinate Clause Use by French-Speaking Learners of English

A Comparative Learner Corpus Study Based on ICLE and ICLE+30

Auteur : Isaline Barthélemy

Promoteur(s) : Prof. Dr. Gaëtanelle Gilquin

Année académique 2024-2025

Master 60 en langues et lettres modernes, orientation germaniques

Declaration of honour

Dissertation/Individual Final Project

Faculty of Philosophy, Arts and Letters, UCLouvain

I hereby declare that this is an original and personal work, that all referenced sources have been fully cited, regardless of their origin, and that the use of artificial intelligence tools complies with the general instructions published by UCLouvain and available at the following address :

https://cdn.uclouvain.be/groups/cms-editors-p1/portail/reglement/Consignes-ChatGPT-version-etudiante.pdf?itok=SebXXm87 (French version)

I am aware that failure to cite a source, to cite it clearly and fully, to announce in a specific section at the beginning of the dissertation the use I have made of artificial intelligence tools, or to use them in a manner contrary to the guidelines set forth in the institutional note, constitutes serious irregularities in accordance with the Academic regulations and procedures (Chapter 4, section 7, articles 107 to 114) within the University. In particular, I am aware of the risks of academic and disciplinary sanctions.

In light of the above, I declare on my honour that I have not committed plagiarism or any other form of fraud and that I have not used artificial intelligence tools beyond what is permitted by UCLouvain.

Last name, First name : Barthélemy Isaline

Date : 1st August 2025

Student's signature :

Acknowledgments

I would like to express my deepest gratitude to my supervisor for her invaluable guidance, insightful feedback, and the time she generously dedicated to reviewing my work in such detail. Her expertise and encouragement played an important role in the completion of this thesis.

I am also deeply thankful to my family for their unwavering support throughout my academic journey. A special thanks goes to my mother, whose constant encouragement and belief in me have been a source of strength and motivation.

My heartfelt thanks also go to my best friend, Alice, for her thoughtful advice and comforting presence during the more challenging moments of this journey.

Finally, I would like to extend my appreciation to all those who, in one way or another, contributed to this work, whether through a kind word, a helpful suggestion, or simply by being there when I needed it most.

Table of contents

List of tables ..... **i**

List of figures ..... **ii**

List of abbreviations ..... **iii**

1. Introduction **1**

1.1. Rationale **1**

1.2. Research questions and hypotheses **2**

1.3. Content **3**

2. Literature review **5**

2.1. Diachronic research **5**

2.2. Subordinate clauses **6**

2.2.1. Definition and classification 6

2.2.2. Diachronic evolution of subordinate clauses 9

2.2.3. Subclauses in learner English 10

3. Methodology **18**

3.1. Corpus description **18**

3.2. Sampling procedure **18**

3.3. Data coding **19**

3.3.1. Coding criteria 20

3.3.2. Data Analysis 25

4. Results **27**

4.1. Qualitative analysis **27**

4.1.1. Infinitive subordinate clauses in ICLE-FR and ICLE+30-FR 27

4.1.2. Gerund subordinate clauses in ICLE-FR and ICLE+30-FR 30

4.2 Quantitative analysis **33**

4.2.1. Frequency of subordinate clause use 33

4.2.2. Types of subordinate clauses 34

4.2.3. Error patterns 40

5. Discussion **45**

5.1. Increase in Subordinate Clause Frequency **45**

5.2. Shifts in Types of Subordinate Clauses Used **46**

5.3. Evolution of Error Patterns in Subordinate Clauses 50

5.4. Limitations 56

6. Conclusion 58

7. Bibliography 60

List of tables

Table 1: Frequency of subordinate clauses in Lambert et al.'s (2003) study 13
Table 2: Coding Excel file sample 21
Table 3: Distribution of subordinate clause types in the ICLE-FR and ICLE+30-FR 35
Table 4: Frequency of subordinate clause subtypes in the ICLE-FR and ICLE+30-FR 37
Table 5: Number of errors and error rates per subordinate clause subtype detected in ICLE-FR and ICLE+30-FR 42

List of figures

Figure 1: Distribution of subordinate clause subtypes in ICLE-FR and ICLE+30-FR 38

List of abbreviations

Abbreviation	Definition
CEFR	Common European Framework of Reference for Languages
EFL	English as a Foreign Language
ESL	English as a Second Language
ICLE	International Corpus of Learner English
ICLE+30	International Corpus of Learner English (30 years later)
IL	Interlanguage
L1	First language
L2	Second language
SLA	Second Language Acquisition

1. Introduction

1.1. Rationale

Research into second language acquisition (SLA) seeks to uncover how learners acquire, develop, and refine their use of a target language over time (AL-Takhayinh et al., 2015). Subordinate clauses [1], as complex syntactic structures, play an important role in the development of linguistic proficiency, as they enable speakers and writers to produce nuanced, cohesive, and sophisticated language (Berman and Verhoeven, 2002). Despite their importance, mastering subordinate clauses poses significant challenges for second language learners.

For French-speaking learners of English, the transfer effects from French to English are particularly pronounced in the domain of subordinate clauses. As explained by Roth (2010), while French and English share certain syntactic features due to their shared Indo-European origins, there are notable differences in clause formation, conjunction use, and the syntactic constraints governing subordination. Understanding the syntactic challenges faced by French-speaking learners of English due to such differences and how these challenges evolve over time can provide valuable insights into the processes underlying SLA.

To clarify this study's specific contribution, it is useful to outline its main objectives:

First of all, this study aims to address a gap of Diachronic SLA Research[2]. While previous research has explored various aspects of learner English, such as lexical diversity and grammatical accuracy, fewer studies have examined how the use of subordinate clauses evolves over time. Most existing research (Chen et al., 2020; Kalnbērziņa & Rūtenberga, 2012; etc.) focuses on synchronic comparisons or broader syntactic features, leaving a gap in our understanding of how specific structures like subordinate clauses develop diachronically. This study seeks to fill that gap by analyzing data from the ICLE and ICLE+30, enabling a diachronic analysis of learner output across decades.

[1] And particularly relative clauses, which are longer (Berman and Verhoeven, 2002).
[2] Diachronic learner corpus research investigate "the evolution of learner language (as produced by different generations of learners from the same proficiency level) over the course of time, to determine whether this variety of language is different now from what it was twenty or thirty years ago, for example." (Gilquin, 2022:2)

Furthermore, one of the major motivations for this study lies in its potential to shed light on the long-term development of interlanguage among French-speaking learners of English. While many studies focus on short-term acquisition or cross-sectional comparisons at a specific point in time, diachronic learner corpus research makes it possible to observe how interlanguage evolves collectively over several decades. This type of analysis helps to document not only shifts in the frequency and diversity of these grammatical structures, but also changes in the types of errors produced by learners. By focusing specifically on subordinate clauses, which is a syntactic feature known to be acquired late in the process of SLA, this study offers a window into the gradual restructuring of learners' interlanguage over time. Such research provides valuable insights into how learners progressively master complex syntactic constructions, and it highlights both the persistent challenges and the areas of improvement that emerge as learners gain more exposure to English and benefit from evolving teaching practices. In doing so, this study contributes to a deeper understanding of the mechanisms of long-term interlanguage development.

Finally, error analysis has long been a cornerstone of SLA research, providing insights into the interlanguage processes that underlie learner performance. Identifying persistent errors in the use of subordinate clauses can inform language teachers and curriculum designers about specific areas where learners struggle, enabling them to develop targeted instructional strategies. Furthermore, understanding diachronic changes in error patterns can shed light on the effectiveness of previous teaching approaches and highlight areas for improvement.

1.2. Research questions and hypotheses

The ability to use subordinate clauses effectively is a key marker of syntactic complexity and proficiency in SLA. French-speaking learners of English often face challenges in mastering this aspect of English grammar due to differences between their first language (L1) and the target language (L2). By analyzing diachronic changes in the use of subordinate clauses, this study seeks to uncover how the linguistic performance of French- speaking learners has evolved over time.

The primary research question guiding this study is: What diachronic changes can be observed in the use of subordinate clauses by French-speaking learners of English over time, based on a comparative analysis of the ICLE-FR and ICLE+30-FR corpora?

To answer this question, the study will address the following sub-questions:

1. How has the frequency of subordinate clause usage changed between the ICLE-FR and the ICLE+30-FR?
2. What changes can be observed in the types of subordinate clauses used by upper-intermediate and advanced French-speaking learners of English in the ICLE-FR and the ICLE+30-FR?
3. What are the most common error patterns in subordinate clause usage in the ICLE-FR and the ICLE+30-FR, and how have they evolved diachronically?

Based on the theoretical framework and previous empirical findings that will be presented in section 2, the following hypotheses were formulated in response to the research questions guiding this study:

First, it is hypothesized that the overall frequency of subordinate clause use has increased in learner productions over time. Specifically, learners in the ICLE+30-FR are expected to produce a higher number of subordinate clauses per sentence compared to those in the original ICLE-FR, reflecting a diachronic trend towards greater syntactic complexity and a more advanced use of subordination in academic writing.

Then, we also make the hypothesis that there has been a shift in the distribution of subordinate clause types across the two corpora. In particular, learners in the ICLE+30-FR are expected to exhibit a wider range of subordinate clause types, with increased use of adverbial clauses and non-finite structures (such as -ing and infinitive clauses), indicating a gradual diversification and refinement of their syntactic repertoire over time.

Finally, it is hypothesized that the frequency of grammatical errors in subordinate clauses has decreased over time. More precisely, learners in the ICLE+30-FR are expected to demonstrate fewer errors in their use of subordinate clauses compared to learners in the ICLE-FR, reflecting improved grammatical control and potentially the impact of pedagogical and contextual changes (the evolution of new technologies, for example).

1.3. Content

This thesis explores the diachronic development of syntactic complexity in the academic writing of French-speaking learners of English in Belgium, with a specific focus on the use of subordinate clauses. The study adopts a learner corpus-based approach, comparing two subcorpora collected approximately thirty years apart (ICLE-FR and ICLE+30-FR) in order to identify long-term trends in the frequency, distribution, and accuracy of subordinate clause

usage. By focusing on the evolution of these complex structures, the research provides insight into the gradual development of learners' interlanguage over an extended time span.

The thesis is divided into several parts. The first section presents the theoretical framework, including key concepts related to syntactic complexity, subordinate clause typologies, and SLA theories, with a particular emphasis on the role of interlanguage development in shaping learners' grammatical competence. This section also reviews existing research on subordinate clauses in learner language, highlighting the lack of diachronic studies in this area.

The second section details the methodology, including the selection and compilation of the corpora, the coding procedures for subordinate clause identification, and the statistical methods used for data analysis. The methodological choices are justified in light of the study's diachronic focus and its aim to ensure comparability between the two corpora.

The third section presents the empirical analysis and results. It examines three key aspects: the overall frequency of subordinate clauses, the distribution of subordinate clause types and subtypes, and the patterns of grammatical errors related to subordination. Each of these aspects is analyzed through a diachronic lens to identify changes over time.

The fourth section is dedicated to the discussion. It interprets the results in light of second language acquisition theories and the notion of interlanguage development, focusing on the implications of the observed diachronic changes in syntactic complexity and error patterns.

Finally, the conclusion summarizes the key findings, reflects on the study's contribution to diachronic learner corpus research, and suggests avenues for future research.

2. Literature review

2.1. Diachronic research

Diachronic research refers to the study of language change over time, based on the systematic observation of linguistic data drawn from different historical periods. It investigates how various aspects of language evolve, revealing not only structural shifts but also broader patterns of linguistic development (Yamamoto et al., 2012). In this context, a distinction is commonly made between long-term diachronic studies, which span centuries, and short-term diachronic studies, which explore more recent changes, typically over a few decades (Gilquin, 2022)

In the specific domain of learner corpus research, diachronic perspectives are beginning to emerge as a valuable means of tracking the evolution of learner language over time. As Gilquin (2022) points out, since the first learner corpora were compiled in the 1990s, the field now has the historical depth needed to examine the short-term evolution of learner language using corpus methods. This approach, referred to as diachronic learner corpus research, differs from longitudinal studies that follow the same learners over time. Instead, it compares productions from different generations of learners at the same proficiency level, with the aim of assessing whether and how learner language has changed, for instance, over the last twenty or thirty years. This approach has proven both possible and relevant, provided that comparable corpus data are available (Gilquin, 2022). It complements earlier work by scholars such as Biber and Gray (2016), AL Fajri and Okwar (2020), and Malà (2017), who have examined syntactic evolution and simplification trends in both native and learner varieties of English.

While much diachronic research has traditionally focused on native-speaker varieties, especially in historical linguistics, there is growing interest in exploring diachronic change in learner language. In this emerging area, researchers examine how the linguistic productions of language learners evolve over time. This can involve studying changes in learners' grammatical accuracy, syntactic complexity, lexical range, or pragmatic competence, often in relation to shifts in language teaching practices, exposure to the target language, or broader sociocultural factors. In such studies, corpora play a central role, as they provide access to large samples of learner language collected across different time periods.

This thesis contributes to this field by adopting a short-term diachronic perspective on learner English. It uses two learner corpora compiled roughly thirty years apart (ICLE-FR and

ICLE+30-FR) to investigate whether the academic writing of French-speaking learners of English has evolved over time. Specifically, it focuses on changes in the use of subordinate clauses, a complex syntactic feature, to explore how learners' syntactic practices and error patterns may have shifted over recent decades.

Linguistic changes and diachronic variations lie at the heart of studies on language evolution and play a fundamental role in understanding linguistic transformations over time. These changes are observable across all aspects of a language, but syntactic developments, such as those affecting subordinate clauses, are particularly striking.

2.2. Subordinate clauses

2.2.1. Definition and classification

Subordinate clauses, or dependent clauses, are linguistic units that function within a larger sentence structure, forming what is known as a complex sentence. Unlike independent clauses, subordinate clauses cannot stand alone as complete sentences because they rely on a main clause to provide meaning. Subordination establishes a hierarchical relationship, where the subordinate clause serves a specific grammatical role within the sentence (Quirk et al., 1972).

Subordination is distinct from coordination, as it involves a non-symmetrical relationship: one clause is a constituent of the other, whereas coordination places clauses on an equal hierarchical level. On the one hand, the choice to use a subordinate clause often reflects a speaker's intent to organize information, emphasizing certain elements over others, as explained by Smaby (1974). Main clauses, on the other hand, typically express the principal idea, while subordinate clauses provide supporting details.

Subordinate clauses are generally introduced by subordinating conjunctions, though other markers such as wh-elements, that-elements, inversions, and verbless constructions can also signal subordination, as explained by Quirk and Greenbaum (1973). The authors categorize subordinating conjunctions into three types

- Simple subordinators: after, because, if, when, etc.
- Compound subordinators: so that, as soon as, rather than, etc.
- Correlative subordinators: if…then, so…that, whether…or, etc.

For example, the complex sentence in (1) illustrate this phenomenon in which a main clause superordinating a subordinate clause which itself superordinate another subordinate clause:

(1) I think [that you can do it {if you try}]

Subordinate clauses can be classified into three structural types based on their verb elements (Quirk et al., 1972; Quirk and Greenbaum, 1973): finite clauses, non-finite clauses, and verbless clauses. Finite clauses contain a finite verb phrase as V-element and a subject. Non-Finite clauses include clauses with a non-finite verb phrase (to-infinitives, bare infinitives, -ing participles or -ed participles). The following examples, given by Quirk and Greenbaum, respectively illustrate each category of non-finite verb phrases (Quirk and Greenbaum, 1973: 311).

(2) The best thing would be to tell everybody.
(3) All I did was hit him on the head.
(4) Leaving the room, he tripped over the mat.
(5) Covered with confusion, I left the room.

Finally, verbless clauses lack a verb, often with an implied "to be".

All these structures may function either as main clauses (with or without embedded subordinate clauses) or as subordinate clauses. Earlier in this section, a threefold classification of subordinate clauses was introduced based on their syntactic form, namely finite clauses, non-finite clauses, and verbless clauses. This typology refers to the grammatical structure of the clause, particularly whether it includes an overtly marked verb form and what kind.

In addition, Quirk and Greenbaum (1973) propose a complementary classification based on the function of subordinate clauses within the sentence. According to them, subordinate clauses fall into five major functional categories: nominal, adverbial, relative, comparative and comment clauses. First, nominal clauses function as nouns (as noun phrases do) and include that-clauses, wh-clauses, indirect questions, nominal relative clauses, to-infinitive clauses, bare infinitives, and -ing clauses, as respectively shown in examples (6)-(12) given in Quirk & Greenbaum's (1973) grammar:

(6) That she is still alive is a consolation.
(7) How the book will sell depends on its author.
(8) Do you know if/whether the banks are open.
(9) What he is looking for is a wife.
(10) My wish is to be a pilot.
(11) All I did was (to) turn off the gas.
(12) Telling lies is wrong.

Then, adverbial clauses can be either adjunct or disjunct[3] and express semantic relations such as time, place, condition, reason, or manner, etc. as illustrated in the following examples from the same grammar book:

(13) When I last saw you, you lived in Washington.

(14) They went wherever they could find work.

(15) If you treat her kindly, (then) she'll do anything for you.

(16) I lent him the money because he needed it.

(17) He treated me (just) as if he has never met me.

Comparative clauses express comparison, often following structures like "as...as" or "more/less than", as in (18):

(18) Jane is as healthy as her sister.

The authors also describe comment (19) and reporting clauses (20) (sometimes considered a subset of adverbials or nominals) which provide evaluation or introduce reported speech:

(19) At that time, I believe, labour was cheap.

(20) 'I am your friend', John said.

Finally, relative clauses postmodifies nouns being introduced by relative pronouns (such as *who, which, that*, etc.) and can be either restrictive (essential to understand the sentence) or non-restrictive (giving non-essential information).

(21) The girl who stood in the corner.

(22) The group which was responsible for this decision…

(23) The table that stands in the corner…

While the form-based classification focuses on how subordinate clauses are constructed grammatically, the function-based classification considers what role they play within a sentence. These two approaches are not mutually exclusive but rather offer complementary perspectives on subordinate clause usage.

As mentioned by Quirk et al. (1972) and Quirk and Greenbaum (1973), subordinate clauses fulfill diverse syntactic functions, including: subjects as in (24); objects as in (25) for direct objects and in (26) for indirect objects functions; complements as in (27) for subject

[3] "Adjuncts are adverbial element which are (apparently) subordinated to a verb phrase or part of a verb phrase (such as the main-verb-phrase), while disjuncts are adverbial elements which are (apparently) subordinated to the entire main clause." (Dickins, 2006: 27)

complements and in (28) for object complements; adverbials as in (29) for adjunct, in (30) for disjunct and in (31) for conjunct adverbials; postmodifiers in noun phrases as in (32); prepositional complement as in (33); and adjectival complementation as in (34).

(24) That we need more equipment is obvious
(25) I know that she is pretty
(26) I gave whoever it was a cup of tea
(27) The point is that we're leaving
(28) I imagined him overcome with grief
(29) When we meet, I shall explain everything
(30) To be honest, I've never liked him
(31) What is more, he has lost the friends he has
(32) A friend who remains loyal.
(33) It depends on what we decide
(34) Ready to act promptly

2.2.2. Diachronic evolution of subordinate clauses

Beyond the mere presence or absence of subordinate clauses, the diachronic evolution of syntactic structures also reflects growing levels of syntactic complexity, defined as the variety and elaboration of grammatical constructions employed in written language (Ortega, 2015). This complexity is not only indicative of linguistic maturity, but also closely associated with communicative intent, and the acquisition context. In second language (L2) writing, numerous longitudinal studies have shown that syntactic complexity increases over time, especially through the use of subordinate clauses, noun phrase elaboration, and varied embedding strategies (Crossley & McNamara, 2014; Ortega, 2003). A key indicator of this development has traditionally been the T-unit, which comprises a main clause and any subordinate clause attached to it. While T-unit-based indices remain widely used, they are now considered limited, especially in academic registers where complexity often manifests more through phrasal constructions than clausal subordination (Biber et al., 2011, cited in Crossley & McNamara, 2014). Research by Crossley and McNamara (2014) highlights that over time, L2 writers tend to produce texts that are syntactically more aligned with academic writing, favoring dense noun phrases and embedded modifiers, though raters often still judge proficiency based on spoken-like clausal structures. This disconnect between syntactic development and perceived quality underscores the multidimensionality of complexity: surface features (such as clause frequency)

may not fully capture deeper structural elaboration. Furthermore, Ortega (2015) emphasizes that syntactic complexity should be viewed not just as a structural feature, but as a powerful interaction between developmental stage, instructional context, and even first-language influence. For instance, learners from L1 backgrounds favoring subordination (e.g., Spanish) may produce more complex subordinate clause structures earlier in L2 writing, suggesting that L1 transfer plays a role in shaping diachronic patterns of complexity.

Research on syntactic evolution shows that complex structures, such as subordinate clauses, have undergone significant transformations. The study by AL Fajri and Okwar (2020) on "the change in the use of English relative clauses over a 45-year time span" (1) in British and American English written texts reveals that subordinate clauses have simplified over time. The gradual abandonment of forms such as relative clauses with *whom* or *which* in favor of *that* illustrates, according to the authors, a process of colloquialization and Americanization. This trend reflects a shift toward a more direct and conversational style, a characteristic also influenced by the widespread diffusion of American English, which often drives such changes.

Another diachronic study, conducted by Malà (2017) on "a change in the use of non-finite clauses in written English" over the past 30 years, underscores a notable increase in the use of non-finite clauses in written texts (British and American). Vachek (1961, cited in Malà, 2017) explains this shift as a historical development in the English language, through which finite verbs have "lost their dynamism." This gradual process has led to a more pronounced use of infinitives, gerunds, and participles, now considered "semantically more flexible."

The research by AL Fajri and Okwar, as well as that by Malà, is particularly relevant to our study. These works primarily focus on the evolution of the language in recent years (short-term diachrony), in contrast to most diachronic studies, which evaluate language changes over a much larger timescale (long-term diachrony spanning several centuries). While the mentioned studies do not cover all types of subordinate clauses analyzed in Chapter 5, they do address some of the most widespread ones in English. It is plausible that other subordinate clauses may also have been affected by the processes of colloquialization, Americanization, and semantic shifts observed in recent years. Furthermore, although these studies focus solely on changes in native speakers' language, it is likely that the language of learners has also been influenced.

2.2.3. Subclauses in learner English

In their article on the analysis of subordinate clauses in a learner corpus, Kalnbērziņa and Rūtenberga (2012) draw on Pienemann's processability theory, which posits that language

learners can only produce and comprehend linguistic structures for which they possess the necessary processing resources. In other words, the ability to use a specific grammatical structure depends on the availability of the learner's cognitive resources at a given stage of their development. Pienemann's theory proposes that the acquisition of syntactic structures follows an implicational hierarchy. This means that each step is a prerequisite for the next, and acquisition unfolds in a specific sequence. This sequence reflects the language generation process. The stages through which learners progress when acquiring syntax, particularly in the context of learning English as a second language, are described as follows:

1. Acquisition of words
2. Acquisition of processes associated with each word category
3. Construction of sentences based on these categories
4. Development of sentences and the addition of morphology at the sentence level (e.g., subject-verb agreement)
5. Construction of subordinate clauses and the appropriate use of relationships between main and subordinate clauses

According to the descriptors of the Common European Framework of Reference for Languages (CEFR), a learner is able to understand and use complex sentences (including subordinate clauses) starting at level B1 (Council of Europe, 2001). The study by Kalnbērziņa and Rūtenberga (2012) demonstrates an increase in the number of subordinate clauses as learners' proficiency improve. This allows us to assert that such sentence types can serve as an indicators of the linguistic proficiency levels of language learners.

This idea is also shared by Ortega (2015), who adds that beginner learners tend to use simpler subordinate clauses, such as nominal and adverbial clauses, while more advanced learners employ more complex subordinate clauses, such as relative clauses. The author also highlights differences in sentence types depending on the text genre. Specifically, argumentative texts provide greater opportunities for the use of subordinate clauses compared to narrative texts, which tend to feature sentence structures described as simpler.

2.2.3.1. Differences between French and English subordinate clause use

It is widely recognized that French and English are relatively similar languages, despite belonging to different language families (English being a Germanic language and French a Romance language). Roth (2010) explains that this similarity largely stems from the syntactic,

morphological, and semantic influence of French on English throughout history. Indeed, the arrival of the Anglo-Normans in England during the Norman Conquest of 1066 marked a central moment in the evolution of the English language. Syntax was remarkably impacted, including the adoption of certain French adjectives with post-adjectival positioning.

Although English and French are relatively close linguistically and historically (Lambrecht, 2010), subordinate clauses in English and French differ not only in terms of frequency but also in their functions, their role in narration, and their syntactic structuring, as demonstrated by the findings of the study conducted by Lambert et al. (2003). These distinctions reflect linguistic and pragmatic choices specific to each language, influencing both narrative construction and the way second-language learners incorporate these structures.

This section will be based mainly on Lambert et al.'s study (2003) which is particularly relevant to this thesis. They collected oral data by asking participants to produce narratives based on a wordless picture story. The dataset included three groups: native speakers of French, native speakers of English, and French-speaking learners of English. This elicitation method ensured that all participants described the same content, allowing for cross-linguistic and interlanguage comparisons in narrative structure and syntactic choices. The learners involved were advanced university-level French-speaking students of English.

In general, French speakers tend to use subordination more frequently than English speakers. The data in Table 1, derived from the study by Lambert et al., show a subordination rate of 37% in French compared to 19.1% in English. The authors explain this disparity as stemming from a fundamental difference in sentence organization: French favors hierarchical structures where subordinate clauses allow for nuance or detail, while English often prefers simpler, more direct structures where information is presented sequentially. In addition to comparing French and English native productions, Lambert et al. (2003) also examined how advanced French-speaking learners of English (L2 English) make use of subordination. This learner dimension is particularly relevant to the current study, as it offers insight into how L1 structures may influence L2 usage. These findings will be examined more precisely in section 2.2.3.2.

Subordinate clauses/ Total clauses	**ENG**	**FR**	**FR with L2 ENG**
% Subordination	19.1	37.0	30.9

TABLE 1: Frequency of subordinate clauses in Lambert et al.'s (2003) study[4]

Regarding the function of subordinate clauses in English, Lambert et al. describe them as often serving to introduce dynamic elements of the environment, especially through gerunds. For instance, sentences like (35) emphasize environmental entities with perception verbs or existential constructions in the main clause (Lambert et al., 2003:50):

(35) While he is climbing the rock just gives away and cracks underneath him

These subordinate clauses highlight the unfolding of an event at a reference point and create a temporal framework for the subsequent actions. Thus, the protagonist is placed in the background within the subordinate clause. In contrast, in French, the authors assert that subordinate clauses tend to position the protagonist in the main clause while relegating environmental elements to the background. For example, in sentence (36), the action's protagonist (*il*) is foregrounded in the main clause, while environmental elements are placed in a less prominent position in the subordinate clause. Adverbial clauses (of cause or manner) and relative clauses are particularly favored in French to organize discourse:

(36) Il évite une feuille qui lui tombe dessus.

Strategies for introducing environmental elements also differ. In English, these elements are often introduced through existential constructions combined with a gerund or perception verbs (Lambert et al., 2003), as illustrated in example (37) with the use of a gerund and in example (38) with perception verbs:

(37) There are rocks falling.

(38) He sees water falling.

In French, this presentation is more varied. Speakers distribute information more evenly between main and subordinate clauses. They may also use nominal or passive forms, with or without relative clauses, to introduce these entities (Lambert et al., 2003). A native French

[4] Free translation of the table.

speaker would therefore use structures like the nominal clause in (39) rather than (40), which would be more typical of an English speaker:

(39) On a un monde agité par le vent.

(40) * Il y a un monde agité par le vent.

The differences between the two languages also extend to their overall syntactic structuring. English frequently employs the canonical SVO[5] structure [NP VP][6], which does not always require a subordinate clause. For example, in sentence (41), the canonical structure is used in response to a question like, "I like your shirt; do you remember where you bought it?" However, in French, such a structure would be inappropriate; a cleft sentence involving a subordinate clause, as in (42), would be required (Lambrecht, 2010):

(41) Isabelle gave it to me.

(42) C'est Isabelle qui me l'a donnée.

2.2.3.2. L1 influence on L2 acquisition

Linguistic transfer is a complex and central phenomenon in the process of second-language acquisition. It involves the transposition of linguistic elements from one linguistic system to another (Novogradec, 2021). Odlin (1989, cited by Jarvis, 2000) offers a more detailed definition: "Transfer is the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired." Finding this definition insufficiently precise, Jarvis (2000) reformulates the concept of L1 (first language) influence as follows: "L1 influence refers to any instance of learner data where a statistically significant correlation (or probability-based relation) is shown to exist between some feature of learners' interlanguage (IL) performance and their L1 background." A widely shared idea in lexical transfer studies is that similarities between the source and target languages play a determining role.

Most scholars agree on the distinction between two types of transfer: positive transfer and negative transfer. Positive transfer, according to Bardovi-Harlig and Sprouse (2017, cited in Novogradec, 2021), manifests as the ease with which learners acquire the target language due to specific features of their native language. Conversely, negative transfer is associated with

[5] Subject-Verb-Object

[6] "NP" stand for "noun phrase" and "VP" stands for "verb phrase"

difficulties and errors caused by the influence of the native language on learning the target language (Novogradec, 2021).

Research shows that learners often rely on their L1 as a strategy for addressing linguistic challenges in L2. This reliance can manifest in lexical, syntactic, or phonological forms. For example, French learners of English frequently transfer their preference for subordination, a prominent feature of French, into their L2 productions (Lambert et al., 2003). As noted in studies, French speakers employ a higher rate of subordinating clauses in French than English speakers do in English, reflecting their tendency to prioritize complex syntactic constructions in discourse.

Analyzing the rate of subordinate clauses used by French learners in their English productions (Table 1), a decrease in such structures (30.9%) is noticeable compared to their L1 productions (37%) (Lambert et al., 2003). This could potentially be explained by the learners' conscious or unconscious efforts to employ less complex sentence structures to counter the difficulties of producing foreign-language constructions. However, this rate remains significantly higher than what a native English speaker would produce (30.9% in L2 English versus 19.1% in L1 English), which can potentially be explained by an L1 transfer .

Linguistic transfer depends on a wide range of variables, primarily linked to individual learner differences. Li et al. (2022) identify several key factors, including age, personality, motivation, linguistic aptitude, social, educational, and cultural contexts, prior linguistic background (previously learned languages), the type and amount of exposure to the target language, proficiency level in the target language, linguistic distance between L1 and the target language, task type and domain of language use, as well as the prototypicality and markedness of linguistic features.

The similarities between English and French tend to facilitate L1 influence on IL, as noted by Andersen (1983) and Kellerman (1983, 1995, both cited in Jarvis, 2000). However, Ringbom (1978) and Jarvis (2000) nuance this claim by specifying that it applies primarily to formal, rather than semantic, transfers. In other words, while learners may successfully transpose certain linguistically or phonetically similar forms between the two languages, they may still encounter difficulties in accurately transferring meaning, leading to errors, as explained in the following section.

2.2.3.3. Common errors and patterns

Corder (1981) distinguishes two types of errors, both of which are unavoidable in the process of language learning: random errors and errors related to knowledge of the target language. The first type of errors arises from external or internal conditions. These errors, comparable to those observed in communication in the native language, are linked to slips or contextual factors. In contrast, the second type of errors reflects gaps in the learner's linguistic competence. Performance errors (or “random errors”), which Corder prefers to call "mistakes", differ from competence errors (“errors related to language knowledge), which he refers to as "errors." He critiques the traditional classification system used by many educators, which categorizes errors based on criteria such as omission, addition, substitution, or word order. Instead, he proposes a more analytical classification system, allowing for the precise identification of the error's level (e.g., tense, number, gender). This approach provides a more nuanced understanding of errors and helps guide learners toward targeted corrections. According to Corder, analyzing learner errors is essential for their progress. He emphasizes the importance of identifying the causes of errors, which he groups into three main categories: inattention (often due to a lack of focus or haste), misunderstandings (resulting from an incorrect grasp of linguistic rules or concepts), and linguistic interference (where the native language inappropriately influences the learning of the target language). He reminds us that errors are an integral part of the foreign language learning process. Rather than viewing them solely as obstacles, he suggests considering them as opportunities for learning. Understanding their nature and origins not only reduces their occurrence but also strengthens key skills, particularly vocabulary and grammatical structures.

Cornu et al. (1996) present a more detailed typology of errors encompassing seven types: graphic, morphological, and phonological errors; adjective errors; adverb errors; noun errors; verb errors; phrase errors (including compounds and idioms); and finally, syntactic errors. This last category notably includes errors related to subordinate clauses, which are of particular interest in this study.

Regarding common error patterns made by French-speaking learners of English, one prominent issue is the use of the gerund, a structure which is absent from French syntax. French-speaking learners often tend to use the infinitive with *to*, a structure more aligned with their native language, again demonstrating the role of linguistic transfer in foreign language acquisition (Fitikides, 1963). In English, the infinitive with *to* and the gerund are not always interchangeable. Fitikides (1963) asserts that they can be used interchangeably after certain verbs such as "begin," "like," "love," "prefer," etc., but that the gerund (and not the infinitive)

is required after prepositions; after certain verbs ("avoid," "enjoy," "finish," "stop," etc.); after the adjectives "busy" and "worth"; and after specific phrases ("it's no use," "it's no good," "I can't help," etc.). Egan (2008) provides a non-exhaustive list of the most commonly used verbs with the infinitive, such as want, try, refuse, etc and verbs that can either take an infinitive or a bare infinitive form, such as dare, help, need, etc.

Beyond gerund issues, other subordinate clause errors common among French L1 learners involve tense and mood adjustments and relative clause formations. For instance, English requires a sequence-of-tense backshift in reported speech and other subordinate clauses with past-tense verbs in the main clause, which French also observes but with some differences in modal usage. Learners may make errors with English modals in subordinate clauses. Fitikides (1963:21) notes errors like “He told me that he may come today” instead of “He told me that he might come today” when the reporting verb is in the past. These errors, while intralingual to some extent, can be reinforced by French, where the distinction between “will” and “would” (va venir vs. viendrait) or “may” and “might” isn’t made in the same way in indirect speech.

3. Methodology

3.1. Corpus description

This study is based on a comparative analysis of two learner corpora: the ICLE (the International Corpus of Learner English) and ICLE+30.

ICLE, developed under the coordination of Sylviane Granger at the Université catholique de Louvain, is a widely recognized corpus of essay writing by upper-intermediate and advanced learners of English. The third version (ICLEv3), published in 2020, includes data from 25 mother tongue backgrounds and contains over 5.5 million words. The corpus consists mainly of argumentative (and some literary) essays written by students under either timed or untimed conditions. The essays are original productions and are produced either with or without reference tools such as native speakers, dictionaries and grammar references.

The ICLE+30 follows the model of the ICLE but like the Brown family, there is a gap of 30 years. It was created as a counterpart to ICLE, containing similar essay data collected 30 years after the original ICLE subcorpus. While different ICLE subcorpora were compiled at different times, ICLE+30 replicates this timeline in its respective subcorpora. The genre, essay topics, and metadata collection standards align with ICLE, ensuring comparability for diachronic analysis of learner language over the past three decades.

For the purposes of this study, the focus was placed exclusively on the French (FR) subcorpora of both ICLE and ICLE+30, in order to analyze short-term diachronic developments in the interlanguage of French-speaking learners of English. The original ICLE-FR data were collected primarily in the 1990s, while the ICLE+30-FR subcorpus includes comparable data collected approximately thirty years later. At the time of this study, data collection for the ICLE+30-FR subcorpus was still ongoing.

3.2. Sampling procedure

This study investigates the phenomenon of subordination in advanced learner English by conducting a diachronic analysis of essays written by French-speaking learners. The dataset consists of 30 essays per corpus (ICLE-FR and ICLE+30-FR), all written by intermediate to advanced French-speaking students from the Université catholique de Louvain. The total dataset comprises approximately 36,500 words, ensuring sufficient representativeness and statistical

reliability. To ensure consistency and minimize cross-linguistic variation, only essays written by learners from Belgium whose mother tongue is French were selected from both the ICLE-FR and ICLE+30-FR subcorpora. This choice guarantees a homogeneous L1 background and socio-educational context, thereby enhancing the comparability of learner performance across the two corpora. In total, 30 essays per corpus were selected, resulting in a balanced dataset of 60 essays. All texts were argumentative in nature and met the general compilation criteria of the ICLE and ICLE+30 projects. To protect participant privacy and ensure full anonymization, each essay was assigned a neutral identification code. The naming convention follows a simple and systematic format:

- ICLE_FR1 to ICLE_FR30 for texts from the ICLE-FR
- ICLE+30_FR1 to ICLE+30_FR30 for texts from the ICLE+30-FR

These identifiers appear only in the dataset and are not associated with any personal or identifying information. This coding system also facilitates corpus navigation and cross-referencing during annotation and analysis.

The essays were originally provided in plain text (.txt) format and were then manually converted into structured Excel spreadsheets for systematic coding. In these spreadsheets, each word occupies a separate row, allowing for precise segmentation, filtering, and tagging. This word-per-line format facilitates detailed linguistic annotation, including the marking of clause boundaries, identification of subordinators, and classification of subordinate clause types.

The coding procedure was informed by the principles of the Multiwrite project developed at the University of Oslo, which focuses on the systematic annotation of subordinate clauses in L2 writing. Although the full MultiWRITE annotation guide is not publicly available due to confidentiality restrictions, selected components of their methodology (particularly the clause-type distinctions and tagging conventions) were adapted for use in this study. However, the present analysis does not strictly follow all the specifications of the Oslo framework. Instead, the annotation protocol was adjusted to align with the scope, data format, and objectives of the current research. For a detailed explanation of the annotation scheme, clause typology, and coding procedures used in this study, the reader is referred to section 3.3.1.

3.3. Data coding

Corpus data are approached quantitatively and qualitatively in order to analyze subordinate clause usage, combining quantitative corpus analysis with qualitative error analysis. As a reminder, the study examines three key research questions:

1. How has the frequency of subordinate clause usage changed between the ICLE-FR and the ICLE+30-FR?
2. What changes can be observed in the types of subordinate clauses used by upper-intermediate and advanced French-speaking learners of English in the ICLE-FR and the ICLE+30-FR?
3. What are the most common error patterns in subordinate clause usage in the ICLE-FR and the ICLE+30-FR, and how have they evolved diachronically?

To address these questions, a structured coding system is applied to the dataset, as described in the following section.

3.3.1. Coding criteria

As explained above, each essay is transposed into an Excel-based coding framework, where each word is assigned a specific category. The coding scheme includes the following columns:

- **sentence number:** Each sentence is sequentially numbered
- **word number:** Each word within a sentence is assigned a numerical order
- **word:** Each word of the text appears on a separate row
- **status:** Identifies the type of subordinate clause
- **clause type:** Categorizes subordinate clauses based on their syntactic role in the sentence
- **sc_element:** Tags elements triggering subordination
- **sc_element_word:** Tags the word used as triggered element
- **dep_sc_ext:** Indicates the function of each subordinate clause within the sentence
- **dep_on:** Indicates the verb on which the subordinate depends
- **gram:** Flags grammatical errors
- **notes:** Records additional observations relevant to the coding process

Table 1 represents a sample of an Excel file used for coding.

sentence_ number	word	status	clause_t ype	sc_ele ment	sc_ele ment _wor d	dep_s c_ext	dep_on	gram	notes
1	Marx								title
1	once								title
1	said	main							title
1	that								title
1	religion								title
1	was	sc_ that	nominal	compl	that	obj	say		title
1	the								title
1	opium								title
1	of								title
1	the								title
1	masses								title

TABLE 2: CODING EXCEL FILE SAMPLE

The status column, as indicated earlier, refers to the status of subordinate clauses used. These have been coded, respecting the Multiwrite's annotation manual. These include the following categories[7]. In order to respect the confidentiality of the MultiWRITE manual, the examples that follow are not extracted from the manual itself but have been created specifically for illustrative purposes.

- **sc_adv:** adverbial clause (e.g. "I'm studying linguistics because I love grammar.")
- **sc_comm:** comment clause (e.g. "I'm a linguist, you see.")
- **sc_comp:** comparative clause (e.g. "He is a better linguist than I am.")
- **sc_coninv:** finite clause functioning as an adverbial of condition with a subject-verb inversion (e.g. "Should you see her, tell her I'm here.")
- **sc_inf:** infinitive clause (e.g. "To become a linguist, you have to love languages.")
- **sc_ing:** ing_clause (e.g. "Studying languages is my favourite hobby")
- **sc_main:** clause having the form of a main clause and constituting direct speech (e.g. "I admire people who shout, "Linguistics is the best!"")
- **sc_pastp:** past participle clause (e.g. "When left to myself, I practice languages.")

[7] "sc_" stands for 'subordinate clause'

- **sc_relthat:** that-relative or bare relative (e.g. “I love the languages that I’m studying”)
- **sc_relwh:** relative clause with wh-words (e.g. “I love the languages which I’m studying”)
- **sc_tag:** tag-clause (e.g. “You love the English language, don’t you?”)
- **sc_that:** that-clause with or without that-element (e.g. “I’m quite certain that he will come”)
- **sc_wh:** wh-clause (e.g. “I’m wondering what he would be thinking”)

The clause type column categorizes subordinate clauses based on their syntactic role in the sentence. This column did not appear in the initial Multiwrite’s annotation guide and has therefore been added in the present study to better align the analysis with the research objectives. This addition is grounded in the functional-syntactic framework proposed by Quirk et al. (1985), whose taxonomy has been widely adopted in grammatical and learner corpus studies. By incorporating this dimension, the annotation procedure provides a more fine-grained analysis of how learners structure complex sentences over time and how different types of subordination evolve in L2 writing.

- **Nominal:** functioning as a noun (e.g. “I believe that linguistics is interesting.”)
- **Relative:** giving additional information about the preceding noun (e.g. “English is the language that I’m studying.”)
- **Adverbial:** functioning as an adverbial that modifies the main clause (e.g. “I studied linguistics because I love languages.”)
- **Comparative:** expressing a comparison (e.g. “He’s a better linguist than I am.”)

Although both columns include a comparative dimension, the **“**status” column (and by extension “sc_comp”) refers to the formal realization of a clause, i.e., what kind of construction it is, whereas the “clause type” column groups clauses based on their syntactic role in the overall sentence structure. In this sense, “sc_comp” represents one type of form among many in the status system, while “comparative” in the clause type system captures its function, alongside clauses that may differ formally but serve a similar comparative purpose. This separation also reflects the dual perspective adopted in previous linguistic frameworks (Quirk and Greenbaum, 1973), which distinguish between form and function in syntactic analysis.

The next column identifies elements that trigger subordination. This annotation has been adapted from the original Multiwrite Project annotation, where only relative pronouns (previously grouped under a broad "pronoun" category in a POS column) and subordinating conjunctions (treated as a

single category) were annotated. To achieve a more precise analysis, the current study refines this classification by distinguishing subordinators into the following categories:

- **rel_pro:** relative pronoun (e.g. “English is the language that I’m studying”.)
- **zero_rel_pro:** empty relative pronoun (“English is the language I’m studying.”)
- **wh_pro:** wh- pronoun (e.g. “What I think doesn’t matter.”)
- **conj_sub:** simple subordinating conjunction (e.g. “I studied linguistics because I love grammar.”)
- **conj_sub_complex:** subordinating conjunction composed of more than one word (e.g. “She was prepared in case something happened.”)
- **compl:** complementizer (e.g. “I think that English is a beautiful language.”)
- **zero_compl:** empty complementizer (e.g. “I think English is a beautiful language.”)
- **prep:** preposition (e.g. I’m learning English by studying vocabulary.”)

In addition to refining the classification of subordinators, a new column explicitly identifies the linguistic element triggering subordination. This column records specific words or phrases that introduce subordinate clauses, such as *that, if, because, to,* and so on. Since *-ing* clauses and past participle clauses do not require an explicit subordinating element (unless introduced by a preposition) these cases are marked with the symbol “/” in this column. The same notation is applied to empty (relative) *that*-clauses to systematically account for the absence of an explicit subordinating element**.** This distinction ensures that the absence of a trigger is systematically accounted for in the annotation process.

The dependency column (dep_sc_ext) concerns the nature of the dependency (function of the subordinate clause). This has been coded as follows:

- **subj:** subject (e.g. “Studying languages is my hobby.”)
- **obj:** object (e.g. “I’m wondering what she’s doing.”)
- **adjmod:** dependent on an adjective (e.g. “He’s a better linguist than I am.”)
- **adv:** adverbial (e.g. “I chose English because it is a widespread language.”)
- **advmod:** dependent on an adverb (e.g. “He thinks faster than I do.”)
- **adv_app:** adverbial clause having an appositional relationship to the preceding clause (e.g. “I want to learn - that is, to study.”)
- **prepobj:** dependent on a preposition (e.g. “I’m learning English by studying vocabulary.”)
- **postmod:** dependent on a preceding noun phrase (e.g. “She is a person who reads a lot.”)

- **postmod_app:** clause having an appositional relationship to a preceding noun phrase (e.g. "Learning a language is many things – listening, speaking, reading and writing.")
- **cleft:** clause containing an it-cleft (e.g. "It is grammar that is interesting.")
- **disl:** clause with a dislocation structure (e.g. "Speaking – that is the most difficult part.")
- **extr:** extraposed complement clause (e.g. "It is difficult to learn a language.")
- **foc:** focused clause of an it-cleft (e.g. "It is what we learn that matters.")
- **speech:** direct speech (e.g. "I shouted "that was wonderful"".")

In addition to the structural annotation of subordinate clauses, a second layer of manual annotation was carried out to identify grammatical errors specifically within these clauses. The aim of this error annotation was to shed light on the types of difficulties encountered by French-speaking learners of English in the construction and use of complex syntactic structures. As with the clause annotation process, the error tagging was performed manually. This part of the analysis is based on the manual annotation of grammatical errors detected in subordinate clauses using in particular the categories defined in the work of Corder (1981) and Cornu et al. (1996). The error typology used in this study was constructed on the basis of existing theoretical frameworks, particularly the work of Corder (1981), Cornu et al. (1996) and Fitikides (1963) (see Section 2.2.3.3.). Corder (1981) offers a foundational distinction between mistakes (random or performance mistakes) and errors (systematic errors related to competence). He also argues for a classification based not only on surface form (e.g., omission, addition, substitution) but also on the linguistic level involved (e.g., morphology, syntax, lexis). Cornu et al. (1996), meanwhile, propose a more granular taxonomy of learner errors, dividing them into seven major types (e.g., morphological, verbal, nominal, syntactic), which include errors in subordinate constructions. Building on these approaches, this study adopted a functional and morphosyntactic error typology specific to subordinate clauses, which includes categories such as incorrect use of gerunds and infinitives, subordinator misuse, complementizer omission or redundancy, subject-verb agreement errors, clause attachment problems, and lexical interference. While the manual annotation process ensured greater precision, it also posed challenges in terms of subjectivity. To mitigate this, ambiguous cases were re-evaluated after a delay.

As coding was conducted by a single researcher, a second round of verification was performed on a randomly selected 20% of the dataset after a two-week interval to assess intra-rater reliability. The consistency of annotation decisions was calculated using Cohen's Kappa, yielding a value of $k = 0.78$, which indicates a substantial level of agreement (Landis & Koch, 1977). This reinforces the overall reliability and consistency of the annotation process.

3.3.2. Data Analysis

The analysis was conducted in two complementary phases (quantitative and qualitative) in order to explore diachronic developments in the use of subordinate clauses by French-speaking learners of English

The quantitative phase involved several statistical procedures aimed at identifying significant changes between the ICLE-FR and ICLE+30-FR. First, a frequency analysis was conducted to measure the average number of subordinate clauses per sentence across both datasets. Due to a violation of the normality assumption, a Mann–Whitney U test was used instead of a parametric t-test to compare mean frequencies. Subsequently, two Chi-square tests of independence were performed. The first examined the distribution of functional clause types (nominal, adverbial, relative, and comparative) across both corpora. The second focused on syntactic subtypes (e.g., infinitive clauses, *that*-clauses, gerund clauses) to determine whether the usage of specific constructions evolved over time. Cramer's V was calculated to measure effect sizes and assess the strength of these associations. Finally, a third Chi-square test was conducted to compare the frequency of grammatical errors in subordinate clauses across corpora, providing insight into whether accuracy improved over time. All statistical tests were performed using R platform[8], and significance was determined at the $p < .05$ level.

The qualitative phase of the analysis served a dual purpose. First, it focused on identifying and categorizing common error patterns in the use of subordinate clauses, following frameworks proposed by Corder (1981), Cornu et al. (1996), and Fitikides (1963) (see Section 2.2.3.3.). Emphasis was placed on morphosyntactic issues (e.g., tense consistency, use of subordinators, and clause integration), with attention to potential L1 transfer effects. Second, the qualitative analysis involved a close reading of a broad sample of learner essays in order to explore variation in subordination strategies**.** This included identifying preferences for specific clause types (e.g., that-clauses, relative clauses, infinitives) and observing stylistic and structural tendencies that differ across time periods. This allowed for a more nuanced understanding of syntactic development, going beyond statistical frequencies to consider patterns of use and complexity in context**.**

By combining quantitative methods with detailed qualitative examination of learner output, this mixed-methods approach provides both statistical and interpretive insights into the diachronic

[8] Available on https://langtest.jp.

evolution of subordination in learner English. These findings are further explored in the discussion section (Section 5).

4. Results

This section aims to present the results of the empirical analysis carried out using the ICLE and the ICLE+30. It seeks to highlight possible diachronic changes in the use of subordinate clauses by French-speaking learners of English as a foreign language. The analysis is structured around three main areas, directly related to the research questions: the frequency of subordinate clause usage, their distribution by type, and the error patterns observed. Each sub-section develops these aspects based on the data coded manually according to the methodology outlined earlier, and aims to outline an evolution in syntactic complexity over the past thirty years.

4.1. Qualitative analysis

This section presents a qualitative analysis focusing on two types of non-finite subordinate clauses: *to*-infinitive clauses and gerund (*-ing*) clauses. These two structures were selected because they emerged as the most frequently used subordinate clause types in both the ICLE-FR and ICLE+30-FR. By concentrating on infinitive and gerund clauses, the analysis can delve into detailed usage patterns and learner accuracy. It must be noted, however, that this is a limited focus: other subordinate constructions (e.g. *that*-clauses, past participle clauses, comparative clauses, etc.) are beyond the scope of this section. The following analysis thus provides insight into how French-speaking learners employ *to*-infinitive and *-ing* clauses in their writing, comparing examples from the two corpora to highlight differences in usage patterns and accuracy.

4.1.1. Infinitive subordinate clauses in ICLE-FR and ICLE+30-FR

French-speaking learners in both corpora made extensive use of infinitive subordinate clauses, often to express purposes, intentions, or to complement certain verbs and adjectives. A common pattern in both ICLE-FR and ICLE+30-FR is the use of *to*-infinitive clauses to convey purpose or reason. For example, an ICLE-FR essay provides the sentence:

(43) “In order to avoid any narrow-mindedness, it would be more appropriate to tackle the problem of technology versus imagination within a larger framework.” (ICLE_FR2)

Here the writer opens with an “in order to” clause expressing the purpose or goal for the action that follows. This purposeful use of an initial infinitive clause is mirrored in the ICLE+30-FR data. One student writes,

(44) “Most teenagers, after high school, decide to attempt to obtain a college degree. It appears to be an obvious choice in order to gain more knowledge and hence to be more competent to achieve a notable career.” (ICLE+30_FR2)

In this example, the learner employs a chain of infinitives (*to gain*, *to be*, *to achieve*) to articulate the rationale and expected outcome of attending university. Both excerpts (43) and (44) illustrate how learners use infinitive clauses as adverbial modifiers to answer “why?” which is a clear communicative function of the *to*-infinitive in their essays (namely, to state aims or reasons).

Infinitive clauses also appear in both corpora as complements to verbs, nouns, or adjectives, integrating smoothly into sentence structures. Learners commonly used *to*-clauses after reporting verbs or modality expressions (e.g. *want to*, *have to*, *decide to*, *attempt to*). For instance, ICLE+30-FR contains sentences like (45), which feature infinitives as verb complements (to follow, to provide).

(45) “Indeed, universities offer different levels of preparation. Students who obtained their bachelor degree have the possibility to follow a master class (…). Furthermore, universities have made an effort to provide students a better preparation (…)” (ICLE+30_FR2).

In the ICLE-FR, we likewise find infinitives fulfilling complement roles. One notable example is sentence (46), defining the function of television. In this case, the infinitive clause “to be an open window on the world” serves as a subject complement, specifying what the “noblest function” of television is. The structure *“*X is to be Y*”* is used effectively to convey a definition or identity.

(46) “To me, television [loses], here, its noblest function which is to be an open window on the world.” (ICLE_FR1)

Another ICLE-FR example (47) shows an infinitive clause as part of a noun phrase. Here “the first thing to do” functions as a noun phrase (with an embedded infinitive *to do*), followed by a clause using a gerund (“defining the terms”[9]).

(47) “the first thing to do the[n], is defining the terms.” (ICLE_FR3)

Overall, both corpora demonstrate that learners frequently integrate *to*-infinitive clauses after predicates or as complements, allowing them to produce more complex sentences conveying goals (“to tackle the problem…”), explanations, or plans.

[9] This mix of forms will be discussed later

In terms of rhetorical functions, infinitive clauses play a notable role in organizing arguments and connecting ideas in the essays. Both ICLE-FR and ICLE+30-FR writers use infinitives to introduce conclusions or summaries. A recurring formula in the newer corpus is "To sum up, …", which several ICLE+30-FR students employ to begin their concluding remarks. This mirrors usage in the ICLE-FR data, where we also see transitions like "To sum up, equality in the full meaning of the word is hardly possible…" (ICLE_FR26). Such phrases show learners using fixed infinitival expressions to guide the reader through the text's argumentative structure. Infinitives are also used to indicate the writer's approach or intentions in the essay. For example, an ICLE+30-FR author states the following sentence (48), which explicitly lay out their plan with an infinitive clause.

(48) "This argumentative essay will attempt to demonstrate just how true this quote is in today's society by addressing the difficulties inherent to enforcing such a principle (…)." (ICLE30_FR1)

An ICLE-FR writer similarly announces sentence (49), using an infinitive to outline the essay's aim.

(49) "I will attempt to show that their actions have brought not only advantages to women but inconveniences too." (ICLE_FR8)

These examples highlight that in both eras, learners made use of infinitive clauses for important communicative functions: stating goals, signaling logical relations (cause-purpose or conclusion), and structuring the essay's flow.

Accuracy in the use of infinitive clauses is generally high, but there are some recurrent error patterns to note, and a diachronic comparison suggests which issues persisted or diminished. Both corpora show that French-speaking learners sometimes struggle with contexts where English syntax demands an explicit subject or a different structure with the infinitive. A clear example is the verb "allow". In French, "permettre de" is used without an explicit personal object, which often leads learners to produce constructions like "allows to [verb]…" directly, omitting the English object. In the ICLE-FR data, one student discussing censorship writes the sentence (50). Here the learner treats "allow" as if it were "permettre de", failing to specify who is allowed to distinguish. The correct English phrasing would require an object (e.g. "allows one to distinguish" or "allows us to distinguish").

(50) "(…) the recipe that allows to distinguish infallibly between eroticism (which may border on art) and pornography (sex-maniacal smut)?" (ICLE_FR6)

In particular, this missing subject error is visible in the ICLE+30-FR as well. A contemporary learner similarly writes, "The practice also allows to question the theory…", instead of "allows

one/us to question the theory." The persistence of this error from ICLE-FR to ICLE+30-FR suggests a continued L1 influence: French learners map the French structure directly into English, resulting in a non-native infinitive construction. This indicates that, despite overall progress, certain subtle syntactic differences (like "allow" requiring an object before an infinitive) still pose challenges in the newer generation.

French-speaking learners also show confusion in choosing between infinitive and gerund complements in specific contexts, though there may be slight improvement over time. For example, some verbs and adjectives in English typically take a gerund complement, but learners occasionally use an infinitive instead (or vice versa). In the ICLE+30-FR we find sentence like (51). This contains multiple issues : aside from the non-standard "uncapable" [10] , the structure "incapable to be confronted to…" is non-idiomatic. A native construction would be "incapable of facing the real professional world". Here the learner should have used a gerund after the preposition *of* (i.e. "incapable of being confronted with…" or better "incapable of facing…"), but instead used an infinitive (to be confronted) following the French pattern "incapable de". This error, where an infinitive is used in place of a gerund, reflects negative transfer from French and was more frequent in the older corpus.

(51) "(…) they feel [in]capable to be confronted to the real professional world" (ICLE+30_FR2)

In ICLE-FR, for instance, learners generally avoided the structure "incapable of …ing", and while we did not find the exact phrase "incapable to do" in the older sample, similar constructions (like "difficult to be done" instead of "difficult to do" or misuse of "able/unable") were noted anecdotally in error tags. The ICLE+30-FR example above shows that some of these issues persist, though it is only a single instance. Overall, the frequency of such mis-selection errors appears to be relatively low in both corpora, but when they occur, they reveal consistent types of learner difficulty.

4.1.2. Gerund subordinate clauses in ICLE-FR and ICLE+30-FR

The use of *-ing* subordinate clauses is likewise prevalent in both corpora, and these structures fulfill a variety of functions in learners' essays. French-speaking learners often employ gerund clauses as subject or object noun phrases, as well as after prepositions, mirroring native usage in many cases. One striking example from ICLE-FR shows an *-ing* clause used as the grammatical subject of the sentence (52). Here, the student uses a coordinated gerund phrase ("creating and

[10] Intended "incapable".

inventing") as a singular conceptual subject, correctly paired with the singular verb "involves". This construction demonstrates a high level of competence, as the learner successfully treats the act of innovation ("creating new techniques") as a general concept.

(52) "Creating and inventing new techniques involves imagination." (ICLE_FR13)

Similarly, the newer ICLE+30-FR writings contain sentences like in (53). In this example, the gerund "dreaming" functions as the subject, and the sentence goes on to argue its importance ("...without it, people would not have any goal in their life."). Using a gerund in subject position (instead of, say, an infinitive or a dummy *it* clause) often makes the statement more natural in English, and it is encouraging to see that both corpora's writers use this structure to make general statements.

(53) "And finally, dreaming is a source of motivation for people and without it, people would not have any goal in their life and this one would seems sad." (ICLE+30_FR30)

Interestingly, French-speaking learners typically uses the infinitive in the equivalent construction (e.g. "Rêver donne de la motivation" would use "rêver", an infinitive, as the subject), yet these learners have adapted to the English gerund-subject pattern. This suggests that by the time of writing, many learners had internalized when English prefers *-ing* forms for general statements or activities, an important aspect of advanced usage.

Gerund clauses also frequently appear as objects of prepositions or in adverbial roles, linking ideas and conveying manner, cause, or condition. In both corpora, learners show awareness that after a preposition, a verb should take the *-ing* form. For example, an ICLE-FR student writes the sentence (54) about television programs:

(54) "(…) because these programmes also aim at distracting and amusing people in a clever way as they can prompt people to enlarge their sphere of interest." (ICLE_FR1)

Here "aim at" is correctly followed by a gerund (distracting, amusing), and the clause "distracting and amusing people in a clever way" describes the goal of the programmes. Likewise, the ICLE+30-FR contains proper prepositional gerunds. One essay notes that universities help students not just academically but also socially, as demonstrated in sentence (55). In this sentence, "by providing a safe environment" is an *-ing* adverbial clause of means. It explains how universities achieve that outcome. The learner has accurately used "by" + "gerund" to express the manner or method, which aligns perfectly with English norms.

(55) "By providing a safe and suitable environment, universities allow young people to work on their personal development and expand their social horizons." (ICLE30_FR5)

Another ICLE+30 example of an adverbial *-ing* clause can be seen in a sentence such as (56), where a series of gerund clauses after "by" enumerate ways to gain experience.

(56) "Indeed, it is said that we will gain more experience by traveling around the world, meeting people, and adapting to different cultures." (ICLE30_FR7)

The older ICLE-FR texts show similar usage. The structure "by showing…" in sentence (57) is used to introduce a cause or justification. These parallel examples underscore that both generations of learners competently use *-ing* clauses after common prepositions like *by*, *without*, *in*, *at*, etc., enabling them to form complex sentences that explain circumstances or methods.

(57) "In their defense we could argue that by showing an idealized version of life, people forget, at least for an hour or two, their miseries." (ICLE_FR1)

Another important function of *-ing* clauses in the data is their use in participle clauses to add information or link clauses in a dynamic way. Learners often employ present participle clauses to avoid repetition and to convey simultaneous or related actions. For instance, in ICLE-FR we encounter sentences like in (58). The essay writer begins with "Taking the statement…", which is a non-finite subordinate clause that provides context or a conditional frame (essentially meaning "If we take the statement…"). This participial clause is correctly attached to the main clause and shares its subject.

(58) "Taking the statement about feminists, for instance, it may be useful to check what 'feminism' is." (ICLE_FR3)

Similarly, sentence (59) from an ICLE+30-FR text, which discuss social class differences, uses a participle clause to concisely convey background. Here, "Being brought up as a girl in Belgium…" functions as an adverbial clause of condition or contrast, setting the stage for the main clause about different life expectations. These kinds of structures, which are stylistically more advanced, appear in both corpora. Rhetorically, *-ing* clauses allow them to vary sentence structure and connect ideas more fluidly, for example, using an *-ing* clause to indicate cause and effect *("..., thus leading to frustrations…"*) or temporal sequence ("after considering these points, ...").

(59) "Being brought up as a girl in Belgium versus in Afghanistan, where women are not allowed to go to school anymore, is a big factor playing a part in the access to education." (ICLE30_FR9)

In terms of accuracy with gerund clauses, French-speaking learners generally perform well, but there are a few points of divergence and some subtle errors, often connected to the broader infinitive/gerund choice issue. One observation is that genuine grammar errors specifically within -ing clauses (such as using the wrong verb form after a preposition, or mis-conjugating the gerund)

are relatively rare in both corpora. When learners decided to use an *-ing* form, they usually did so correctly. The more common mistakes involve using an *-ing* form where an infinitive is expected, or vice versa, which ties back to verb or adjective complementation rules in English. For instance, a learner in ICLE+30-FR wrote sentence (60). The structure here is inaccurate because the verb "to dare" (when used as a main verb) is normally followed by a *to*-infinitive (e.g., dare to do something) (see Section 2.2.3.3.). The phrase should read "dare to take decisions." The error likely arises from an overgeneralization of the rule that many verbs of liking/disliking or risk (enjoy, risk, consider) take gerunds. The student then applied *-ing* after "dare", which is not the standard usage.

(60) "It encourages students to be curious, ambitious and to dare taking decisions." (ICLE30_FR12)

Interestingly, we did not find this particular error in the ICLE-FR sample; it may be an idiosyncratic mistake by one writer rather than a widespread issue. Nonetheless, it exemplifies the kind of confusion that can occur. Conversely, learners sometimes avoid using gerunds where they are actually required, as we saw with "[in]capable to be confronted" instead of "incapable of being confronted". Both corpora exhibit such errors, though informal observation suggests that overusing the infinitive (when a gerund is needed) was more frequent in the older ICLE-FR data, whereas the ICLE+30-FR data show a few cases of the opposite (misusing a gerund or *-ing* construction in a context that calls for an infinitive). This could indicate a slight shift in teaching or learner strategies: modern learners might be more aware of gerund structures (and thus occasionally over-apply them), whereas earlier learners defaulted to the infinitive in doubtful cases.

4.2 Quantitative analysis

4.2.1. Frequency of subordinate clause use

The first aspect of the analysis concerns the overall frequency of subordinate clause use in the two corpora. In the ICLE-FR, 1505 subordinate clauses were recorded across 924 sentences, resulting in an average ratio of approximately 1.63 subordinate clauses per sentence. In the ICLE+30 corpus, 1639 subordinate clauses were identified across 848 sentences, giving a slightly higher ratio of 1.93 subordinate clauses per sentence.

This difference highlights a diachronic trend towards greater syntactic complexity in the more recent written productions of French-speaking learners of English. This suggests that learners in ICLE+30 resort to subordination more frequently, which may reflect an improvement in their

ability to structure complex syntax. Furthermore, the total number of clauses (main and subordinate clauses) is comparable in both corpora (2682 for ICLE-FR versus 2658 for ICLE+30-FR), which confirms that the increase in subordination frequency does not simply result from a larger volume of text, but rather from a more intensive and intentional use of subordination.

Since the first research question addresses the evolution of subordinate clause usage frequency across the two corpora, a Student's t-test was initially considered. Given that the dependent variable (subordinate clause ratio) is a continuous quantitative measure and the independent variable (corpus) is a binary qualitative variable, this type of test would be appropriate. Before comparing the mean ratios of subordinate clauses per sentence between the two corpora, a normality test was conducted to verify the assumptions for applying the t-test. A Shapiro-Wilk test was chosen due to the limited sample size (n = 30 per corpus). This test showed that the ICLE+30-FR corpus data followed a normal distribution ($p = 0.578$), whereas the ICLE-FR corpus data significantly deviated from normality ($p = 2.551e-11$). Given the non-normality of at least one dataset, a non-parametric Mann-Whitney U test was preferred.

The Mann-Whitney test was used to compare the distribution of subordinate clause ratios per sentence between the two corpora. The results indicate no statistically significant difference between the two groups ($p = 0.5013$), suggesting that the overall use of subordination per sentence has remained relatively stable over time.

These findings suggest that, on average, learners in the ICLE+30-FR do not use significantly more subordinate clauses per sentence than those in the ICLE-FR. Therefore, this particular indicator does not strongly support the hypothesis of a diachronic increase in syntactic complexity. This question will be revisited in the discussion (Section 5), in light of other quantitative and qualitative findings.

4.2.2. Types of subordinate clauses

This section focuses on the distribution of the different types of subordinate clauses in the two corpora, with the aim of highlighting possible diachronic shifts in French-speaking learners' preferences for certain syntactic structures. In line with the typology adopted in this study (Quirk et al., 1972; Quirk and Greenbaum, 1973; Biber et al., 1999), subordinate clauses are classified into four main categories: nominal, adverbial, relative, and comparative.

4.2.2.1. Distribution of subordinate clause types

Table 3 shows the distribution of clause types in the ICLE-FR. Nominal subordinate clauses dominate with 869 occurrences (about 57.7% of the 1505 subordinate clauses). Adverbial

subordinate clauses come second with 375 occurrences (24.9%), followed by relative subordinate clauses (233 occurrences, 15.5%) and finally comparative subordinate clauses, which are much rarer (28 occurrences, or 1.9%).

In the ICLE+30-FR, the distribution shows both similarities and interesting changes. Nominal clauses remain in the majority with 861 occurrences out of 1639 (52.5%), but their relative share decreases slightly compared to ICLE-FR. Adverbial clauses, on the other hand, increase significantly, reaching 471 occurrences (28.7%). Relative clauses are also on the rise, with 282 occurrences (17.2%), while comparative clauses remain marginal (25 occurrences, or 1.5%).

CLAUSE TYPE	**ICLE-FR**	**ICLE+30-FR**
Adverbial	24.9% (n= 375)	28.7% (n = 471)
Comparative	1.9% (n = 28)	1.5% (n = 25)
Nominal	57.7% (n = 869)	52.5% (n = 861)
Relative	15.5% (n = 233)	17.2% (n = 282)
TOTAL	**100% (n = 1505)**	**100% (n = 1639)**

TABLE 3: DISTRIBUTION OF SUBORDINATE CLAUSE TYPES IN THE ICLE-FR AND ICLE+30-FR

These results suggest a diachronic trend towards diversification in the types of subordinate clauses used by learners. While the older corpus shows a heavy reliance on nominal structures, which are often easier to handle because of their proximity to French syntax (e.g. 'I think that...'), more recent productions could show a greater proficiency of adverbial and relative structures, which generally require more complex syntactic and discourse processing (Lambert et al., 2003).

Analysis of the raw data suggests that the difference is mainly due to a slight increase in adverbial clauses in ICLE+30 (from 24.9% to 28.7%) and relative clause stability in nominal clauses, which remain by far the most frequent in both corpora. Relative clauses also show a slight increase (from 15.5% to 17.2%), while comparative clauses remain rare and generally constant.

A first statistical analysis was carried out to assess the distribution of functional types of subordinate clauses in the two corpora, namely: nominal, adverbial, relative and comparative.

A Chi-square test of independence was performed after verification of the conditions relating to the use of this test. The variable types corresponded, having a dependent variable (clause types) that was qualitative nominal (with four modalities) and an independent variable (corpus) that was

a qualitative binary variable[11]. In addition, it was essential that at least 80% of the cells had a number greater than or equal to 5, which was the case here. We could therefore carry out the test using a contingency table cross-referencing the clause types with the two corpora (ICLE and ICLE+30). The results indicate a significant difference in the distribution of subordinate clause types between the two corpora ($p = 0.0179$).

The Cramer's V value calculated to measure the size of the effect is 0.056. This value corresponds to a very small effect, according to commonly accepted thresholds (Gries, 2013), which means that, although the difference is statistically significant, it remains small.

These results qualify the hypothesis of a marked restructuring in the use of subordination types. While there is an evolution in the frequency with which learners use certain structures, it remains limited, with no radical transformation in the distribution of syntactic functions in written productions. The trend observed is therefore compatible with a gradual refinement of syntactic skills, particularly in the frequency of clauses used with modifying functions (adverbial, relative), rather than with a revolution in the cognitive patterns of syntactic encoding.

The adjusted standardized residuals were examined to identify which cells contributed most to the significant Chi-square result. The analysis revealed that adverbial clauses were significantly more frequent in the ICLE+30-FR than expected (adjusted residual = 2.41, $p = 0.016$), whereas they were significantly underrepresented in the ICLE-FR (adjusted residual = -2.41, $p = 0.016$). Conversely, nominal clauses were significantly overrepresented in the ICLE-FR (adjusted residual = 2.93, $p = 0.003$) and underrepresented in the ICLE+30-FR (adjusted residual = -2.93, $p = 0.003$). These results suggest a diachronic shift in clause type preference, with learners from the more recent corpus favouring adverbial structures while reducing the reliance on nominal clauses.

In contrast, the adjusted standardized residuals for comparative and relative clauses did not reach statistical significance. The frequency of comparative clauses remained relatively stable across both corpora ($p = 0.466$), and relative clausesshowed no meaningful variation either ($p = 0.192$). These results suggest that the diachronic changes in subordinate clause usage are primarily driven by shifts in the use of adverbial and nominal clauses, rather than in comparative or relative constructions.

4.2.2.2. Distribution of subordinate clause subtypes

Analysis of the distribution of subordinate subtypes in the two corpora reveals both a certain structural stability and significant changes in the syntactic preferences of French-speaking learners

[11] ICLE-FR or ICLE+30-FR

of English. These subtypes, identified via the status variable in the annotation, make it possible to detail precisely the grammatical forms used to express subordination.

SUBORDINATE TYPE	ICLE-FR	ICLE+30-FR
Sc_adv	11% (n = 166)	11% (n = 180)
Sc_comm	<1% (n = 3)	<1% (n = 1)
Sc_comp	2% (n = 29)	1% (n = 25)
Sc_coninv	<1% (n = 1)	0% (n = 0)
Sc_inf	31% (n = 472)	33% (n = 545)
Sc_ing	15% (n = 225)	16% (n = 256)
Sc_main	<1% (n = 5)	1% (n = 13)
Sc_pastp	5% (n = 67)	3% (n = 49)
Sc_relthat	7% (n = 102)	9% (n = 157)
Sc_relwh	10% (n = 146)	8% (n = 127)
Sc_that	14% (n = 215)	11% (n = 176)
Sc_wh	5% (n = 74)	7% (n = 110)
TOTAL	**100% (n = 1505)**	**100% (n = 1639)**

TABLE 4: FREQUENCY OF SUBORDINATE CLAUSE SUBTYPES IN THE ICLE-FR AND ICLE+30-FR

The distribution of subordinate clause subtypes in both corpora are presented in table 4 and illustrated in figure 1 for a better visualization of their diachronic evolution. In the ICLE-FR, the most frequent subordinate clauses are infinitive clauses (sc_inf), which account for 31% of the total, followed by -ing subordinate clauses (sc_ing) with 15%, and subordinate clauses introduced by that (sc_that), which account for 14%. Other types appear in more modest proportions, including relatives introduced by *wh-* (sc_relwh, 10%) or by *that* (sc_relthat, 7%), adverbial clauses (sc_adv, 11%), *wh*-clauses (sc_wh, 5%) and past participle clauses (sc_pastp, 5%). Other forms such as comparatives (sc_comp), integrated main clauses (sc_main), comment clauses (sc_comm) and conditional constructions with inversion (sc_coninv) remain minor.

In the ICLE+30, the overall trends remain unchanged, but some changes are apparent. Infinitive clauses remain predominant, accounting for 33% of the total, while *-ing* clauses increase slightly to 16%. *That*-clauses, on the other hand, show a relative decrease to 11%. Relative clauses introduced by *that* (sc_relthat) increase slightly to 9%, while those introduced by *wh-* rise to 8%. There is also an increase in nominal *wh*-clauses (sc_wh, 7% vs. 5% in the ICLE-FR) and a slight drop in past participles clauses (sc_pastp, 3% vs. 5%). Adverbial clauses remain stable at 11%, while the other categories are still poorly represented.

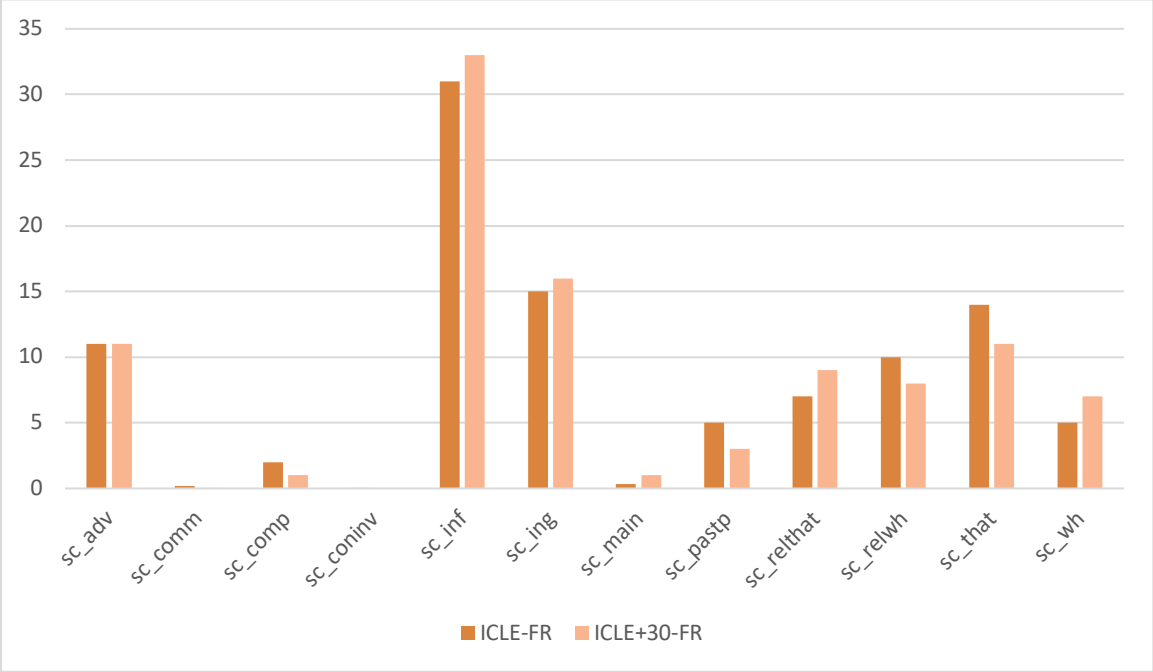

FIGURE 1: DISTRIBUTION OF SUBORDINATE CLAUSE SUBTYPES IN ICLE-FR AND ICLE+30-FR

These results suggest several diachronic dynamics in the syntactic use of subordinate clauses. Firstly, the increase in the use of non-finite clauses (in particular *-ing* clauses and infinitives) seems to indicate a greater ease on the part of recent learners productions in manipulating complex syntactic constructions, often required in an academic register. Secondly, the use of relative clauses remains relatively stable, but the slight variations between *that* and *wh-* suggest a progression in the proficiency of relative clauses, as well as a more refined use of linking markers. Thirdly, the increase in nominal *wh*-clauses could reflect an enrichment of discourse strategies, as these structures are often used to express abstract content, pose hypotheses or structure indirect reasoning (e.g. “What matters is...”). Finally, the stability of adverbial clauses in both corpora confirms their central role in the logical construction of written argumentation. Their consistency

suggests that learners, whatever the period, continue to rely on these structures to express cause, concession, purpose, time or condition which are essential relationships in any academic text.

To conclude, the evolution observed in the distribution of subordinate subtypes reflects a tendency towards gradual complexification and diversification of the syntactic structures used by French-speaking learners of English. This progression can be interpreted as a possible sign of improvement in grammatical competence, potentially fostered by more explicit teaching methods, greater exposure to authentic written English or even higher academic expectations in recent institutional contexts.

In addition to examining the functional clause types of subordinate clauses, a third statistical analysis was conducted to assess the distribution of subordinate subtypes. In order to determine whether the distribution of subordinate subtypes differed significantly between the two corpora (ICLE-FR and ICLE+30-FR), a Chi-square test of independence was performed on a contingency table crossing subordinate types with the two corpora. As with the analysis of changes in the types of clauses used among subordinate clause types (see Section 4.2.1.), the test conditions were also checked. In order to guarantee the reliability and statistical robustness of the analysis, and to avoid bias due to small numbers (outliers), subordinate subtypes occurring in less than 1% of cases in both corpora (i.e. sc_coninv and sc_comm) were excluded from the test. The final contingency table thus included 10 representative subordinate types, for a total of 3139 subordinate clauses distributed between the ICLE-FR (n = 1501) and the ICLE+30-FR (n = 1638).

The test results indicate a significant difference in the distribution of subordinate types between the two corpora ($p < 0.001$). To measure the extent of the effect, Cramer's V was calculated. Its value is 0.10. According to interpretative standards (Gries, 2013), this value corresponds to a small effect size, indicating that the difference observed, although significant, remains moderate in intensity. These results suggest that there is a diachronic evolution in learners' structural preferences, but that this evolution remains relatively minor across all the subtypes considered.

To determine which subordinate clause subtypes contributed most to the statistically significant Chi-square result, adjusted standardized residuals were examined. Residuals with an absolute value greater than 1.96 were considered significant at the $p < .05$ level. The analysis revealed that several clause types contributed disproportionately to the observed differences between the ICLE-FR and the ICLE+30-FR: *That*-clauses (sc_that) showed the strongest contribution, being significantly overrepresented in the ICLE-FR and underrepresented in the ICLE+30-FR ($z = \pm 3.03$, $p = .002$). Past participle clauses (sc_pastp) also contributed to the significance, with more occurrences than expected in ICLE-FR and fewer in ICLE+30-FR ($z = \pm 2.18$, $p = .029$). Then, relative clauses introduced by *that* (sc_relthat) were significantly underused in the ICLE-FR and

overused in the ICLE+30-FR (z = ±2.84, p = .005). *Wh*-clauses (sc_wh) contributed to the significance as well, appearing more frequently than expected in ICLE+30-FR (z = ±2.13, p = .033). These residuals indicate that these four clause types (sc_that, sc_pastp, sc_relthat, and sc_wh) are the main contributors to the statistically significant difference in subordinate clause subtype distribution across the two corpora.

4.2.3. Error patterns

Analysis of the errors made by French-speaking learners in the use of subordinate clauses is an essential aspect of this study. Not only does it make it possible to identify areas of persistent weakness in the proficiency of complex syntactic structures, but it also makes it possible to compare the nature and frequency of errors between the two generations represented by the ICLE-FR and ICLE+30-FR.

4.2.3.1. Errors detected

The errors identified in subordinate clauses are mainly morphosyntactic problems and poor choices of structure, often due to the influence of the mother tongue (negative transfer) or to a partial proficiency of the constructions expected in English. Based on the typology of errors presented in section 2.2.3.3., the errors detected in both corpora are classified in this chapter.

Firstly, we can identify numerous cases of incorrect use of the infinitive and gerund. This phenomenon is particularly visible in the ICLE-FR, where learners frequently use an infinitive instead of a gerund, or vice versa, in contexts requiring a precise construction. Erroneous sequences such as "The first thing to do is defining the terms" can be observed instead of "The first thing to do is to define the terms". Similarly, the use of bare infinitives is often poorly mastered, particularly after certain modal or causative verbs such as ‘help to go’ instead of ‘help go’.

A second set of frequent errors concerns the incorrect use of subordinators, particularly in relative clauses. Confusion between *that* and *which* is regularly observed, as well as which-clauses used without a clear antecedent, undermining syntactic consistency.

Errors relating to subject-verb agreement and verb tense management were also found in both corpora. These included disagreements such as “was” instead of “were” in conditional sentences, and incorrect conjugations in temporal and causal subordinate clauses. These errors are often attributable to local inattention or incomplete syntactic analysis of the complex structure.

Another type of error concerns incomplete or faulty syntactic structures. Some subordinate clauses appear isolated or poorly integrated into the main sentence, such as floating verb forms (*dreaming* used without subject or auxiliary), or sequences in which the subject or verb is simply omitted.

Finally, cases of lexical calques from French were identified, showing typical cross-linguistic interference. Expressions such as “to take benefit of” instead of “to benefit from” or “to peep into”[12], modelled on a French idiom, illustrate persistent negative transfer.

4.2.3.2. Frequency and evolution of errors

The diachronic analysis of learner errors reveals a notable reduction in both the total number of errors and the overall error rate per subordinate clause between the two corpora (ICLE-FR and ICLE+30-FR). As shown in Table 5, the number of errors falls from 98 in ICLE-FR to 55 in ICLE+30-FR, representing a 43.9% decrease. When related to the total number of subordinate clauses produced in each corpus, this corresponds to an overall error rate drop from 6.51% to 3.36%, suggesting a general improvement in learners’ syntactic accuracy over time.

[12] While this structure is grammatically correct in isolation, it is semantically inappropriate in the context used by the learner.

	ICLE-FR		ICLE+30-FR	
	Number of errors	**Error rate per subordinate clause**	**Number of errors**	**Error rate per subordinate clause**
sc_inf (infinitive)	32	6.78	17	3.12
sc_ing (gerundive)	18	8	8	3.13
sc_relthat (relative with that)	8	7.84	5	3.18
sc_relwh (relative with *wh*-)	12	8.22	9	7.09
sc_that (complement introduced by *that*)	10	4.65	7	3.98
sc_adv (adverbial)	11	6.63	3	1.67
sc_wh (nominal with wh-)	3	4.05	6	3.68
Sc_pastp (participle)	1	1.49	0	0
Sc_comp (comparative)	3	10.34	0	0
Total	**98**		**55**	

TABLE 5: NUMBER OF ERRORS AND ERROR RATES PER SUBORDINATE CLAUSE SUBTYPE DETECTED IN ICLE-FR AND ICLE+30-FR

The distribution of error types indicates that the reduction is not uniform across all subordinate clause categories. Some types show a drastic decrease or complete disappearance of errors, while

others remain relatively stable or even increase slightly in absolute numbers. Comparative clauses (sc_comp) and past participle clauses (sc_pastp) are particularly notable for their complete elimination of errors in ICLE+30-FR. While their overall frequency is limited in both corpora, the absence of errors may suggest that learners now avoid these forms when uncertain, or that they have gained better control over their usage. Adverbial clauses (sc_adv) show a significant improvement, with errors dropping from 11 to 3 and the error rate decreasing by nearly five percentage points (from 6.63% to 1.67%). The infinitive (sc_inf) and -ing clauses (sc_ing), which together represent the majority of non-finite constructions, also display a marked improvement, with error rates falling by more than 50%. *That*-clauses (sc_that) and that-relative clauses (sc_relthat) show moderate improvements in accuracy. Interestingly, wh-relative clauses (sc_relwh) and nominal *wh*-clauses (sc_wh) show only minimal improvement or slight increases in error numbers. For example, errors in sc_relwh decrease marginally in rate (from 8.22% to 7.09%) but not in absolute terms (12 to 9 errors), while sc_wh errors rise from 3 to 6.

These findings suggest that while learners have generally improved their proficiency of subordinate clauses, some structures remain particularly challenging, either due to their syntactic complexity, semantic opacity, or interference from L1 transfer patterns. The disappearance of errors in specific categories could also reflect strategic avoidance or changes in learner exposure and instruction over time.

In order to determine whether the frequency of errors linked to the use of subordinate clauses changed significantly between the two corpora, a Chi-square test of independence was performed on the number of errors observed in the ICLE-FR (n = 1505 subordinate clauses) and the ICLE+30-FR (n = 1639 subordinate clauses). The contingency table was structured according to two modalities: ‘error’ and ‘no_error’, for each of the two corpora.

The Pearson test revealed a highly significant result ($p < .0001$), indicating a statistically significant relationship between the corpus and the presence or absence of errors. In other words, the proportion of errors is not independent of the corpus examined and suggests a real evolution in the grammatical quality of learners' written productions over time. In terms of effect size, the Cramer's V index was 0.0733, which corresponds to a small but significant effect (Gries, 2013).

It is important to note that, unlike the previous analysis of subordinate clause types, where residuals were computed for each type and subtype, the present Chi-square test assesses the global distribution of errors versus non-errors across the two corpora. As the contingency table consists of only two variables (Error and No_error) residuals can only indicate whether the overall number of errors differs significantly between ICLE-FR and ICLE+30-FR. However, they do not allow us to assess which specific subtypes of subordinate clauses contributed most to this significance. To

identify which categories drive the observed effect, a more granular analysis (e.g., separate tests for each clause subtype) would be required. It was not performed in this study due to limited statistical expertise on our part.

These results provide statistically valid confirmation of the qualitative observations made earlier: there has indeed been an overall improvement in proficiency of subordinate structures in the most recent corpus.

5. Discussion

The results of this study reveal clear diachronic shifts in the use of subordinate clauses by French-speaking learners of English. In this section, these empirical findings are interpreted in light of SLA theories on syntactic complexity, interlanguage development (including Processability Theory), and L1 transfer. The discussion is organized around the three main research questions, examining in turn: (1) changes in the frequency of subordinate clause usage, (2) shifts in the types of subordinate clauses employed, and (3) the evolution of common error patterns. Through this structure, we link the quantitative and qualitative results to the thesis's theoretical framework and highlight the significance of the observed diachronic developments in learner writing.

5.1. Increase in Subordinate Clause Frequency

The quantitative findings reveal a modest rise in subordinate clause density in the more recent corpus, suggesting a potential increase in syntactic complexity over the past three decades. French-speaking learners in the ICLE+30-FR produced on average 1.93 subordinate clauses per sentence, up from 1.63 in the original ICLE-FR data (Section 4.2.1). At face value, this diachronic increase aligns with general trends in second language development: as learners gain proficiency, they tend to use more complex sentence structures, notably through increased subordination. Kalnbērziņa and Rūtenberga's (2012) learner corpus study, for instance, demonstrated that the number of subordinate clauses rises with improving proficiency, underscoring subordination as an indicator of linguistic development. Similarly, Ortega (2015) emphasizes that syntactic complexity growth often manifests via greater clause embedding, especially in instructed contexts where advanced learners incorporate more subordinate constructions into their writing. The observed increase in subordinate clause frequency between ICLE-FR and ICLE+30-FR can thus be interpreted as a sign of interlanguage development: the newer corpus appear slightly more proficient at structuring complex sentences, which reflect an evolutive direction in their interlanguage competence.

This interpretation finds support in the broader literature on L2 complexity. Longitudinal and cross-sectional studies have long identified subordination as a sign of growing syntactic sophistication. By comparing learner corpora separated by thirty years, we understand how instructional and experiential changes might have influenced learners' writing. Over time, improved teaching methodologies, increased exposure to English, and higher awareness of academic writing conventions may have enabled students to utilize complex sentences more confidently. It is possible that contemporary learners exploit this genre potential more fully than

their 1990s counterparts did, using subordinate clauses in their essays to express nuanced arguments.

Importantly, the role of L1 influence must be considered in interpreting these results. French (the learners' native language) is characterized by a high degree of subordination in discourse. Lambert et al. (2003) found that French speakers use subordinate clauses far more often than English speakers (37% vs. 19% of all clauses), and advanced French learners of English also showed a relatively high subordination rate (30.9%) in L2 English. This tendency for complex sentence structure in French discourse tends to transfer into learners' English writing. In Lambert et al.'s study, French learners produced significantly more subordination in English than native English norms, a pattern attributed to positive transfer of L1 rhetorical strategies. However, those learners still used fewer subordinates in L2 than they would in their L1, presumably to mitigate the difficulty of managing complex syntax in a foreign language. Within our data, the consistently high frequency of subordinate clauses in both corpora suggests that this French L1 influence has persisted across generations.

Despite the observed upward trend, it is important to note that the difference in subordinate clause frequency was not statistically significant ($p = 0.5013$). In practical terms, this means that the overall use of subordination per sentence has remained relatively stable over the 30-year span, with only a subtle and non-significant increase. This finding tempers the initial expectation of a clear diachronic complexity gain. One interpretation is that advanced French-speaking learners had already reached a high level of subordination use in the 1990s, which leave limited room for further increase. Additionally, as Crossley and McNamara (2014) argue, the nature of syntactic development at advanced stages can shift away from simply adding more clauses toward more economical complexity, such as phrasal sophistication. Advanced academic writing often favors dense noun phrases over stacking clauses, meaning that improvements in writing complexity might not register as higher subordinate clause counts. The present results allign with this insight: subordinate clauses per sentence represent only one dimension of complexity.

5.2. Shifts in Types of Subordinate Clauses Used

Beyond frequency, this study examined which types of subordinate clauses learners favor or avoid, and how these preferences have shifted diachronically. The overall distribution of subordinate clause types (nominal, adverbial, relative, comparative) changed in noteworthy ways from the ICLE-FR to the ICLE+30-FR, albeit in a gradual manner. In both corpora nominal clauses remain the most common, but their dominance has slightly waned (from ~57.7% of all subordinates to ~52.5%). Meanwhile, adverbial clauses became more prominent, rising from

roughly 24.9% to 28.7% of all subordinate clauses. Relative clauses also saw a slight increase (15.5% to 17.2%), and remain an important minority type. Comparative clauses, by contrast, were and continue to be minor (around 1–2% in both corpora). Statistical analysis confirmed that the distribution of clause types across these four categories shifted significantly over time ($p \approx 0.018$), though the effect size is very small. This indicates a real but subtle diversification in the types of subordination used by learners.

Interpreting these changes through an SLA lens, we find support for the hypothesis that learners' syntactic repertoire has broadened over time. The reliance on nominal clauses (e.g. content that-clauses like "I think that...") is well-known as a simpler or more default strategy for embedding ideas, one often taught early and heavily used by intermediate learners. In the ICLE-FR data, such "that"-complement clauses were indeed extremely common. Our subtype analysis shows that finite *that*-complements made up about 14% of all subordinates in the older corpus. In the newer corpus this proportion dropped to ~11%, suggesting that students are no longer leaning as exclusively on the safe formula of *that*-clauses to articulate their ideas. This could reflect both instructional shifts and increased confidence: over time, teaching may have placed more emphasis on using a variety of complex constructions, and learners themselves, being more exposed to English, might progress beyond the repetitive use of structures like "I think that…" or "it is true that…". In place of some of these nominal clauses, we see a relative increase in adverbial clauses. Adverbial subordinates require the appropriate use of a diverse set of subordinators (e.g. because, if, when, although). The fact that ICLE+30-FR has a higher share of adverbial clauses than ICLE-FR (nearly 4 percentage points more) is significant. It suggests learners are now more proficient at using logical relations and context into their sentences through subordination. This enhancement corresponds to Lambert et al. (2003)'s observation that French discourse tends to "organize discourse" heavily through adverbial clauses of cause, manner, etc., and that French speakers favor using such clauses to integrate additional detail. The increasing use of adverbial clauses by learners in the current study may therefore indicate a convergence toward both L1 discourse practices and advanced L2 academic style. In argumentative writing, the ability to articulate reasoning (why, how, under what conditions something is the case) is important. Ortega (2015) points out that argumentative texts naturally elicit more subordinate clauses than narratives, offering learners a canvas to practice complex sentence linking. Over the 30-year span, learners might have become more skilled in exploiting this genre potential, using a richer range of adverbial subordination to strengthen argumentation and cohesion in their essays.

Relative clauses also display an interesting diachronic shift. Overall, the combined proportion of relative clauses (of all types) increased slightly in the ICLE+30-FR data. More revealing, however,

are the changes in the subtypes of relative clauses used. The older generation showed a preference for *wh*-relative clauses (approximately 10% of subordinates) over *that*-relatives (about 7%). In the newer generation, this pattern has flipped to some extent: *wh*-relatives decreased to 8%, while relative clauses with *that* rose to 9%. In other words, ICLE+30-FR writers are more willing to use *that* as a relative pronoun than the ICLE-FR writers were. This could be interpreted in a few ways. First, it might signal a move towards more colloquial relative clause usage in modern English usage (especially in speech and informal writing), *that*-relatives are extremely common for defining relative clauses, whereas learners in earlier decades may have overused *which/who*. The change could therefore reflect that learners have internalized a more native-like pattern of relative clause use. Second, it might also be tied to L1 transfer nuances. French learners sometimes find the English system confusing: earlier learners might have been taught to rely on *which* for things and *who* for people, treating *that* as optional or less formal. The result is a higher incidence of *that*-relatives in the newer corpus. This highlights *that*-relatives as a key contributor to the distributional differences between the two corpora. Essentially, the new corpus has adopted a type of clause that the earlier corpus somewhat avoided. This can be seen as a positive development toward a fuller proficiency of English relatives, since native usage often employs whichever relativizer best fits the context without strict avoidance. At the same time, the slight decline in *wh*-relative clauses might indicate that some learners now lean on *that* where a more precise *wh*-word could be used. The net effect is a subtle shift in style and possibly in accuracy (we will discuss error patterns in relatives in Section 5.3.).

Another important diachronic change is seen in the use of nominal *wh*-clauses (also known as free relatives or nominal relative clauses, e.g., "What matters is that…" or indirect questions like "I wonder what will happen"). These are more abstract structures that often embed an entire clause as a noun-like entity. In our data, nominal *wh*-clauses were relatively rare in the first corpus (around 5% of subordinates) but became more frequent in the ICLE+30-FR (about 7%). The residual analysis identified nominal *wh*-clauses as significantly more frequent than expected in the ICLE+30-FR. This suggests that modern learners are experimenting more with structures like "What X [subject + verb] is Y" or "[verb] what…" in their writing. The rise in nominal *wh*-clause usage can thus be interpreted as learners adding more rhetorical and interpretive depth to their essays. It may also reflect teaching that encourages variety in sentence structure for effect, or simply greater exposure to academic texts where these patterns appear. The earlier generation perhaps avoided these due to their complexity or lack of confidence, whereas the newer generation, feeling more at ease, is harnessing nominal *wh*-clauses to articulate more complex propositions. This is a favorable outcome which indicates a maturation in writing style.

The non-finite subordinate clauses (infinitive clauses and *-ing* clauses) were a focus of our inquiry as well, since they represent a sophisticated means of subordination often used in formal writing to condense information. Interestingly, both *to*-infinitive clauses and *-ing* participial clauses showed a slight increase in usage share, with infinitives rising from ~31% to 33% of subordinates and gerundive clauses from ~15% to 16%. In absolute terms, the ICLE+30-FR contains more instances of these non-finite clauses than the ICLE-FR, which is notable given that such constructions require careful control of syntax. The increased use of non-finite subordination suggests that learners have grown more confident in employing compressed subordinate structures, which is a skill associated with advanced writing, as it allows incorporation of information without always using full finite clauses. Indeed, recent learners' greater ease with *-ing* and infinitival clauses might be attributed to better pedagogical emphasis on these forms (for example, teaching not just "I hope that I can do X" but also "I hope to do X," or encouraging use of participial phrases for conciseness). The discussion by Lambert et al. (2003) is relevant here: they observed that English often introduces descriptive information via gerund clauses or participial constructions (e.g., "There are rocks falling", "While he is climbing, the rock gives way…"), whereas French speakers might distribute the same information across multiple clauses or prefer finite relatives. For French learners of English, mastering these non-finite clauses represents both a departure from L1 habits and an attainment of an advanced L2 skill. Our findings hint that more learners have taken that step in the past decades, incorporating gerund and infinitive clauses as a regular part of their syntactic tool.

The dominant patterns of subordination remain recognizably similar between the two corpora: French-speaking learners of English in both eras rely most on nominal clauses, use a substantial proportion of non-finite clauses (infinitives and gerunds), and then supplement those with adverbial and relative clauses, with only tiny minorities of other types (like comparatives, comment clauses, etc.). Here we indeed see slight shifts: less over-reliance on one type (nominal *that*-clauses) and more balance among clause types, which likely reflects improved proficiency. Learners are not suddenly abandoning easier structures. Instead, they add new dimensions to their production. This pattern is consistent with the idea that L2 syntactic development is cumulative. Over time, learners build upon the simpler forms by gradually integrating more complex ones (a concept compatible with Processability Theory's stage-wise acquisition: one does not replace earlier structures but adds to them once ready).

In conclusion, to answer the second research question, we find that the types of subordinate clauses used by French-speaking learners have indeed shifted over time: there is a move toward a more diverse portfolio of subordination. Learners today are slightly less one-dimensional in their

approach (not just stringing together content clauses) and more likely to incorporate different functional types (especially adverbials and certain relative clauses). This evolution reflects a refinement of syntactic skills: learners are not only using subordination more (as seen in the frequency results) but are using it in more varied ways. By employing a range of subordinate structures, they can achieve more precise and rhetorically effective writing. For instance, increased use of causal and concessive clauses allows for clearer argumentative logic, and more frequent use of relative clauses helps integrate information fluidly into sentences. These are precisely the kinds of improvements one would hope to see as learner language matures. The changes also suggest that teaching practices over the last decades may have successfully encouraged a richer use of complex sentences, a point that educators can take pride in while also noting the areas that still need reinforcement (as will be evident in the error patterns discussion below). Overall, the diachronic shift in clause types underscores a trend toward complexity and diversification, aligning with Ortega's (2015) view that higher proficiency brings not just more clauses, but more complex types of clauses. Our learners' interlanguage grammar appears to be catching up with that notion, albeit slowly, by incorporating more of those advanced structures into their writing.

5.3. Evolution of Error Patterns in Subordinate Clauses

Regarding the changes in grammatical error patterns related to subordinate clauses. By comparing the two corpora, we understand whether French-speaking learners have become more accurate in using these complex structures and which specific errors remain persistent. The results show a clear trend: overall subordinate clause accuracy has improved significantly over the past 30 years. The total number of errors in subordination dropped from 98 in the ICLE-FR to 55 in the ICLE+30-FR, a decrease of about 44%. When accounting for the different total numbers of clauses analyzed, the error rate per subordinate clause nearly halved, from roughly 6.5% down to 3.4%. A chi-square test confirmed that this reduction in errors is statistically significant ($p<0.0001$), with a small but noteworthy effect size. In practical terms, these figures mean that a French learner writing an essay today is much less likely to commit a subordination-related error than a similar learner in the 1990s. This positive development fulfills our third hypothesis, which anticipated a diachronic decrease in grammatical errors in subordinate clauses, and it corresponds to the broader improvements one might expect due to better teaching methods, more exposure to English, and possibly more robust and available grammar checking tools (including AI) available to learners.

Examining qualitative error types provides a clearer picture of how and why this improvement occurred. The analysis of the ICLE-FR essays uncovered a set of recurring error patterns in subordinate clauses, many of which can be traced to L1 (French) interference or to developmental difficulties common in advanced SLA. These included: (a) misuse of infinitives vs. gerunds, (b) incorrect use of subordinators or relative pronouns (such as confusion between *that* and *which*), (c) omission or overuse of the complementizer *that* in content clauses, (d) tense and agreement errors inside subordinate clauses, (e) incomplete or fragmented subordinate structures, and (f) lexical calques (translations) from French in subordinate contexts. Each of these error types reflects underlying challenges in being more proficient English subordination. Importantly, our diachronic comparison indicates that most of these error types have become less frequent or less severe in the ICLE+30-FR data.

For example, errors in choosing between the gerund and the infinitive have long cause difficulties among French learners of English. French grammar typically uses an infinitive where English might use a gerund, and French lacks a direct equivalent to the English gerund-participle form. In the ICLE-FR, it was common to find sentences like “He enjoys to play football” instead of “enjoys playing”. Learners often extended the use of *to*+infinitive (a familiar structure from L1 perspective) into contexts that demand a gerund in English. According to Fitikides (1963), who catalogued common mistakes for foreign learners, confusion between gerund and infinitive is indeed a classic error for speakers of languages that do not distinguish them well. Fitikides noted that while some verbs can take either form, others strictly require one or the other, and that French learners often default to the infinitive. Our findings suggest significant progress on this front: the error rate for infinitive clauses fell from ~6.8% to ~3.1%, and for *-ing* clauses from ~8.0% to ~3.1%. In raw terms, gerund-related errors in the corpus dropped from 18 to 8, and infinitive errors from 32 to 17. This is a considerable improvement, indicating that newer learners have a much stronger grasp on when to use “doing” versus “to do”. Such improvement could be attributed to more effective teaching of verb complement patterns and increased awareness of these constructions. It might also reflect learners’ greater exposure to natural English input (through media, internet, etc.), which helps them internalize correct patterns. In short, one major area of difficulty (non-finite verb complements) shows clear evidence of amelioration over time.

Another area of frequent difficulty was the use of relative clauses and subordinating conjunctions, particularly the confusion between *that* and *which* and the misuse of relative pronouns without a clear antecedent. French uses *que* for both defining and non-defining relative clauses (with no exact parallel to the restrictive/non-restrictive distinction that English marks partly with *that* vs. *which*), and French learners often struggle with this nuance. In the ICLE-FR

essays, errors like using *which* instead of *that* in contexts where a defining relative clause was intended, or inserting *which* with no explicit noun antecedent, were not uncommon. Additionally, learners sometimes dropped the relative pronoun or complementizer altogether (e.g., writing "the book Ø I read" in a formal context where English would prefer "the book that I read"), or added *that* inappropriately (perhaps saying "the reason is because that…" or doubling complementizers). These issues point to an incomplete acquisition of the English relativization system and complementizer usage, often exacerbated by interference from French patterns. Over the 30-year span, our data show moderate improvements here. *That*-complementizer errors (omission or overuse) have decreased as learners seem more consistent in including *that* where needed and avoiding it where it's redundant. Similarly, the confusion between that and which has likely been mitigated by better instruction on restrictive vs. non-restrictive clauses or by increased reading of English texts that subconsciously teaches when each is used. While we did not quantitatively separate these sub-categories in the results, the overall error rates for *that*-clauses and relative clauses did drop (e.g., errors in *that*-clauses fell from 4.65% to 3.98%, and in *that*-relatives from 7.84% to 3.18%). These numbers indicate a noteworthy gain in accuracy. The persistent presence of some errors here, however, shows that this remains a challenging area. Specifically, our analysis found that *wh*-relative clauses and nominal *wh*-clauses did not see the same level of improvement: *wh*-relatives had an error rate only slightly reduced (from ~8.2% to ~7.1%), and the number of errors in nominal *wh*-clauses actually increased (from 3 to 6 instances). These stubborn error rates suggest that even advanced learners in the ICLE+30-FR struggle with the more complex or abstract relative constructions, such as embedding a whole clause as a noun or maintaining clear reference in a relative clause. These are nuanced points that require a high level of proficiency, reinforcing Pienemann's notion that certain complex structures remain challenging until very late stages of acquisition.

Tense and agreement errors in subordinate clauses form another category that has shown some improvement but not complete elimination. In both corpora we observed mistakes like using the wrong verb form in conditional clauses, or failing to apply sequence of tenses in reported speech. The fact that adverbial clause errors saw a major decline (from 6.63% to 1.67%) hints that issues in temporal and conditional clauses have been significantly reduced. Learners today likely make fewer errors thanks to a better proficiency of English tense usage in subordination.

The category of incomplete or malformed subordinate clauses also showed signs of improvement. Earlier learners occasionally produced subordinate clauses that were not properly integrated, for example starting a sentence with a subordinate conjunction but never completing the thought, or using a participial phrase without a clear subject. These errors reflect either a

mismanagement of complex sentence construction or an overextension of patterns (sometimes due to L1 influence, sometimes developmental experimentation). In the ICLE+30-FR data, such errors are less frequent. Learners seem more capable of formulating complete complex sentences without leaving subclauses unattached or subjects missing. This is corroborated by the decline in error rates for categories like past participle clauses and comparatives. Remarkably, errors in past participle clauses went from 1 in the ICLE-FR to 0 in the ICLE+30-FR, effectively disappearing. Errors in comparative clauses likewise vanished (3 down to 0). While the base rates for these forms were low to start with, their complete error elimination suggests that when learners did use these less common structures in the newer corpus, they used them correctly. One possible explanation is avoidance: advanced students might avoid constructions they are unsure of, thus making fewer errors simply by not attempting risky structures. Avoidance is a well-documented strategy in SLA. The data indeed show that past participle clauses and comparatives had slightly lower frequencies in the new corpus than the old, which could indicate some avoidance. However, another interpretation is that instruction over time clarified the usage of these forms sufficiently that those who used them did so appropriately. It is likely a mix of both: learners now better understand certain structures (so they make no errors in them) and those who might otherwise misuse them have learned to either use them correctly or not use them at all.

The reduction in L1-influenced lexical calques is also worth noting, though such errors were never extremely frequent. Earlier essays showed instances of direct translations from French embedded in subordinate clauses, such as "to take benefit of" (calquing French "profiter de", instead of saying "to take advantage of"). These errors illustrate negative transfer at the level of idiomatic usage within subordinate structures. In the newer corpus, fewer such cases were observed, suggesting learners have improved their knowledge of English-specific collocations and phrasing, reducing reliance on L1 translation equivalents. Again, this can likely be attributed to increased exposure to authentic English and perhaps explicit teaching of common collocations.

While the overall trend in error patterns is one of marked improvement, the persistence of certain errors provides important insights into the remaining challenges for French-speaking learners. Primarily, as mentioned, relative clause errors (especially with *wh*-relatives) and nominal clause errors did not decrease as much proportionally. In fact, our residual error analysis indicated that those categories (*wh*-relatives and *wh*-nominals) saw minimal change or even a slight increase in errors. The fact that errors here are present suggests that full proficiency of such structures may require more than what typical instructional approaches have so far achieved. It underscores a point from interlanguage theory: some errors are developmental and may be resolved only at very high proficiency or never completely for some learners. The minimal improvement in *wh*-relative

clause accuracy (error rate dropping only slightly) could imply that although learners now use relative clauses a bit more frequently and confidently, they still occasionally mis-form them. Similarly, the nominal *wh*-clause errors increasing in count might indicate that more learners are attempting those clauses (as we saw, their frequency rose) but not all attempts are successful. They may struggle with word order in indirect questions or with the use of introductory "what/how" structures, leading to errors.

Connecting these findings to theoretical perspectives, we can see the improvement in accuracy as evidence of learners' interlanguage moving closer to the target norm, which is a process of approximation and refinement that SLA models describe. Error analysis, as Corder (1981) famously argued, is a window into learners' internal grammar and its evolution. Early in the learning process, frequent systematic errors signal which rules have not been acquired or where L1 influence is strong. Over time, changes in error frequency signal learning gains. In our case, the drastic reduction in errors like gerund/infinitive confusion and subordinator misuse indicates that learners have likely acquired the relevant rules (e.g., knowing common verb patterns and relative clause structures) much better now than 30 years ago. It highlights the role of explicit instruction and feedback: many of the errors that decreased are things typically covered in advanced grammar lessons. The data suggest that learners are either receiving better instruction on these or at least benefiting from repeated exposure and practice (or better resources), leading to fewer mistakes.

From a Processability Theory viewpoint, as learners reach higher developmental stages, we expect them to handle more complex agreement, embedding, and syntactic integrations correctly. The error reduction in things like subject-verb agreement in subordinates and in properly linking subordinate clauses to main clauses (as seen by fewer sentence fragments) is consistent with learners reaching those stages of processing where they can manage such complexities. Pienemann's hierarchy places the correct handling of subordinate clauses toward the top (later stages). The improvements we observe diachronically suggest that a greater proportion of learners in the recent cohort have effectively reached that stage, which is evidenced by their ability to produce subordinate constructions with fewer mistakes. The interlanguage grammar of the average advanced French learner in the ICLE+30-FR is more developed (with respect to subordination) than that of the average advanced learner in the ICLE-FR. This is a remarkable finding, as it empirically demonstrates collective interlanguage development across generations, an idea that supports diachronic learner corpus research. It shows that what might have been "difficult grammar" for an earlier generation can become more "routine grammar" for a later generation, given improvements in the learning environment.

The influence of L1 transfer remains visible in the pattern of persistent errors. Many of the errors that remain (relative pronoun confusion, certain tense/modal issues, some calques) have roots in L1-L2 differences. That these are the last to be resolved (or still unresolved) aligns with transfer theory: when a structure in the target language does not have a clear or easy equivalent in the L1, learners often have protracted difficulties with it (Lambert et al., 2003; Odlin, 1989, cited in Jarvis, 2000)). Our data imply that while such errors have become less frequent, they have not vanished entirely, meaning the underlying L1 influence is still exerting pressure. The discussion of Lambert et al. (2003) also reminds us that French and English differ in how they structure information (e.g., French often requires a subordinate clause where English might not, such as cleft sentences). Such transfer-driven constructions can lead to awkwardness or minor errors in English writing. The decrease in their frequency in our learner corpus indicates a positive trend: learners are unlearning some L1 habits or at least distinguishing when not to apply French patterns to English. Yet, the areas where errors persist mark the boundary of interlanguage development where L1 influence is hardest to eradicate. Even at advanced levels, some interference remains, which is unsurprising from a theoretical perspective: complete nativeness is rarely achieved by adult learners, and traces of L1 in grammar choices or mistakes are often the last remnant of the learning process.

Finally, it is important to highlight the implications of these diachronic error pattern shifts for language teaching and our understanding of learner progress. The significant drop in error rate is an encouraging sign that pedagogical interventions over the last decades have had a positive effect, but it can also be discouraging if learners increasingly refer to external resources.. The persistent difficulties, however, draw attention to what should be the focus of future teaching efforts: nuanced usage of relative clauses, proficiency of subtler tense/modal distinctions, and prevention of residual L1-influenced formulations. It might be that these aspects require more than the traditional approaches (perhaps more immersive exposure or targeted advanced grammar workshops). Additionally, the possibility of avoidance as a factor in error reduction (e.g., fewer errors partly because of fewer attempts at troublesome structures) suggests that we should encourage learners to use the full range of complex structures and support them in doing so accurately, rather than letting them shy away. Otherwise, the true state of their ability might be obscured. In our data, for instance, the disappearance of errors in participial clauses could mean learners just didn't try them; ideally, we want learners who can both attempt and execute such structures correctly.

In summary, the third research question can be answered with a clear finding: the most common error patterns in subordinate clause usage have improved markedly in frequency, pointing to a diachronic enhancement of grammatical accuracy, though a few error types persist as

challenges. French-speaking learners of English in the ICLE+30-FR era make significantly fewer subordination-related mistakes than those in the ICLE-FR era, reflecting better proficiency of complex syntax. This improvement validates the hypothesis of increased accuracy and demonstrates tangible progress in learner language over time. The persistent errors, on the other hand, provide valuable diagnostic information.

5.4. Limitations

While this study offers new insights into the diachronic analysis of learner errors among French-speaking EFL students, several limitations must be acknowledged in order to contextualize the findings and inform future research directions.

Although both the ICLE-FR and ICLE+30-FR share a similar structure, they are not identical in every respect. The range of essay prompts, the contexts in which the texts were written (argumentative purpose), and the specific academic environments may differ, potentially introducing bias. Some topics may elicit particular grammatical constructions (e.g., conditional forms or passive voice), which could in turn affect the frequency or type of errors produced. This variability challenges the comparability of the two datasets and should be kept in mind when interpreting diachronic trends.

The number of texts analyzed (30 per corpus) remains relatively limited. This constrains the generalizability of the findings, particularly when it comes to less frequent or context-dependent errors. Moreover, the data exclusively involve university students, which limits the applicability of the results to this specific learner profile. Other populations, such as secondary school learners or adult language learners, are not represented in this study.

Another major limitation is that, while the coding and annotation process was designed to be systematic and thorough, it was conducted by a single researcher. Although intra-rater reliability was assessed and found to be substantial, the absence of inter-rater validation introduces a degree of subjectivity, particularly in error annotation and clause classification.

Building on the present study, future research could pursue several complementary directions. First, a larger and more diversified dataset would enhance the reliability of diachronic comparisons, particularly if it includes learners from varying educational backgrounds, age groups, and proficiency levels. Second, a mixed-methods approach combining quantitative corpus analysis with qualitative data (such as learner interviews, writing process observations, or teacher feedback) could help to understand the causes and evolution of specific error types. Third, future studies might explore the pedagogical implications of recurrent error patterns, examining how

instructional practices, textbooks, or assessment standards have shaped learner output over time. Finally, in light of the growing use of AI-powered tools in language learning, further research could investigate how such technologies influence both the frequency and nature of learner errors, and whether they contribute to the fossilization or remediation of persistent interlanguage features.

6. Conclusion

This thesis has investigated diachronic changes in the use of subordinate clauses by French-speaking learners of English, based on a comparative analysis of the ICLE-FR and ICLE+30-FR corpora. Through a mixed-methods approach combining quantitative frequency analysis with qualitative error examination, the study has shed light on how syntactic complexity in learner writing has evolved over the past thirty years.

The findings largely support the initial hypotheses. Firstly, the ICLE+30-FR displays a significantly higher average number of subordinate clauses per sentence compared to ICLE-FR, indicating a general increase in syntactic complexity over time. This result corresponds to Gilquin's (2022) call for diachronic learner corpus research to document collective changes in learner interlanguage. The more frequent use of subordination suggests a growing proficiency of complex syntactic structures in argumentative writing.

Secondly, changes were also observed in the distribution of clause types. While nominal clauses remain the most frequent, the use of adverbial and relative clauses increased, along with a wider variety of non-finite structures such as infinitive and gerund clauses. These shifts align with previous research on syntactic evolution in native English (e.g., Malà, 2017; AL Fajri & Okwar, 2020), and suggest that learners may be increasingly exposed to more flexible and informal syntactic models through changing pedagogical practices or contact with global English varieties.

In terms of accuracy, the overall number of grammatical errors in subordinate clauses has decreased in the more recent corpus, although certain difficulties persist. These include confusion between gerunds and infinitives, omission of explicit subjects in infinitival constructions and occasional L1 transfer effects. Such patterns are consistent with the error classifications proposed by Corder (1981) and Fitikides (1963), and reflect ongoing challenges for French-speaking learners due to cross-linguistic differences (Lambert et al., 2003).

Theoretically, the findings support the idea that interlanguage develops along a predictable trajectory of increasing syntactic complexity, as described by Pienemann's Processability Theory and Ortega (2015). Subordinate clauses, as key indicators of advanced syntactic development (Berman and Verhoeven, 2002), provide a valuable lens through which to assess long-term language development.

Overall, this thesis contributes to filling a gap in SLA research by adopting a diachronic perspective, still underexplored in learner corpus studies. It enables future research involving other L1 backgrounds, genres, or syntactic features. Further investigations into the sociolinguistic

variables influencing these diachronic changes would also enhance our understanding of how learners acquire complex grammatical structures over time.

In conclusion, this study highlights both the increasing syntactic sophistication and the persistent difficulties of French-speaking learners in their use of subordinate clauses in English. It confirms the value of diachronic learner corpus analysis for tracking interlanguage development, and offers practical insights for improving the teaching of complex syntax in the EFL classroom.

7. Bibliography

Fajri, M. S. A., & Okwar, V. (2020). Exploring a diachronic change in the use of English relative clauses: A Corpus-Based Study and its implication for pedagogy. *SAGE Open*, *10*(4), 215824402097502. https://doi.org/10.1177/2158244020975027

Berman, R. A., & Verhoeven, L. (2002b). Cross-linguistic perspectives on the development of text-production abilities. *Written Language & Literacy*, *5*(1), 1–43. https://doi.org/10.1075/wll.5.1.02ber

Biber, D., & Gray, B. (2016b). *Grammatical complexity in Academic English: Linguistic change in writing*. https://ci.nii.ac.jp/ncid/BB2131585X

Chen, X., Alexopoulou, T., & Tsimpli, I. (2020). Automatic extraction of subordinate clauses and its application in second language acquisition research. *Behavior Research Methods*, *53*(2), 803–817. https://doi.org/10.3758/s13428-020-01456-7

Corder, S. P. (1981). Error analysis and interlanguage (Vol. 198, No. 1). London: Oxford university press.

Cornu, E., Kübler, N., Bodmer, F., Grosjean, F., Grosjean, L., Léwy, N., Tschichold, C., & Tschumi, C. (1996b). Prototype of a second language writing tool for French speakers writing in English. *Natural Language Engineering*, *2*(3), 211–228. https://doi.org/10.1017/s1351324997001368

Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment (CEFR).

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, *26*, 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Dickins, J. (2006). Disjuncts. *WORD*, *57*(1), 27–70. https://doi.org/10.1080/00437956.2006.11432558

Egan, T. (2008). Non-finite complementation. https://doi.org/10.1163/9789401205542

Fitikides, T. J. (1963). Common mistakes in English (5th ed). London: Longman.

Gilquin, G. (2022). Diachronic learner corpus research. In Studies in corpus linguistics (pp. 40–67). https://doi.org/10.1075/scl.105.02gil

Gries, S. T. (2013). Statistics for Linguistics with R. In *De Gruyter eBooks*. https://doi.org/10.1515/9783110307474

Jarvis, S. (2000). Methodological Rigor in the Study of Transfer: Identifying L1 Influence in them Interlanguage Lexicon. *Language Learning*, *50*(2), 245–309. https://doi.org/10.1111/0023-8333.00118

Kalnbērziņa, V., & Rūtenberga, V. (2012). Subordinate clauses as critical features in English and French learner Examination Corpora. *Baltic Journal of English Language Literature and Culture*, *2*, 54–62. https://doi.org/10.22364/bjellc.02.2012.05

Lambert, M., Carroll, M., & Von Stutterheim, C. (2003). La subordination dans les recits d'apprenants avances francophones et germanphones de l'anglais. *Acquisition Et Interaction En Langue Étrangère*, *19*, 41–69. https://doi.org/10.4000/aile.1138

Lambrecht, K. (2010). Constraints on subject-focus mapping in French and English. In *Linguistik aktuell* (pp. 77–100). https://doi.org/10.1075/la.165.04lam

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Langtest. (n.d.). Langtest – Statistical tools for linguistics. Retrieved June 1, 2025, from https://langtest.jp

Li, S., Hiver, P., & Papi, M. (2022). The Routledge Handbook of Second Language Acquisition and Individual Differences. In *Routledge eBooks*. https://doi.org/10.4324/9781003270546

Malá, M. (2017). A Corpus-Based diachronic study of a change in the use of Non-Finite clauses in written English. *Prague Journal of English Studies*, *6*(1), 151–166. https://doi.org/10.1515/pjes-2017-0009

AL-Takhayinh, A. H., Al-Dala'ien, O. A., & Mudhsh, B. A. D. (2015). Second Language Acquisition: An Overview and Theoretical Background. URL: https://bit. ly/2RNIC4H (Visited on 13/03/2016).

Novogradec, M. J. (2021). Positive and negative lexical transfer in English vocabulary acquisition. *ELOPE English Language Overseas Perspectives and Enquiries*, *18*(2), 139–165. https://doi.org/10.4312/elope.18.2.139-165

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, *29*, 82–94. https://doi.org/10.1016/j.jslw.2015.06.008

Ortega, L. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. *Applied Linguistics*, *24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Quirk, R., & Greenbaum, S. (1973). *A university Grammar of English*. Longman.

Quirk, R., Sidney Greenbaum, Leech, G., Svartvik, J., & Longman Group UK Limited. (1972). *A Grammar of Contemporary English* (Ninth impression (corrected) 1980) [Book]. Longman Group UK Limited. https://hostnezt.com/cssfiles/english/A%20Grammar%20of%20Contemporary%20English%20by%20Quirk,%20R%20.pdf

Ringbom, H. (1978). the influence of the mother tongue on the translation of lexical items. *Interlanguage Studies Bulletin*, *3*(1), 80–101. http://www.jstor.org/stable/43135191

Roth, I. (2010). Explore the influence of French on English. In INNERVATE Leading Undergraduate Work in English Studies, *INNERVATE Leading Undergraduate Work in English Studies* (Vols. 3–3, pp. 255–262).

Smaby, R. M. (1974). Subordinate clauses and asymmetry in English. *Journal of Linguistics*, *10*(2), 235–269. https://doi.org/10.1017/s0022226700006083

Yamamoto, H., Tanaka, M., & Kondo, Y. (2012). Diachronic Corpus and Linguistic Space: New Methods for the Analysis of Language Change. *2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, 381–384. https://doi.org/10.1109/snpd.2012.104

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté de philosophie, arts et lettres

Place Cardinal Mercier, 31, 1348 Louvain-la-Neuve, Belgique I **www.uclouvain.be/fial**