Degree Project
Level: bachelor’s

Writer/Reader Visibility in EFL Writing

A Corpus-based Analysis of Young Swedish Students' Writing Development

Author: Francis Mwangi
Supervisor: Joseph Lee
Examiner: Annelie Ådel
Subject/main field of study: Linguistics
Course code: EN2035
Credits: 15
Date of examination: 10 January, 2023

At Dalarna University it is possible to publish the student thesis in full text in DiVA. The publishing is open access, which means the work will be freely accessible to read and download on the internet. This will significantly increase the dissemination and visibility of the student thesis.

Open access is becoming the standard route for spreading scientific and academic information on the internet. Dalarna University recommends that both researchers as well as students publish their work open access.

I give my/we give our consent for full text publishing (freely accessible on the internet, open access):

Yes ☑  No ☐
Abstract
This corpus-based study explores writer-reader visibility (WRV) features in the writing by young Swedish learners of English. Specifically, using Petch-Tyson’s (1998) framework, this study examines the use of WRV features in essays written by young Swedish learners in lower and upper secondary school, and compares their use to that of Swedish university-level learners. The findings align with Hasund and Hasselgård’s (2022) observation based on young learners that “the tendency to be visible writers starts early” (p. 19). Yet, as the writing proficiency or experience of this learner population increases, their visibility decreases by relying less on WRV devices. This study contributes to the understanding of how young learners’ use of WRV features in their writing develops and provides insights into writing instructions for young learners of English.

Keywords: Corpus Linguistics, Writer/Reader Visibility, English as a Foreign Language, Language Development, Second Language Writing
# Table of contents

1 Introduction ............................................................................................................................. 3  
  1.1 Aim of the Study ............................................................................................................ 5  
2. Theoretical Background ......................................................................................................... 6  
  2.1 Swedish Secondary School English Education Writing Goals ...................................... 6  
  2.2 Defining Writer/Reader Visibility .................................................................................. 8  
  2.3 Writer/reader Visibility in Specialist Writing .............................................................. 11  
  2.4 Writer/reader visibility in L2 student writing .............................................................. 13  
3. Corpora and Methodology ................................................................................................... 17  
  3.1 Corpus and Corpus Linguistics ..........................................................................................18  
  3.2 Description of the Corpora ........................................................................................... 18  
    3.2.1 Swedish Learner English Corpus (SLEC) ........................................................... 19  
    3.2.2 The Swedish Subcorpus of International Corpus of Learner English (SWICLE) .......................................................... 20  
    3.2.3 Comparability of the SLEC and SWICLE .......................................................... 21  
    3.2.4 Corpora Used in the Study .................................................................................. 22  
  3.3 Methods of Analysis ..................................................................................................... 24  
4 Results and Discussion .......................................................................................................... 29  
  4.1 Overall WRV Features ................................................................................................. 29  
  4.2 First- and second-person pronouns .............................................................................. 31  
  4.3 The Collocation I think ................................................................................................ 34  
  4.4 Modal Adverbs ............................................................................................................. 36  
  4.5 Modal Auxiliaries ......................................................................................................... 38  
  4.6 Questions ...................................................................................................................... 41  
5 Conclusion ............................................................................................................................. 42  
References ................................................................................................................................ 45
1 Introduction

Over the past decades, considerable attention has been devoted to understanding the presence of the writer and reader in academic texts under the rubric of writer/reader visibility (henceforth WRV). Petch-Tyson (1998) characterises WRV as the manner in which the involvement of both the reader and writer is encoded in the discourse. Numerous studies have analysed published research articles (RAs) to examine how RA writers integrate WRV features in their texts (e.g., Dahl, 2004; Hyland, 2018). Hyland (2018) found that, in the strategic use of WRV features in RAs, writers can establish an impression of authority, integrity, and credibility, allowing them to emphasise or diminish the visibility of their readers within the text. In her cross-cultural analysis of these features in RAs written in English, French, and Norwegian, Dahl (2004) found that while discipline is also highly relevant, such features are more closely related to language. While these studies have been valuable in understanding WRV features in professional writing, they offer little insight into how students, particularly second language (L2) students, utilise the features in academic writing.

Therefore, to gain greater insight into novice writers’ use of these features, researchers have analysed WRV elements particularly in L2 students’ academic writing. Ådel (2008), for example, examined three factors that influence the use of involvement features in writing: time, interaction with secondary sources and register awareness. Paquot et al. (2013) investigated the use of writer/reader visibility features in argumentative and discipline-specific texts written by French and Norwegian learners. The findings of these studies suggest that L2 learners, no matter their mother tongue, tend to use a personal style of writing that is exhibited through the use of linguistic features that make them more visible or involved in their texts (Ådel, 2008; Paquot et al., 2013). Writing in a personal style implies that students use linguistic features and devices that show their involvement or presence and that of their intended audience, or depending on
their communicative goals, they may choose to be less visible. The WRV phenomenon can be established through the writer’s use of features that are the focus of this study, namely: first- and second-person pronouns (singular and plural), the collocation *I think*, modal adverbs, modal auxiliaries, and questions. The extensive research on WRV in L2 student writing provides valuable insights into the challenges L2 learners experience in writing academic texts. By examining how L2 students use WRV features, these studies help educators and researchers better understand L2 students’ challenges and how to support their development as writers in a second language.

However, despite the insights provided by research on L2 students’ use of WRV features, it is important to note that most WRV studies have focused on adult university students. As Thomson (2018) observes, there is little research on young learners of English and their use of WRV features in their academic writing, or the development of these elements in their writing. The term “young learner” in the English Language Teaching (ELT) profession generally refers to individuals under 18, although interpretations and usage can vary (Ellis, 2013). However, Ellis (2013) emphasises that the term is often used vaguely, lacking specificity in age range and leading to confusion. Additionally, in the context of teaching English as a foreign language (EFL), the term “young learner” can encompass both age and prior language experience (Sehan, 2018). While a 15-year-old student with no prior English exposure might be considered a young learner, Sehan (2018) points out that there is no universal consensus among experts on this definition. As a result, the age classification for EFL young learners is applied, both for those learning English as a first language (EFL) and those learning it as a foreign language. Considering the little research on young learners of English, one notable exception is Hasund and Hasselgård’s (2022) study of WRV features in the writing of young Norwegian learners of English. Yet, as they point out, it is unclear whether young Norwegian learners’ use is culture-
specific and unique to this specific learner group. Thus, they underscore the need to examine young learners from different first languages. Yet again, there is still a significant gap in our understanding of how young Swedish L2 learners use WRV features in their writing and whether their use changes over time.

1.1 Aim of the Study

Building on Hasund and Hasselgård (2022), this corpus-based study investigates the developmental characteristics of the use of WRV features in young L1 Swedish English learners’ writing. More precisely, it compares the presence of these features in argumentative essays written by Swedish students at the lower and upper secondary school levels. Additionally, the study draws comparisons between the WRV features employed in these students’ essays and those found in the writings of Swedish university-level learners. The research was guided by the following questions:

1. What are the frequencies of writer/reader visibility features in L1 Swedish lower and upper secondary EFL students’ argumentative essays?

2. To what extent do L1 Swedish lower and upper secondary EFL students’ argumentative essays differ in their use of writer/reader visibility features?

3. To what extent do L1 Swedish lower and upper secondary EFL students’ use of writer/reader visibility features differ from advanced adult Swedish learners?

Through this research, understanding the variation in WRV features in young learners’ writing may not only contribute to the understanding of language development but also lead to more effective teaching and support strategies. Additionally, understanding the changes in WRV features can help teachers provide age-appropriate support in young learners’ writing as well as instructions to help them develop their writing skills (Fulcher, 2014; Nunan, 2003).
Furthermore, as Lee et al. (2019) point out, it could “help students develop appropriate register awareness” at various stages in their writing development (p. 152).

2. Theoretical Background

In this section, concepts relevant to this study and previous research on WRV features are presented. Section 2.1 explains English education in Sweden for lower and upper secondary schools, while section 2.2 defines writer/reader visibility (WRV). Section 2.3 synthesises WRV in specialist writing, and, lastly, section 2.3 presents research on writer/reader visibility in L2 student writing.

2.1 Swedish Secondary School English Education Writing Goals

According to the Swedish National Agency for Education, secondary school English language writing goals align with the Common European Framework of Reference for Languages (CEFR) levels (Skolverket, 2023). The framework provides a standardised way of describing proficiency across different languages, from A1 (beginner) to C2 (proficient) (Council of Europe, 2023). In Swedish secondary schools, the teaching approach involves primarily conducting classes in English, exposing students to diverse forms of written and spoken English. Students are encouraged to connect the content with their personal experiences, hence encouraging interaction in both speech and writing. The emphasis is on developing comprehensive communicative skills, encompassing the understanding of spoken language and texts, as well as the expression and interaction in speech and writing. Additionally, correctness in language use in both spoken and written forms is a key aspect of students’ development (Skolverket, 2023).
Specifically for lower secondary students, according to the Swedish National Agency for Education, it is anticipated that by the conclusion of Year 9, students should be able to express themselves in a straightforward, comprehensible, and relatively coherent manner to meet Grade E criteria. Grade E indicates that a student has fulfilled all components of the knowledge requirements for Grade E (Skolverket, 2023). The grading system in Swedish lower secondary school is a six-grade scale from A to F. Grades A to E signify passing grades, with F denoting a failing grade, and A being the highest achievable grade. To enrich and diversify their communication skills, students are encouraged to refine and make simple enhancements to their expressions. In written interactions across various contexts, students are expected to articulate their thoughts simply and understandably. To some extent, they should also adapt their communication based on the purpose, audience, and situation. Moreover, students are expected to demonstrate the ability to choose and apply functional strategies that address issues and enhance their interactions (Skolverket, 2023).

According to CEFR, the minimum pass grade for Swedish upper secondary school English language writing is at the B2 level. Students are expected to demonstrate the ability to write clear and detailed texts across diverse subjects, organising their thoughts logically and coherently. They should employ a variety of vocabulary and grammatical structures, adopt an appropriate style for the intended audience, and effectively correct their own writing (Council of Europe, 2023). The Swedish National Agency for Education has defined specific learning outcomes for upper-secondary English at different proficiency levels. In English 5, students are expected to comprehend the main content and basic details of written English in diverse genres. Moving to English 6, the expectations include understanding the main content and essential details in written English across various genres and in more formal contexts. In the advanced level of English 7, students are challenged to grasp the main content and essential details and
even implied meanings, of written English across various genres of an advanced nature. The grading system ranges from E (basic understanding) to A (advanced understanding), with specific criteria for each grade level (Skolverket, 2023).

2.2 Defining Writer/Reader Visibility

Writer/reader visibility (WRV) features are linguistic devices that indicate the presence of the writer and/or reader in the text. As Petch-Tyson (1998) explains, these features signal that the “participation of reader and writer are coded in the discourse” (p. 108). They can be used to create a sense of engagement and interaction between the writer and the reader. From a collective approach of metadiscourse, Ädel (2006) states that WRV features reveal the writer’s and reader’s (or speaker’s and hearer’s) presence in the text, either by referring to the organisation of the text or by commenting on the text in other ways. Metadiscourse, a term extensively employed in contemporary discourse analysis, pragmatics, and language teaching, pertains to the strategies writers and speakers employ to interact with readers and listeners through their use of language (Hyland, 2017). Additionally, Hyland (2018) mentions that WRV serves to indicate the writer’s acknowledgement of a reader.

The extent to which writers and/or readers are made visible in texts can include the use of linguistic devices. For instance, a more visible writer may use first-person pronouns I and we and express individual opinions, hence signifying personal involvement and perspective:

(1) However there are things I think isn’t that useful. Some people love shopping, however I think it’s mostly a waste of money. (SLEC\(^1\) H_8_M_22_111)

In this example the student writer indicates personal involvement and perspective by using the first-person I. Additionally, the student expresses a contrasting opinion on shopping, stating

\(^1\) SLEC stands for the Swedish English Learner Corpus.
that, despite some people enjoying it, they believe it is mostly a waste of money. This contributes to WRV, as it provides insight into the student’s personal stance on the topic. By contrast, as Palmer (2001) notes, a less visible writer may adopt a more formal and impersonal tone by using a passive voice. Additionally, when writers intend to make themselves less visible, it may be done by using, for instance, modality. Specifically, Biber and Conrad (2009) define modals as “grammatical words that express the speaker’s or writer’s attitude or stance towards the proposition that they are making” (p. 257). Example 2 illustrates the use of modals:

(2) Settling down, staying in one place, might mean that they would have to face up to their responsibilities, to face their own life and try to do something about it. (SWICLE\textsuperscript{2} SWUL8017)

Here, modal adverbs and auxiliaries show uncertainty or soften their statements (Palmer, 2001). The phrase \textit{might mean} suggest a potential outcome but does not assert it with absolute certainty. The use of modal auxiliary \textit{might} implies that the student leaves room for alternative scenarios, hence softening the writer’s imposition on the reader’s thoughts.

To further illustrate the use of modal adverbs, an upper secondary student from the SLEC writes,

(3) Now you’re probably thinking, “If you do not get paid the amount that you need what do you do? (SLEC G\_1\_S\_F\_21\_133)

The Modal adverb \textit{probably} introduces a sense of uncertainty or likelihood, influencing the reader’s perception according to the modal degree of certainty (Biber & Conrad, 2009). The phrase \textit{Now you’re probably thinking} suggests that the reader likely has a particular thought or question in mind. The modal adverb \textit{probably} conveys a degree of likelihood rather than absolute certainty, acknowledging the potential variability in the reader’s thoughts.

---

\textsuperscript{2} SWICLE refers to the subcorpus Swedish International Corpus of Learner English.
Additionally, WRV can be expressed by direct or indirect forms of writing, whereby writers can choose to either explicitly state their points or use indirect communication strategies like rhetorical questions to make their arguments less obvious (Thompson, 2001). For example, in a sentence from the SLEC, we find the following question:

(4) If you do not get paid the amount that you need what do you do? (SLEC G_1_S_F_21)

The student in this example prompts the reader to consider the potential challenges or issues associated with not receiving the necessary payment. The rhetorical question serves as a subtle way to convey the argument without explicitly stating it. By presenting the question, the writer implicitly suggests that not getting paid the required amount could pose a problem, encouraging the reader to reflect on possible responses or solutions. This indirect approach allows the writer to guide the reader’s thoughts and considerations without imposing a clear and direct stance, fostering a more reflective and open-ended engagement with the topic. Consequently, writers can use such rhetorical questions to reinforce their point (Hyland, 2005). However, it should be noted that the current study examined all questions in the corpus, not only rhetorical questions.

As for the reader, including you (or other second-person pronouns) directly addresses the reader, thus increasing reader involvement and interaction. Writers may explicitly address the reader using phrases like a student in the SLEC:

(5) I want to tell you how I think you can have a good life. (SLEC H_9_F_22_13)

The pronoun you enhances a sense of engagement with the reader.

In summary, WRV features are important for effective communication, as they allow the participating writers to tailor their writing to the needs and expectations of their readers (Petch-Tyson, 1998).
2.3 Writer/reader Visibility in Specialist Writing

Extensive research has been devoted to examining WRV features in specialist texts to gain a deeper understanding of how professionals in various fields engage their readers and convey their ideas effectively (Bhatia, 1995; Hyland, 2018; Swales, 1992). As Hyland (2018) contends, by establishing an impression of authority, integrity, and credibility through the use of WRV features, “writers are able to either highlight or downplay the presence of their readers in the text” (p.63). This implies that, when writers are transparent about their presence in the text, it might build credibility and trust. By acknowledging their role, expertise, or biases in their texts, writers may establish authenticity and reliability, which are crucial elements in professional writing.

To understand WRV features in specialist writing, Shaw and Vassileva (2009) studied how academic rhetoric has evolved across four European countries: Britain, Bulgaria, Denmark, and Germany, over the course of the twentieth century. They found that while there is a general trend of decreasing first-person pronoun usage in most languages, the opposite is true for British English. This suggests that British English may be evolving in a different direction than other languages. However, the corpus of academic articles is limited; the authors only analysed a corpus of 91 articles, which is not representative of all academic writing in the four cultures that were studied. Additionally, the corpus is not balanced across the four cultures. It contains more articles from Britain than from any of the other three cultures. This may make it difficult to generalise the relative development of academic rhetoric in each culture. In a study investigating academic writing of four disciplines, Harwood (2005) explores the use of personal pronouns I and inclusive and exclusive we. Harwood explains that inclusive we encompasses both the writer and the reader, whereas exclusive we pertains exclusively to the writer and
individuals associated with the writer. The study found that inclusive pronouns serve various purposes, including describing or critiquing common disciplinary practices and creating a positively polite tone of solidarity. Additionally, inclusive pronouns served to aid in organizing the text, highlighting the writer’s claims and findings, outlining the paper’s structure, and signalling current issues within the discipline. Dahl (2004) examined the use of textual metadiscourse in research articles written in English, French, and Norwegian from the disciplines of economics, linguistics, and medicine. Dahl’s study, which Ädel (2022) finds to show a reverse pattern, found that articles in English and Norwegian were found to use more textual metadiscourse than articles in French, and this difference was consistent across all three disciplines (Dahl, 2004).

In a more recent study, Ädel (2022) investigates the variation in WRV features in humanities research articles across different languages, regional varieties, and disciplines. The study found that WRV features varied across disciplines and were most used in Linguistics. Additionally, there was a difference in first-person pronoun usage between English and Swedish academic writing, whereby “the Swedish academics used writer visibility through ‘I’ orientation to a greater extent than their English counterparts, making it abundantly clear that an ‘I’ ban does not apply to academic Swedish” (Ädel, 2022, p. 59). Authors of American English articles used less we compared to the British English articles that exhibited a clear preference for we, mirroring the pattern observed in Swedish writing. Furthermore, the use of WRV features was found to be closely tied to academic discipline, particularly in the field of Linguistics. However, first-person pronouns were found to be inconsistent indicators of genre in research articles. Ädel (2022) explains the inconsistency is a result of the substantial variation observed in the use of first-person pronouns across different academic fields, indicating that first-person pronouns are not consistently used in humanities research articles and thus cannot be considered reliable
indicators of genre or register. In contrast, second-person pronouns are consistently avoided in the articles. This suggests that making generalizations about the typical use or non-use of personal pronouns in the articles requires careful consideration of the specific context (Ådel, 2022).

In summary, these studies demonstrate how researchers use WRV features in their writing to enhance credibility, promote their work, and achieve academic recognition (Hyland, 2003a). While these studies have shed significant light on how WRV features are used in professional writing, they offer little insight into how L2 students employ these features in their writing.

2.4 Writer/reader visibility in L2 student writing
Considerable research has been conducted to better understand L2 students’ use of WRV features in their writing. Understanding how L2 students employ these features can provide insights into their writing proficiency and development. The model initiated by Petch-Tyson (1998) to study L2 learners’ use of WRV features used the International Corpus of Learner English (ICLE). She analysed the writing of L2 students from four language and cultural backgrounds: French, Dutch, Swedish, and Finnish. The study compared their argumentative essay writing to that of native English-speaking students to determine if there were any differences among the language/cultural groups in how they expressed the involvement of the writer and reader in the discourse. The study noted that learner writers exhibited a more noticeable presence in their discourse compared to native writers, attributing this to cultural and persuasive factors. Additionally, the learner writers consistently used WRV features more frequently than the native writers, emphasising interpersonal involvement and potentially resulting in their writing being seen as less conventional in the argumentative genre (Petch-Tyson, 1998, p. 116).
Building on Petch-Tyson’s work, several researchers have analysed WRV features in different L2 learners’ writing. For instance, Paquot et al.’s (2013) study on Norwegian and French learners’ argumentative texts found that the “overuse” (see section 3.3 for definition) of WRV features may be due to some factors, including transfer from the learners’ first language, a lack of awareness of the conventions of academic writing in English or a desire to engage the reader and make the text more personal and interesting. Additionally, this overuse may be due to such studies having a common idea of focusing on persuasive writing that is characterised by using evidence and logical reasoning to support the writer’s claims (Ädel, 2008; Aijmer, 2002; Paquot et al., 2013). This type of writing aims to persuade the reader to adopt a particular point of view (Hyland, 1994). As a result, “personal references and subjective attitudes are certainly hard to avoid” (Reeski, 2004, p. 3). Another study on L2 learners’ writing is McCrostie (2008), which investigated the degree of writer visibility in argumentative academic essays produced by Japanese university-level learners of English. The study involved a comparison between essays written by Japanese learners and those composed by native English speakers. The results revealed a noteworthy disparity, indicating that the essays from Japanese learners exhibited a significantly higher presence of WRV features compared to those of native English speakers. This implies that Japanese learners of English tend to be more overtly visible as writers in their academic writing, in contrast to their native English-speaking counterparts. However, the study compares Japanese learner writing to that of native English speakers without specifying the linguistic background or variety of English of the latter. Different varieties of English may promote varied conventions for WRV, raising questions about the applicability of the findings to a broader context.
Specifically focusing on Swedish learners, researchers have compared these L2 students’ writing and native English-speaking students’ texts. For example, Ädel (2008) investigated argumentative essays from Swedish advanced learners of English and compared them with those of native speakers of British English and American English. The learner writers demonstrated a high degree of visibility, positioning themselves at the extreme end of the writer-reader orientation scale, fostering an intimate and informal connection with their readers. In contrast, Ädel found that American writers fell in the middle, and British writers leaned towards extreme impersonality. In addition, learner writers were found to actively engage their readers, involving them more intensely in the discourse compared to native speakers. However, the study found that British writers maintained a more detached stance, with limited visibility, while still moderately acknowledging readers. Similarly, Aijmer (2002) investigated the use of modality in the essays of advanced Swedish learners and compared them to essays written by native English speakers. The study found that Swedish learners of English adapt their writing style from the Swedish language, which exhibits the overuse of WRV features (p.72). However, there were some differences in the use of modality between the learners and native speakers. For example, learners tend to use more deontic modals (which express obligation or permission) and fewer epistemic modals (which express certainty or doubt). Learners also tend to use modal expressions less accurately than native speakers.

In addition to the cultural factors that influence the style of writing (Paquot et al., 2013), the setting of tasks to write persuasive texts, timing and access to secondary sources may also influence the use of WRV features (Ädel, 2008). The setting of tasks to write argumentative texts has revealed that L2 learners tend to overuse WRV features (Hasselgård, 2009; Virtanen, 1998) These studies were conducted to gain a better understanding of how L2 students used
WRV features in their writing. Examining how L2 students employed these features has provided insights into their language acquisition and proficiency.

However, the majority of studies on L2 student writing have focused on adult university student writing. Little research exists on young L2 student writing and whether their behaviour changes over time (Hong & Cao 2014). Hong and Cao (2014) examined the use of interactional metadiscourse in descriptive and argumentative English essays by young L1 Chinese, Spanish, and Polish learners of English. The study reveals notable differences in the use of boosters, attitude markers, self-mentions, and engagement markers among the three groups of L2 learners. According to Hong and Cao (2014), hedges serve the purpose of lessening or softening the strength of propositions (e.g., *could*). In contrast, boosters function to heighten or magnify propositional strength (e.g., *absolutely*). Self-mentions explicitly indicate the writer's presence in a text (e.g., *I, me*). Engagement markers are employed to directly appeal to the reader or pose questions (e.g., *you*). The study also reveals that there were significant distinctions between the descriptive and argumentative essays in the utilisation of hedges and self-mentions. In addition to Hong and Cao (2014) exemplifying one of few studies on young learners, Hasund and Hasselgård’s (2022) examined WRV features in writing Norwegian lower secondary school learners of English (ages 13-16). Using data from the TRAWL (Tracking Written Learner Language) corpus, the study analysed how WRV features developed over time (from year 8 to year 10) and how their use differed between argumentative and expository writing. The findings show that young Norwegian learners overuse WRV features compared to advanced university students, but their use decreases over time as they become more proficient in English. Additionally, they found that the use of WRV features and the frequency of use varied depending on genre. As highlighted by Recski (2004), argumentative essays found in the International Corpus of Learner English (ICLE) pose a challenge when it comes to steering
clear of personal references and subjective attitudes. This difficulty arises because learners are explicitly encouraged to take a persuasive approach to express their personal opinions (Paquot et al., 2013).

While Hasund and Hasselgård’s research on young Norwegian learners is insightful, the study only investigates lower secondary schools. Thus, it is unclear whether, and to what degree, L2 students’ use of WRV features in their writing changes from lower to upper secondary school. By investigating the changes in WRV features in young learners’ writing, we can not only gain insights into their writing development, but we can also identify areas young L2 students need support at various stages of their development.

The aim of this corpus-based study is to investigate the evolving usage of WRV features in the writing of young Swedish learners of English. To be more specific, it examines the occurrence of these features in argumentative essays composed by Swedish students in both lower and upper secondary schools. Furthermore, the study makes comparisons between the WRV features utilised in these students’ essays and those observed in the writings of university-level Swedish learners’ texts.

3. Corpora and Methodology

Sections 3.1 and 3.2 describe the corpus and corpus linguistics and description of the corpora respectively, while sections 3.2.1 and 3.2.2 respectively describe the Swedish Learner English Corpus (SLEC) and the Swedish subcorpus of the International Corpus of Learner English (SWICLE). Lastly, section 3.2.3 discusses the comparability of the SLEC and SWICLE, and section 3.2.4 describes the corpora used in the study.
3.1 Corpus and Corpus Linguistics

This study uses corpus linguistic methods to analyse WRV features in young L1 Swedish learners’ argumentative essays in English. A corpus is a relatively large, principled collection of naturally occurring spoken and/or written discourse stored electronically, and corpus linguistics is a computer-based approach to the examination of a corpus considered suitable for addressing specific research questions related to language (McEnery & Hardie, 2011). Corpus linguistics offers several benefits for linguistic research. It allows researchers to analyse large amounts of authentic texts, thus providing empirical evidence of how different linguistic features are used in real-world language (McEnery & Hardie, 2011). Additionally, to understand how language usage evolves and changes, corpus linguistics allows researchers to examine language variation across time, regions, and social contexts (Biber & Conrad, 2009; Leech et al., 2009). However, for comparisons of these types of variables to be valid, the corpus material that they are based on needs to be representative and comparable. This method enables the identification of patterns, regularities, and trends in language use, helping researchers uncover linguistic phenomena that may go unnoticed in smaller datasets (Sinclair & Carter, 2004). These benefits demonstrate the significance of corpus linguistics as a methodological approach that enhances the scope of linguistic research.

3.2 Description of the Corpora

The data used for this study consist of two sets of corpora. The primary corpus is the Swedish Learner English Corpus (SLEC) and, for comparative purposes, the Swedish subcorpus of the International Corpus of Learner English (SWICLE) was used (Granger et al., 2020)
3.2.1 Swedish Learner English Corpus (SLEC)

Kaatari et al. (2024) compiled the Swedish Learner English Corpus (SLEC), a corpus of argumentative texts written by Swedish upper and lower secondary school students. They add that, “SLEC provides rich metadata on the students’ background, making it possible to empirically study relations between the linguistic properties of student texts and various extralinguistic and learner variables” (p.2). SLEC is the first corpus of its kind to include detailed information about students’ extramural activities, or English-language activities that students engage in outside the classroom (e.g., gaming, reading, social media, watching, etc.).

The ages of the students range between 12–18 years old from school years 8-12: lower secondary from years 8-9 and upper secondary between years 10-12. The total number of writers/texts in the first release of the corpus is 1,098, consisting of 481,155 words with an average length of 438.2 words per text (p.3). Table 1 describes the SLEC.

Table 1. Description of the SLEC

<table>
<thead>
<tr>
<th>Type</th>
<th>Year</th>
<th>No. of Texts</th>
<th>%</th>
<th>Word Count</th>
<th>Mean Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>8</td>
<td>169</td>
<td>15.4</td>
<td>81,645</td>
<td>483.1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>192</td>
<td>17.5</td>
<td>86,300</td>
<td>449.5</td>
</tr>
<tr>
<td>Upper</td>
<td>10</td>
<td>421</td>
<td>38.3</td>
<td>166,084</td>
<td>394.5</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>254</td>
<td>23.1</td>
<td>112,908</td>
<td>444.5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>62</td>
<td>5.7</td>
<td>34,218</td>
<td>551.9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1,098</td>
<td>100</td>
<td>481,155</td>
<td>438.2</td>
</tr>
</tbody>
</table>

Kaatari et al. say that the SLEC encourages teachers to get involved in students writing. The SLEC developers asked teachers to dedicate one lesson to the writing task, and some chose to incorporate an additional preparatory lesson. Although the prescribed writing time was a minimum of 50 minutes, teachers were encouraged to extend it if necessary. The metadata included documents whether students had a preparatory lesson and recorded the writing time. All students used computers via a provided link for writing and were explicitly instructed not
to use any aids except for a spell checker (Kaatari et al., 2024). The collection of the corpus texts was undertaken by teachers who gathered texts within their respective classrooms. To enhance comparability across texts, the SLEC includes only argumentative texts centred on the same topic. The given prompt directed students to compose an argumentative text addressing the following question: “What are the most crucial aspects for leading a good life?” (p.12).

3.2.2 The Swedish Subcorpus of International Corpus of Learner English (SWICLE)

The International Corpus of Learner English (ICLE) is a computerised collection of essays written by upper-intermediate and advanced learners of English as a foreign language (Granger et al., 2020). Initially compiled in 2002, the corpus has since been updated twice, with the most recent version being released in 2020. The latest version, ICLEv3, contains over 5.5 million words from learners from 25 different mother tongue backgrounds (para. 2). Swedish is part of the 25 L1 subcorpora of the ICLE, named the SWICLE consisting of argumentative texts from advanced learners from Swedish universities who are approximately of the same age (20-30). According to Ädel (2008), only “a small part of the essays (35%) are untimed” (p. 47). The timed essays (65%) are part of an examination, while untimed are written at home. There is no specific time except for the writing conditions described as “the essays were written under strict time constraints” (p.44). Just under half of the essays (44%) utilised reference materials during the writing process.

Comprising 355 essays totalling 201,467 words (Granger et al., 2020), this corpus was used to compare young Swedish and adult Swedish learners of English. As for topics, the corpus featured conventional argumentative themes that provide scope for a more personalised viewpoint. The topics included such as Integration or assimilation?, The use and abuse of state media, In defence of popular culture, Unemployment and the future, Reflections on being a
Students who are in their second year of university studying English in Sweden have typically undergone nine years of English education in school before embarking on their university studies (Ädel, 2006). Ädel (2006) points out that Swedish learners’ English proficiency is assessed primarily through their ability to compose essays on diverse topics, many of which are unfamiliar to them prior to the timed writing session.

3.2.3 Comparability of the SLEC and SWICLE

Considering that the current study is a corpus-based comparison, it is important that both corpora have EFL argumentative texts. According to Granger et al., (2020) “the majority (94%) of the ICLE texts are argumentative essays” (p.13) which is the focus of this study. The SLEC entirely contains argumentative texts on the same topic for comparability. Kaatari et al. (2024) mention that the aim of creating SLEC was to “make SLEC comparable to widely used learner corpora such as ICLE and LOCNESS” (p.3).

To attain even better comparability, both corpora have a database containing every learner’s profile information (variables) that can be used to filter the texts. This allows users to sort the texts used in the analysis on common variables. For instance, considering both corpora are comprised of a number of variables that differ, such as age and proficiency, the current study limited the variables to only Swedish L1 and argumentative texts. However, it is worth noting that SLEC had a preparatory lesson before a 50-minute timed duration to write their texts (Kaatari et al., 2024). On the other hand, since the SWICLE has the smallest proportion of untimed essays, caution was taken when comparing it to SLEC. According to Ädel (2008), time...
setting can significantly impact involvement features in writing, especially first-person singular pronouns. Therefore, care was taken when comparing any similar features between both corpora.

Table 2 provides a description of the lower secondary and upper secondary schools (henceforth LS and US) and university (SWICLE) students.

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>US</th>
<th>SWICLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of texts</td>
<td>361</td>
<td>737</td>
<td>355</td>
</tr>
<tr>
<td>Total word count</td>
<td>167,945</td>
<td>313,210</td>
<td>201,467</td>
</tr>
<tr>
<td>Average word length</td>
<td>465.2</td>
<td>424.9</td>
<td>567.5</td>
</tr>
</tbody>
</table>

As Table 2 shows, the US corpus is the largest in terms of total word count, followed by SWICLE and LS. The SWICLE has the highest average word length, suggesting potentially more complex language use on average, considering it has the lowest number of writers. The US has the lowest average word length with the highest number of writers. The LS, despite the lowest total word count and being the assumed lowest proficient, also has a higher average word length compared to US.

3.2.4 Corpora Used in the Study

Considering the scope of this study, it was necessary to create manageable data. Randomly, 50 texts were chosen from each group: 50 lower secondary (LS), 50 upper secondary (US), and 50 SWICLE texts. After selecting 50 files for each group to be used in the study, LS had a total word count of 20,724 words, while US included 23,387 and SWICLE comprised of 28,771 tokens. To exclude texts that did not meet the required common variable, both the SLEC and SWICLE have an encoding system that enables manual selection or “filtering” texts (Granger et al., 2020, p. 55)
Certain difficulties were encountered when files were randomly selected from both corpora. An error message indicating file corruption was consistently received, necessitating the files to be re-selected. Despite this challenge, the file selection process was straightforward due to the coding utilized in both corpora. For instance, in the case of lower secondary, files with codes beginning with H8M and H9F were randomly picked, and for upper secondary, files with codes starting with G1M, G2F, and G3M were selected. The file codes ending with F & M are for identifying texts by female and male writers. Care was taken to ensure an equal representation of texts from each class level and gender. Table 3 presents the corpora used in the study.

Table 3. Description of Corpora Used in the Study

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>US</th>
<th>SWICLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of texts</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Total word count</td>
<td>20,724</td>
<td>23,387</td>
<td>28,771</td>
</tr>
<tr>
<td>Average word length</td>
<td>414</td>
<td>467</td>
<td>575</td>
</tr>
</tbody>
</table>

As table 3 shows, this study utilizes three corpora, namely LS, US, and SWICLE, each consisting of 50 randomly chosen texts. The SWICLE corpus stands out with the highest word count and average word length, indicating potentially more extensive and complex texts. In contrast, the LS corpus has the smallest word count and average word length. The US corpus falls in between, suggesting a moderate size and complexity.

3.3 Methods of Analysis

The current study adopted a mixed-methods approach, utilizing both quantitative and qualitative analysis. Qualitative methodologies entail description and analysis rather than, for instance, the enumeration of features. Conversely, quantitative approaches involve linguistic variables that are often measurable or countable (Wray & Bloomer, 2013). The quantitative
element focused on determining the frequency of WRV features within the corpus, while the qualitative component examined how WRV features were employed within the corpus. Considering the current study employed both methods alongside the terms “overuse” and “underuse” to explain the findings, it is important to define the terms. In a previous study that emphasised the need for qualitative analysis alongside quantitative methods, Ädel (2014) defines “overuse” and “underuse” in learner corpus research, specifically in the context of Contrastive Interlanguage Analysis (CIA). These terms (“overuse” and “underuse”) refer to instances where a linguistic phenomenon is more or less frequent, respectively, in learner writing compared to native-speaker writing. Going beyond mere errors, Ädel (2014) points out that the terms form part of a systematic approach to identify features of non-nativeness in learner language. Furthermore, the study highlights the contextual and rhetorical considerations involved in assessing “overuse” and “underuse”, stressing the broader implications of these linguistic patterns beyond frequency alone. However, it is important to note that while the current study did not involve native speakers, the terms are used to “tell us whether a linguistic phenomenon occurs more or less often in corpus X than in corpus Y” (Ädel, 2014, p. 69).

To analyse WRV features, Petch-Tyson’s (1988) framework was used because it is a widely-used framework, and it allows for comparisons with other studies. The WRV items analysed in the current study are the same as those investigated in Hasund and Hasselgård’s (2022) study. Table 4 lists the features that are analysed in the study.

Table 4. List of WRV features analysed

<table>
<thead>
<tr>
<th>Feature</th>
<th>Items</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>first/second-person pronouns</td>
<td><em>I, me, myself, my, we, us, ourselves, our, you, yourself, yourselves, your</em></td>
<td>Money makes <em>me</em> happy. (LS-H_8_M_22_106)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>I am young and healthy. (US-G_3_S_21_28)</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>As <em>I</em> stated before, <em>you</em> have to have a vision. (SWICLE-SWUL5010)</td>
</tr>
</tbody>
</table>
The collocation *I think* also that hobbies is an important factor. (LS-H_8_F_22_94)

US- *I think* you should learn to enjoy your own company. (US-G_1_Y_F_22_28)

However, *I think* that assimilation is the way to go. (SWICLE-SWUL2010)

### Modal Adverbs

- actually, certainly, completely, definitely, frankly, maybe, naturally, obviously, of course, perhaps, possibly, probably, really, totally, unfortunately

*Actually* money is not direct key to happiness. (LS-H_8_M_22_119)

*Maybe* you want to be an astronaut. (US-G_3_S_F_22_1)

*Perhaps*, it is impossible to separate these concepts. (SWICLE-SWUL4016)

### Modal Auxiliaries

- can, could, may, might, must, ought, shall, should, will, would

People *can* make you feel more secure. (LS-H_8_M_22_110)

Without music, life *would* be a lost cause. (US-G_2_Y_M_22_17)

This *should* not be seen as a result of prejudice. (SWICLE-SWUL7006)

### Questions

identified through question marks (?)

- what about what we feel inside? (LS-H_9_F_22_75)
- US- Do you live a good life? (US-G_2_Y_M_22_17)
- how did technology evolve? (SWICLE-SWUL4039)

To search for WRV features, AntConc (Anthony, 2014) version 4.2.0, a corpus analysis toolkit, was used. Each of the items in Table 4 was electronically searched. After every search, a manual examination of each example in textual contexts was done to ensure every item fit the WRV categories. For example, using the CASE function on AntConc, the lowercase letter “i” may be concordanced as a first-person pronoun when a student leaves a space when writing “in”.

Additionally, when the CASE function was used on the LS corpus, and a search for lowercase “i” was conducted, 114 hits were obtained, in contrast to 471 without the case function.

Therefore, the texts were manually examined to identify instances where *I* was not used as a pronoun using the KWIC (keyword in context) tool for exclusion, but no such instances were
found because all occurrences of lowercase “i” were used as pronouns. To demonstrate, this example is from LS:

(6) Now i think that health was important and that and also education. (SLEC H_8_M_22_125)

The lowercase letter “i” can be included or excluded by using the CASE function. In this case, it was included because it functions as a pronoun. However, while searching for lowercase “i” in the SWICLE, 7 cases were found using the initials “i.e.” and therefore excluded as pronoun use. For example, the use of “i.e” from SWICLE:

(7) This may be an instance of what some psychologists term “the original sin”, i.e. a pattern of behaviour which is repeated throughout the generations.(SWICLE SWUL8017)

The AntConc cluster tool was used to search for the collocation I think. This tool allows users to search for the frequencies of groups of words (clusters).

Considering that a modal auxiliary can be used both deontically and epistemically (Palmer, 2001), the KWIC tool and File View features on Antcon were used to determine their context usage. According to Palmer (2001), “deontic modality relates to obligation or permission emanating from an external source” (p.9). Epistemic modality pertains to the speaker’s stance regarding the truth of the statement conveyed in the expression (Palmer, 2001). This means that epistemic modality is concerned with the speaker’s degree of certainty or uncertainty about the proposition being expressed. For instance, the modal could can be used both deontically or epistemically. For example, from SLEC,

(8) If you have "left over" money you could use it for good purposes like donating it, to enjoy life. You could also save it. If you save your money you could use it for emergencies, or to travel (SLEC H_8_F_22_97)

In this example, the modal verb could is used deontically in the first instance:
(9) If you have "left over" money you could use it for good purposes. (SLEC H_8_F_22_97)

In this instance, the student is using could to express permission. They are saying that it is permissible to use leftover money for good purposes. This is a deontic use of could because it expresses a rule or obligation. In the second instance, the modal verb could is used epistemically:

(10) If you save your money, you could use it for emergencies. (SLEC H_8_F_22_97)

In this instance, the student is using could to express possibility. They are saying that there is a possibility that left-over money can be used for emergencies. This is an epistemic use of could because it expresses a degree of certainty or probability. In the third instance,

(11) If you save your money, you could use it to travel (SLEC H_8_F_22_97)

This is also an epistemic use of could. The speaker is expressing the possibility that left-over money can be used to travel. In the fourth instance,

(12) you could enjoy life (SLEC H_8_F_22_97)

The student is expressing the possibility that left-over money can be used to enjoy life. In all four instances of the use of the modal verb could, the speaker is making suggestions about how to use left-over money. However, in the first instance, the speaker is stating a rule or obligation. In the other instances, the speaker is expressing a degree of certainty or probability.

However, there were difficulties searching for question marks on Antconc whereby all REGEX concordances did not yield any hits despite a manual search that found the presence of question marks. Tania and Tama (2017) define Regular Expressions (regex) as a string processing tool, widely employed in various computer programming tasks, including the investigation of diverse text formats. It is a pattern that defines a sequence involving letters, digits, or punctuation marks. Alternatively, to search for question marks, a different corpus tool was used:
Sketch Engine. Sketch Engine is an online software application that enables the analysis of occurrences of a specific word or phrase within a corpus (Cambridge University Press, 2009). It provides a summary of word types and patterns associated with the search term. Unlike Antconc, Sketch Engine does concordance for question marks, by using the regular expression ‘\?’. 

After locating and analysing every WRV device within the three corpora, items were normalised to occurrences per 1,000 words (henceforth ptw) to determine any differences in the three corpora. Normalisation is a process that aims to enable precise comparisons between corpora that vary in size (Bestgen, 2019). Corpora can vary significantly in size, and without normalisation, larger corpora may dominate the analysis due to their size. Normalisation ensures that comparisons are fair by accounting for the influence of corpus size on frequency counts.

4 Results and Discussion

Section 4 presents and discusses the results. Section 4.1 provides overall WRV features while 4.2 focuses on first- and second-person pronouns. Additionally, section 4.3 discusses the collocation *I think* and sections 4.4 and 4.5 present modal verbs and modal auxiliaries, respectively. Lastly, section 4.6 discusses questions.

4.1 Overall WRV Features

Table 5 shows the distribution of the overall WRV features in the three corpora.
Table 5. Distribution of WRV Features in the Three Corpora

<table>
<thead>
<tr>
<th>Features</th>
<th>LS (ptw)</th>
<th>%</th>
<th>US (ptw)</th>
<th>%</th>
<th>SWICLE (ptw)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person pronouns</td>
<td>824 (34.89)</td>
<td>26.99</td>
<td>832 (35.54)</td>
<td>29.38</td>
<td>895 (31.01)</td>
<td>49.04</td>
</tr>
<tr>
<td>2nd person pronouns</td>
<td>1263 (60.92)</td>
<td>41.38</td>
<td>1176 (50.18)</td>
<td>41.53</td>
<td>128 (4.44)</td>
<td>7.01</td>
</tr>
<tr>
<td><em>I think</em></td>
<td>96 (4.63)</td>
<td>3.15</td>
<td>63 (2.69)</td>
<td>2.22</td>
<td>45 (1.56)</td>
<td>2.47</td>
</tr>
<tr>
<td>Modal adverbs</td>
<td>191 (5.38)</td>
<td>6.26</td>
<td>133 (2.93)</td>
<td>4.7</td>
<td>158 (3.85)</td>
<td>8.66</td>
</tr>
<tr>
<td>Modal auxiliaries</td>
<td>656 (31.58)</td>
<td>21.49</td>
<td>599 (25.58)</td>
<td>21.15</td>
<td>549 (18.99)</td>
<td>30.08</td>
</tr>
<tr>
<td>Questions</td>
<td>22 (1.06)</td>
<td>0.73</td>
<td>29 (1.24)</td>
<td>1.02</td>
<td>50 (1.73)</td>
<td>2.74</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3052 (138.46)</strong></td>
<td><strong>100</strong></td>
<td><strong>2832 (118.16)</strong></td>
<td><strong>100</strong></td>
<td><strong>1825 (61.58)</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

*Note: ptw = per 1,000 words*

Overall, as seen on the “total” row, LS has the highest frequencies of all features, while the SWICLE has the lowest. The US falls in between the two groups. This suggests that the LS is the most informal and interactive of the three corpora, while the SWICLE corpus is the most formal. As Paquot et al. (2013) observes, “Scandinavian writers favour a less formal, more interactive writing style” (p. 8). Yet, the results show that they appear to rely less on this “less formal, more interactive” expression as their writing develops.

Notably, LS and US show some similarities in the usage of first-person pronouns, yet the SWICLE contains fewer than both groups. Despite SWICLE writers’ overall lower frequencies, they often exhibit distinct patterns in the different feature categories. The patterns that make SWICLE stand out are the lower percentage of second-person pronouns (7.01) and a high percentage of first-person pronouns (49.04). The table shows that both LS and US surpass SWICLE writers in the use of all features, except in questions, which show marginal differences. Therefore, the SLEC corpus study suggests that Swedish learners depend more on
WRV features in their writing at earlier stages of development, and as their writing proficiency or experience increases, their reliance on WRV devices decreases.

The overall differences between the categories of features show LS dominating over the US across all features except for the questions and first-person pronouns, which show minimal differences. The LS overuse of WRV features is observed in Norwegian young learners by Hasund and Hasselgård (2022): “The tendency to be visible writers starts early” (p. 19). Additionally, in consideration of the 50-minute timed writing in SLEC, the use of more features in LS may be explained by Ådel’s (2008) explanation that learners write with less involvement when they are not given a time limit. With timing in mind, it is worth noting that the comparison with Hasund and Hasselgård’s study is flawed by the fact that the timing factor was not considered in their longitudinal collection of data. In comparison, similar to the Norwegian young learners (Hasund & Hasselgård, 2022), modal auxiliaries are the third most used feature, followed by modal adverbs, *I think* and questions respectively.

There is a substantial difference in the use of second-person pronouns. LS and US show higher percentages, while SWICLE has a significantly lower percentage, indicating a notable shift in expressing perspective with the advancement in the writers’ proficiency. The use of *I think* phrases decreases across the three corpora, with the lowest frequency in SWICLE. The frequency of modal adverbs is higher in LS compared to US, and SWICLE shows a moderate frequency. The LS has the highest frequency of modal auxiliaries, followed by US, and SWICLE shows the lowest frequency. Questions are utilised with a relatively low frequency across all three corpora, with SWICLE writers using them more frequently, followed by US and least LS.
4.2 First- and second-person pronouns

Table 5 presents the raw and normalised frequencies of first- and second-person pronouns in the three corpora.

Table 5. Distribution of First- and Second-Person Pronouns in the Three Corpora

<table>
<thead>
<tr>
<th>pronoun</th>
<th>LS (ptw)</th>
<th>%</th>
<th>US (ptw)</th>
<th>%</th>
<th>SWICLE (ptw)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>471 (22.72)</td>
<td>57.16</td>
<td>334 (14.28)</td>
<td>40.14</td>
<td>338 (11.7)</td>
<td>37.77</td>
</tr>
<tr>
<td>me</td>
<td>91 (4.39)</td>
<td>11.04</td>
<td>82 (3.50)</td>
<td>9.86</td>
<td>32 (1.10)</td>
<td>3.58</td>
</tr>
<tr>
<td>myself</td>
<td>3 (0.14)</td>
<td>0.36</td>
<td>7 (0.29)</td>
<td>0.84</td>
<td>13 (0.45)</td>
<td>1.45</td>
</tr>
<tr>
<td>my</td>
<td>112 (5.40)</td>
<td>13.6</td>
<td>112 (4.78)</td>
<td>13.46</td>
<td>69 (2.39)</td>
<td>7.71</td>
</tr>
<tr>
<td>we</td>
<td>80 (3.86)</td>
<td>9.70</td>
<td>164 (7.01)</td>
<td>19.71</td>
<td>254 (8.82)</td>
<td>28.38</td>
</tr>
<tr>
<td>us</td>
<td>28 (1.35)</td>
<td>3.4</td>
<td>48 (2.05)</td>
<td>5.77</td>
<td>77 (2.67)</td>
<td>8.60</td>
</tr>
<tr>
<td>ourselves</td>
<td>3 (0.14)</td>
<td>0.36</td>
<td>5 (0.21)</td>
<td>0.60</td>
<td>5 (0.17)</td>
<td>0.56</td>
</tr>
<tr>
<td>our</td>
<td>36 (1.73)</td>
<td>4.37</td>
<td>80 (3.42)</td>
<td>9.62</td>
<td>107 (3.71)</td>
<td>11.96</td>
</tr>
<tr>
<td>Total</td>
<td>824 (34.89)</td>
<td>100</td>
<td>832 (35.54)</td>
<td>100</td>
<td>895 (31.01)</td>
<td>100</td>
</tr>
<tr>
<td>2nd person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you</td>
<td>979 (47.23)</td>
<td>77.51</td>
<td>884 (37.7)</td>
<td>75.17</td>
<td>85 (2.95)</td>
<td>66.41</td>
</tr>
<tr>
<td>yourself</td>
<td>23 (1.10)</td>
<td>1.82</td>
<td>49 (2.09)</td>
<td>4.17</td>
<td>3 (0.1)</td>
<td>2.34</td>
</tr>
<tr>
<td>yourselves</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>your</td>
<td>261 (12.59)</td>
<td>20.67</td>
<td>243 (10.39)</td>
<td>20.66</td>
<td>40 (1.39)</td>
<td>31.25</td>
</tr>
<tr>
<td>Total</td>
<td>1263 (60.92)</td>
<td>100</td>
<td>1176 (50.18)</td>
<td>100</td>
<td>128 (4.44)</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 6 shows LS and US learners use first- and second-person pronouns more frequently than the older learners in SWICLE. Specifically, LS students use the pronouns you and I more frequently than US and SWICLE students, which are the most frequent pronouns in each corpus. The overall use of first-person pronouns is relatively consistent between LS and US, with SWICLE showing a slightly lower percentage. This suggests a moderate similarity in the use of first-person pronouns among young learners, with a subtle decrease as they advance in proficiency. The differences in the frequency of I across the three corpora may be attributed to various factors, including proficiency level, writing development and experience. Additional variables that may have influenced the findings is that lack of secondary materials tends to influence more involvement features (Ådel, 2008). Also, LS exhibition of higher frequency of
"I may reflect their lack of secondary materials which focus on simple expressions of personal experiences. For example, from LS:

(13) In this text I will discuss what things I think are necessary for a good life and why I think these things are necessary. (SLEC H_8_M_22_105)

This example could be rephrased to preserve the original meaning of the text while eliminating any first-person references, thereby rendering it more objective and impersonal in tone. By contrast, the US texts show a decrease in the frequency of I, suggesting a shift towards less involvement as learners progress in their proficiency and gain writing experience. In comparison to SWICLE, however, we find that the advanced learners demonstrate the lowest frequency of I, indicative of improved register awareness. The gained proficiency allows them to convey ideas using diverse pronouns such as we, reducing the need for frequent use of I.

The use of we shows a clear developmental trend across the three corpora. There is a noticeable increase in the use of we as learners progress from LS to US to SWICLE. This suggests a developmental trend where advanced and proficient learners use we more frequently. This trend demonstrates that advanced and proficient learners use the inclusive pronoun more frequently, reflecting a nuanced evolution in language expression as learners advance in proficiency and experience. An example from SWICLE demonstrates the use of we:

(14) Unfortunately, not all things we discover are entirely good. (SWICLE, SWUL4003)

In this context, we is used to refer to a collective group that includes both the writer and the audience. Instead of using I, the use of we suggests a shared experience or perspective between the speaker and the reader in the act of discovering things, and it implies a sense of
shared responsibility or involvement in the statement. Thus, the writer achieves a better approach to persuade the reader.

There is an overall decline in the use of the second-person pronoun as learners progress, demonstrating that young learners tend to utilise the pronouns more often than their more proficient counterparts. Specifically, there is a significant decrease in the use of the pronoun you as learners progress from LS to US and a more substantial decline in SWICLE. This suggests a developmental trend where advanced and proficient learners use you less frequently. Proficient learners may engage more frequently in formal and academic writing (Virtanen, 1998), where the use of the second person can be minimised. This shift could be influenced by academic conventions that favour a more detached perspective (Ädel, 2008). Additionally, diverse cultural backgrounds may influence their writing choices (Paquot et al., 2013). Some cultures or academic settings may encourage a more indirect or formal communication style, impacting the use of you. Similar to you, there is a consistent decrease in the use of the determiner your as learners progress from LS to US and an even more significant decline in SWICLE. The gain in proficiency may lead inexperienced writers to start engaging in more frequent use of formal and academic writing contexts. In such contexts, there might be a preference for more diverse language choices, leading to a decrease in the direct use of your.

By contrast, SWICLE registers considerably lower frequencies of second-person pronouns. Despite the SLEC students displaying higher frequencies of I, the differences between SLEC-US and SWICLE students’ overall use of all first-person pronouns is quite low, considering this is the period advanced students tend to show less visibility compared to young learners (Hasund & Hasselgård, 2022). The low frequency in the SWICLE use of I demonstrates their
awareness of register, despite their lack of access to supporting materials during their essay writing. According to Ädel (2008), the lack of materials may restrict the scope of the student’s arguments to their own personal experiences, which may lead to an informal and narrative style rather than a formal and argumentative one. Additionally, despite the SWICLE students’ lack of time to formalise their writing, which may influence the overuse of I in their essays, they, however, used I less frequently than both SLEC groups.

In summary, the overall differences and trends in pronoun usage hint at the evolving writing styles and linguistic choices of advanced learners compared to young learners. Advanced learners in SWICLE appear to exhibit distinct patterns in the use of first and second-person pronouns, reflecting potential developmental changes in their language proficiency and communicative strategies.

4.3 The Collocation I think

The results of the collocation I think show that LS learners used it most frequently (4.63 ptw) while the SWICLE students used it least frequently (1.56 ptw). The frequency of I think in the US was 2.69 ptw. A notable decrease in the usage of the phrase I think is observed as learners transition from the LS to US and finally to the SWICLE corpus. This trend suggests that advanced and proficient learners tend to rely less on this phrase compared to their less proficient counterparts. As they gain proficiency, the inexperienced writers (LS) rely less on I think to express personal opinions and engage with the reader (Ädel, 2008; Paquot et al., 2013). The gained proficiency and experience allow them to express opinions and beliefs with more varied and nuanced language. As a result, learners may employ a broader range of expressions, reducing reliance on repetitive phrases like I think. Considering that proficient learners often engage in more formal and academic writing contexts, writers may tend to structure their
sentences in a way that conveys their thoughts without explicitly stating *I think*. For example, in cases where a student could have said *I think*:

(15) But it is my belief that, in an ever increasingly complex and competitive society…

In this example, a less proficient student could have started the sentence with, “I think that…”. This possible difference in diverse sentence structure that tends to occur with gained proficiency may contribute to the observed decrease in *I think* usage in the SWICLE.

Similar to the Norwegian study (Hasund & Hasselgård, 2022), the current study also found that the usage of the phrase *I think* is notably common across the three corpora and contributes greatly to the frequent appearance of both the pronoun *I* and the verb *think* (Hasund & Hasselgård, 2022, p. 455). However, a closer look at other words that collocate with *think* shows that US has the highest frequency of other words that collocates with *think*. The other collocate the US used other than *I think* was *to think* of which SWICLE used the second most frequently as well. LS used *I don’t think* the most after *I think*. Furthermore, similar to pronoun *I*, *I think* may be used to signal the young learners’ stance on an issue, or to distance themselves from the views of others or shift the learner’s own perspective to the perspective of another person (Aijmer, 2001).

4.4 Modal Adverbs

Table 7 shows the distribution of modal verbs in the three corpora. Modal adverbs are features used to express a wide range of stances, including possibility (e.g., *perhaps*), obligation (e.g., *certainly*) or certainty (e.g., *definitely*) (Biber & Conrad, 2009). While the current study did not tabulate the modal adverbs depending on their attitudes, some of the adverbs mark attitude rather than possibility, obligation, or certainty. For example, *frankly, naturally, and*
unfortunately are attitude markers. However, these attitude markers did not yield any significant frequencies and therefore no focus was placed on them.

Table 7. Distribution of Modal Adverbs in the Three Corpora

<table>
<thead>
<tr>
<th>Modals</th>
<th>LS (ptw)</th>
<th>%</th>
<th>US (ptw)</th>
<th>%</th>
<th>SWICLE (ptw)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>actually</td>
<td>7 (0.33)</td>
<td>7.37</td>
<td>5 (0.21)</td>
<td>7.14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>certainly</td>
<td>1 (0.04)</td>
<td>1.05</td>
<td>2 (0.08)</td>
<td>2.86</td>
<td>12 (0.41)</td>
<td>10.62</td>
</tr>
<tr>
<td>completely</td>
<td>1 (0.04)</td>
<td>1.05</td>
<td>3 (0.12)</td>
<td>4.29</td>
<td>6 (0.20)</td>
<td>5.31</td>
</tr>
<tr>
<td>definitely</td>
<td>3 (0.14)</td>
<td>3.16</td>
<td>0</td>
<td>0</td>
<td>3 (0.10)</td>
<td>2.65</td>
</tr>
<tr>
<td>frankly</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>maybe</td>
<td>31 (1.49)</td>
<td>32.63</td>
<td>11 (0.47)</td>
<td>15.71</td>
<td>6 (0.20)</td>
<td>5.31</td>
</tr>
<tr>
<td>naturally</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6 (0.20)</td>
<td>5.31</td>
</tr>
<tr>
<td>perhaps</td>
<td>2 (0.96)</td>
<td>2.11</td>
<td>2 (0.08)</td>
<td>2.86</td>
<td>18 (0.62)</td>
<td>15.93</td>
</tr>
<tr>
<td>obviously</td>
<td>0</td>
<td>0</td>
<td>2 (0.08)</td>
<td>2.86</td>
<td>2 (0.06)</td>
<td>1.77</td>
</tr>
<tr>
<td>of course</td>
<td>1 (0.04)</td>
<td>1.05</td>
<td>3 (0.12)</td>
<td>4.29</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>possibly</td>
<td>1 (0.04)</td>
<td>1.05</td>
<td>3 (0.12)</td>
<td>4.29</td>
<td>1 (0.03)</td>
<td>0.88</td>
</tr>
<tr>
<td>probably</td>
<td>8 (0.38)</td>
<td>8.42</td>
<td>10 (0.42)</td>
<td>14.29</td>
<td>20 (0.69)</td>
<td>17.7</td>
</tr>
<tr>
<td>really</td>
<td>37 (1.78)</td>
<td>38.95</td>
<td>29 (1.24)</td>
<td>41.43</td>
<td>26 (0.90)</td>
<td>23</td>
</tr>
<tr>
<td>totally</td>
<td>3 (0.14)</td>
<td>3.16</td>
<td>0</td>
<td>0</td>
<td>5 (0.17)</td>
<td>4.43</td>
</tr>
<tr>
<td>unfortunately</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8 (0.27)</td>
<td>7.09</td>
</tr>
</tbody>
</table>

As shown in Table 7, LS students use modal adverbs more frequently than the US and SWICLE writers. Interestingly, US contains fewer adverbs than LS and SWICLE, especially in totally and definitely where US had zero occurrences. The findings show similar patterns to Hasund and Hasselgård’s (2022) study of L1 Norwegian learners. In both L1 Norwegian and Swedish student writing, there is a high frequency of modal adverb usage among the youngest learners, which is then reduced among the older secondary students before rising again in the advanced learners. The pattern in the Norwegian study was attributed to the limited size of the young learners’ subcorpus and the tendency of certain individuals to utilise adverbs frequently. However, the corpus size explanation may not apply to the SLEC corpus, which is considerably bigger. Yet, there was a similarity with certain writers in the LS being prolific modal adverb users. However, the current study did not further investigate and account for every possible
outlier in the corpus, which may reveal extreme values that could stand out greatly from the overall pattern of values.

Really is a highly used modal adverb across all learner groups, indicating its significance in expressing stance. There is a noticeable decrease of really where it is used most frequently in the LS than US and least frequently in the SWICLE. This trend suggests that there is a development in the use of the word, with less proficient learners depending on it more frequently than more proficient learners. This can be explained by proficient learners who tend to diversify their language use. Instead of relying heavily on a single adverb like really, they explore alternative expressions and synonyms to convey a stance such as totally which rises in the SWICLE. There is a consistent decrease in the frequency of maybe as learners progress from LS to US to SWICLE. Maybe is used most frequently in the SWICLE followed by the US and least frequently in the LS. This developmental trend demonstrates that as proficiency increases, learners may employ more varied ways to convey uncertainty or probability. To reduce the need for repetitive use of maybe, the varied ways are seen in the SWICLE opting to use probably which noticeably increases in the frequency as learners progress from LS to US to SWICLE. Interestingly, perhaps is used most frequently in the LS and least frequently in the US before slightly rising again in the SWICLE. Despite this pattern, perhaps is the third most used modal adverb within the SWICLE after really and probably.

Analysing the usage of the modal adverbs within each corpus, the LS and US rely primarily on really and maybe. The SWICLE students demonstrate register awareness by using a wider variety of hedging devices (e.g. perhaps) compared to LS who rely on emphatic markers (e.g. really) to express a high degree of certainty or emphasis. According to Ádel (2008, p. 36), informal writing exhibits linguistic features indicative of involvement, including first-person
references and emphatic particles. In contrast, formal writing is marked by linguistic features that convey detachment. These results suggest that there is less reliance on informality among the students’ writing as they progress in proficiency. This might be a result of learners displaying reduced engagement when writing without a time constraint. In comparison to SLEC, which incorporated measures to promote adherence to teacher guidelines, the flexibility to extend writing time, and prior exposure to the topics, SWICLE writers, as noted by Ādel (2008), faced some constraints. These limitations encompass insufficient time to refine their text and give it a more polished, “written-like” quality.

4.5 Modal Auxiliaries

Table 8 presents the distribution of modal auxiliaries in the three corpora.

<table>
<thead>
<tr>
<th>Modals</th>
<th>LS (ptw)</th>
<th>%</th>
<th>US (ptw)</th>
<th>%</th>
<th>SWICLE (ptw)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>can</td>
<td>303 (14.62)</td>
<td>46.19</td>
<td>283 (12.10)</td>
<td>47.25</td>
<td>93 (3.23)</td>
<td>16.94</td>
</tr>
<tr>
<td>can’t, cannot</td>
<td>6 (0.28)</td>
<td>0.91</td>
<td>8 (0.34)</td>
<td>1.34</td>
<td>16 (0.55)</td>
<td>2.91</td>
</tr>
<tr>
<td>could</td>
<td>59 (2.84)</td>
<td>8.99</td>
<td>29 (1.24)</td>
<td>4.84</td>
<td>37 (1.28)</td>
<td>6.74</td>
</tr>
<tr>
<td>may</td>
<td>12 (0.57)</td>
<td>1.83</td>
<td>23 (0.98)</td>
<td>3.84</td>
<td>21 (0.72)</td>
<td>3.83</td>
</tr>
<tr>
<td>might</td>
<td>32 (1.54)</td>
<td>4.88</td>
<td>30 (1.28)</td>
<td>5.01</td>
<td>30 (1.04)</td>
<td>5.46</td>
</tr>
<tr>
<td>must</td>
<td>7 (0.33)</td>
<td>1.07</td>
<td>13 (0.55)</td>
<td>2.17</td>
<td>56 (1.94)</td>
<td>10.20</td>
</tr>
<tr>
<td>ought</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5 (0.17)</td>
<td>0.91</td>
</tr>
<tr>
<td>shall</td>
<td>1 (0.04)</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>should</td>
<td>41 (1.97)</td>
<td>6.25</td>
<td>42 (1.79)</td>
<td>7</td>
<td>60 (2.08)</td>
<td>10.93</td>
</tr>
<tr>
<td>will</td>
<td>144 (6.94)</td>
<td>21.95</td>
<td>110 (4.70)</td>
<td>18.36</td>
<td>134 (4.65)</td>
<td>24.41</td>
</tr>
<tr>
<td>won’t</td>
<td>7 (0.33)</td>
<td>1.07</td>
<td>4 (0.17)</td>
<td>0.67</td>
<td>1 (0.03)</td>
<td>0.18</td>
</tr>
<tr>
<td>would</td>
<td>44 (2.12)</td>
<td>6.71</td>
<td>57 (2.43)</td>
<td>9.52</td>
<td>96 (3.3)</td>
<td>17.49</td>
</tr>
</tbody>
</table>

Total Modals | 656 (31.58) | 100 | 599 (25.58) | 100 | 549 (18.99) | 100 |

Overall, there is noteworthy variation in the use of modal auxiliaries across the three learner groups. LS writers use the modals most frequently, while US writers use them less, and SWICLE writers use them the least. It is apparent that there are fewer can, could, might, and will in US and SWICLE. Excluding will, it is expected that SWICLE would increase the number of these hedging modals due to register awareness that is a result of gained proficiency. However, the frequency of the hedging modals in SWICLE declines instead of increasing. In
contrast to SLEC, which had measures to encourage teacher guidelines, the option to extend writing time, and prior familiarity with the topics, SWICLE writers, as per Ädel (2008), had some limitations. The limitations include a lack of sufficient time to formalise their text and make it sound “written-like.” Additionally, they lacked access to expert opinions in their writing compared to SLEC who had teacher guidelines before writing. Also, there are more obligation modals, *would* and *must* in US and SWICLE. While it is expected *would* to increase, it is interesting why *must* is increasing as well. Additionally, the frequency of *should* is slightly lower in the US, but SWICLE includes the highest frequency (though not to a great extent). Both *should* and *must* are obligation modals. The overuse of necessity or obligation modals like *must* and *should* is a method of embracing a direct and emphatic persuasive style. For example in SWICLE:

(16) Therefore, however difficult it may be, Sweden must moderate in the struggle between assimilation and integration and move towards, not only allowing but encouraging a society to emerge where the "indigenous" people can and want to live in "neighborly love" with immigrants. (SWICLE SWUL5009)

In this usage of *must*, it potentially conveys a clear connection to the topic at hand, where students seem to have strong opinions (Aijmer, 2002). The topic is on *Assimilation or Integration* whereby the modal verb *must* is used deontically. It expresses a necessity or obligation, suggesting that Sweden is under a moral or practical duty to moderate the struggle between assimilation and integration and move towards encouraging a society where “indigenous” people can coexist with immigrants in “neighbourly love.”

Comparatively, in Hassund and Hasselgård’s (2022) study, the overall total of modals displayed a small difference in change in relation to the students’ development in proficiency. However,
Table 8 shows the total frequencies of the modal auxiliaries decline as the L1 Swedish students develop in proficiency. Considering that most of the modal auxiliaries are used epistemically, the results suggest that the students’ reliance on using the modal auxiliaries to communicate their uncertainty or probability regarding the statements declines as they progress in proficiency. Interestingly, one of the similarities found between L1 Norwegian and Swedish learners (in the current study) is that there might be cultural or L1 transfer among the youngest students (Paquot et al., 2013; Petch-Tyson, 1998). This transfer is seen when they dominate the use of can/could will/would and should (Swedish cognates kan/kunde, vill/ville and skulle) of which Hassund and Hasselgård as well mention Norwegian cognates “kan/kunne, vil/ville and skulle” (p.10). Additionally, L1 transfer is seen among both countries’ young learners, whom previous studies consider to be not fully aware of language registers (Lee et al., 2019). They are seen to use less might which, similarly to Swedish, Hasund and Hasselgård mention that the modals may and might have “no direct counterparts in Norwegian” (p.10). However, the distribution of may which is similar to the function of might is used less frequently by LS. Additionally, the use of might among Swedish students across the three corpora demonstrates a small difference in frequency, which gives this argument about lack of register among young learners less weight.

4.6 Questions

The questions, as identified by question marks, show marginal differences. LS used questions the least frequently (1.06 ptw) while the SWICLE used questions the most frequently (1.73 ptw); the frequency of questions in US is 1.24 ptw. The frequency of questions appears to increase with proficiency. However, the slightly higher question marks in SWICLE may be as a result of the task that can influence how the texts begin in their introductions. For example, a text from SWICLE started:
The question is do we really need everything that it brings us, and what does this technology do to our minds, hopes and dreams? (SWICLE SWUL5010)

This observation is supported by a study on university-level ESL student writing, where the majority of students used questions in the introduction as a strategy to hook the readers (Lee et al., 2019). By employing an indirect approach (rhetorical question), the students subtly steer the reader’s thoughts and evaluation without overtly asserting a definitive position, hence encouraging a more contemplative and open-ended involvement with the subject (Henkemans et al., 2002). As a result, the rhetorical questions are used to underscore their point (Hyland, 2005). While research shows that interrogatives are infrequent in academic writing, these findings suggest that, as the students develop proficiency, they rely on their register awareness by exhibiting indirectness when using rhetorical questions in their writing.

To conclude this results and discussion section, it is important to emphasise some additional variables that may have influenced the findings. As highlighted by Recski (2004), argumentative essays found in the International Corpus of Learner English (ICLE) pose a challenge when it comes to steering clear of personal references and subjective attitudes. This difficulty arises because learners are explicitly encouraged to take a persuasive approach and express their personal opinions (Paquot et al., 2013). As mentioned in section 3.2.2, the ICLE featured conventional argumentative themes in the topics such as Integration or assimilation?, The use and abuse of state media, In defence of popular culture, Unemployment and the future, Reflections on being a mature student and Christmas 1997 - traditional or innovative?”. In comparison, the SLEC had only one essay prompt for all students: What are the most crucial aspects for leading a good life? To compare SWICLE and SLEC, the essay prompt that provides a greater scope for a personalized viewpoint is the one from the SLEC. The topic tends
to encourage the learners to express their personal beliefs, values, and experiences, making it more conducive to a personalized and subjective perspective. The topics from the SWICLE, while still allowing for personal opinions, may involve more objective analysis, research, or discussion of broader societal issues. For example, SWICLE topics like *Reflections on being a mature student* and *Christmas 1997 - traditional or innovative* may involve personal perspectives. However, the degree of expressing personal perspective is likely higher in the SLEC topic, depending on the writer’s interests and experiences. It is important to note that despite the anticipated influence of these variables on the outcomes, the present study did not investigate or account for variations of WRV characteristics in every topic covered in the SWICLE.

5 Conclusion

This corpus-based study aimed to investigate the evolving use of WRV features in the writing of young Swedish learners of English. Specifically, it compared the presence of these features in argumentative essays written by Swedish students at the lower and upper secondary school levels. The study also compared the WRV features employed in the young students’ essays and those found in the writings of Swedish university-level learners. The results show that lower secondary students use more WRV features than upper secondary students, except for questions, where the difference is minimal. This aligns with Hasund and Hasselgård’s (2022) observation that “the tendency to be visible writers starts early” (p. 19). Yet, as their writing proficiency or experience increases, their visibility decreases by relying less on WRV devices. Additionally, as Paquot et al. (2013) obvers, “Scandinavian writers favour a less formal, more interactive writing style” (p. 8). Yet again, the results show that they appear to rely less on this “less formal, more interactive” expression as their writing develops.
Among the features analysed, pronouns were the most common ways in which both lower and upper secondary demonstrated WRV in their writing. Yet, while second-person pronouns were the most frequent in both lower and upper secondary corpora, first-person pronouns showed the highest and most evenly distributed frequencies across the three corpora, with the pronoun *I* being the most dominant. Although the frequency pattern of first- and second-person pronouns does not align with Hasund and Hasselgård’s (2022) findings, the predominance of the pronoun *I* supports their conclusion that young learners tend to be more interactive in their writing but their need to be visible writers reduces as they develop in their proficiency. Additionally, the low frequency of *I* in the SWICLE and more modals that hedge their statements demonstrates that as the students grow in proficiency, their awareness of the register develops as well. That is, they develop the need to express themselves in a formal and argumentative way rather than an informal and narrative style. The modal auxiliary *can* was the most frequently used in LS, and its frequency was reduced in US and SWICLE.

This study’s main limitation lies in its quantitative approach, which restricts the analysis of the subtle differences between modals that necessitate a more in-depth examination to determine their contextual usage. Future studies can improve by combining quantitative analysis with qualitative methods to triangulate findings. A mixed-methods approach allows for a more comprehensive exploration of the data and facilitates a deeper understanding of the patterns observed quantitatively. Additionally, the study is limited by its failure to consider the distribution within the sample. Consequently, it remains unclear to what degree the findings accurately portray the writers as a collective, and whether there are individual writers who serve as outliers, exhibiting an extreme overuse or underuse of certain features under examination. Furthermore, given that both SLEC and SWICLE involve timed essays, it is important to
acknowledge a potential flaw in the comparison to Hasund and Hasselgård’s (2022) study, as their longitudinal data collection did not account for the timing factor. This limitation can be addressed in the future by utilizing corpora that distinguish between timed and untimed essays.

Despite these limitations, this study of WRV features in Swedish young learners’ writing might contribute to the understanding of language development as well as the development of writing instructions for young learners of English. Specifically, it can provide insights into how young learners bridge the gap between L1 and L2 English writing skills. This can be done by investigating the extent to which learners transfer writing strategies from their L1 to L2. For example, certain WRV features may be more prevalent in the writing of learners from specific linguistic backgrounds. Understanding these transfer patterns can guide educators in addressing language-specific challenges. Additionally, analysing WRV features in the writing of young learners can reveal patterns of interlanguage development. By examining how learners use WRV features, researchers can identify stages of proficiency and uncover common challenges faced during the transition from L1 to L2 writing. Furthermore, it can help writing instructors provide age-appropriate support in young learners’ writing, as well as help students develop appropriate register awareness by informing the students about the genre writing requirements.
References


http://ci.nii.ac.jp/ncid/BA67607594


https://doi.org/10.1075/scl.6
