



FACULTY OF EDUCATION AND BUSINESS STUDIES  
Department of Humanities

---

# An Error Analysis of Subject-Verb Agreement by Swedish Learners of English

A Corpus-Based Study of the Difficulties Surrounding Subject-Verb  
Agreement for Swedish Students

Viktor Molin

2020

Student thesis, Bachelor degree, 15 HE  
English  
Upper Secondary Teacher Education Programme  
Engelska för ämneslärare (61-90)

Supervisor: Henrik Kaatari  
Examiner: Kavita Thomas

---

## **Abstract**

This study investigates subject-verb agreement errors produced by students in upper secondary school in Sweden. The research was conducted in an attempt to understand why students acquiring English as their L2 struggle with learning subject-verb agreement, and to categorize the grammatical environments which cause students to produce errors related to subject-verb agreement. There are two extralinguistic variables at the center of the study: (i) gender and (ii) education level. Error analysis was the method of choice for this study, and the study used essays produced by students from the Uppsala Learner of English Corpus to identify and classify subject-verb agreement errors according to linguistic and extralinguistic variables. The linguistic variables were separated into four clusters: (i) subject type, (ii) subject number, (iii) distance and (iv) verb. The variables in each cluster were: (i) common noun, pronoun, existential clause; (ii) singular, plural; (iii) subject-postmodifier, other distance, no distance; (iv) auxiliary verb, regular, irregular. In total, 117 texts were included in the analysis with 57 texts being written by female students, and 60 from male students, to ensure that gender could be used as a variable. Previous research suggests that female students are slightly better academically than male students and suggests that subject-verb agreement is an issue across all levels of education. These findings were supported by this study, as the normalized frequencies for gender show that not only do female students perform better than male students, but they also improve across level while male students get worse. The grammatical environments that proved to be most difficult for the participants to produce correctly were *common nouns* 57.7% (n=75), *plural subjects* 54.6% (n=71), *irregular verbs* 59.2% (n=77) and *distance* 30.8% (n=40).

**Keywords:** Error Analysis, SLA, Swedish, ELT, Subject-Verb Agreement

## Table of Contents

1. Introduction .....	1
1.1 Aim and Scope .....	2
2. Literature and Research Review .....	3
2.2 Theoretical Basis .....	3
2.2.1 Previous Research .....	3
2.2.2 Second Language Acquisition .....	5
2.3 Grammatical Background.....	6
2.3.1 Subject .....	7
2.3.2 Verb .....	9
2.3.3 Distance .....	11
3. Method & Material .....	12
3.1 Primary Sources.....	12
3.2 Selection of Data .....	13
3.3 Method of Analysis .....	14
3.4 Validity and Reliability .....	16
4. Results .....	19
4.1 Linguistic Variables.....	19
4.2 Extralinguistic variables .....	21
4.3 Cross-analysis of the variables .....	23
4.4 Discussion.....	26
5. Conclusion .....	29
References .....	30
Primary References.....	30
Secondary References.....	31

# 1. Introduction

During the 19th century, the national curriculum in Sweden for English language teaching (ELT) focused on grammar because it was the classic way of teaching the English language (Köhlmyr, 2003: 1). This has since changed and now ELT focuses more on helping the students to be understood by learning all-around communicative skills (Skolverket, 2011). Grammar teaching in Sweden has thus gone from being the primary focus of ELT, to being taught passively through other activities such as reading and listening comprehension (Köhlmyr, 2003: 1; Harmer, 2007: 239). While explicit grammar teaching through course books and presentations can be effective, certain aspects of grammar are simply too complex and abstract for students to learn explicitly (Lightbown & Spada, 2013: 48). However, even simple grammatical concepts that are easy to teach and learn may contain complex and abstract components as well. An example of this is subject-verb agreement and the third-person singular 's'. Lightbown & Spada (2013: 48) state that "A simple 'rule of thumb' such as 'put an -s at the end of a noun to make it plural' may be a better target for instructions". This sort of instruction may be suited for explicit grammar teaching through introductions, but what about the other aspects of subject-verb agreement? With the complexity of certain subjects, it may be difficult to learn when to use a singular or plural verb form (see section 2.3.1). However, empirical studies show that even simple cases of subject-verb agreement present problems for Swedish students in the production of texts in English at all education levels (Köhlmyr, 2003: 173; Källkvist & Peterson, 2006: 112). Research has hypothesized that one of the reasons behind the struggle of learning subject-verb agreement is due to Swedish verbs not agreeing with the subject in number, despite having verb inflection (Jackson et al.: 2018). Since not all verb forms are affected by subject-verb agreement, students may not learn the rules behind subject-verb agreement, and it may lead to the students using the incorrect verb forms (see section 2.3.1). To investigate the complexity of subject-verb agreement acquisition, the study will analyze in what grammatical environments Swedish learners produce subject-verb agreement errors in English. To assess this, essays produced by students will be analyzed for cases of subject-verb agreement errors through the use of error analysis. The reason for choosing subject-verb agreement as the grammatical concept to use for this study is due to it being one of the most common errors produced by all Swedish learners of English,

regardless of level (Källkvist & Peterson, 2006: 112). The hypothesis is that the errors become less frequent over the course of the students' education due to their exposure to the English language through education and other sources. In addition to looking at education level as an extralinguistic variable, gender will also be included in the analysis to study its effect on the production of subject-verb agreement errors.

## 1.1 Aim and Scope

There are two aims for this study: (i) to establish the grammatical environments in which Swedish learners produce subject-verb agreement errors and (ii) analyze the extent to which the extralinguistic variables gender and education level affect error rate. The grammatical environments that are relevant for this study are common noun, pronoun, existential clause, singular, plural, subject-postmodifier, other distance, no distance, auxiliary verb, regular verb and irregular verb (more on these in section 3.3)

Due to this study analyzing subject-verb agreement exclusively, the scope of the study will focus on cases that can be interpreted. What this means is that cases of spelling errors will be excluded to ensure the integrity of the study. There will be some exemplification of errors that are based on other issues that manifest themselves as subject-verb agreement errors, such as the omission of a subject, verb or auxiliary verb. Nevertheless, the focus will be on subject-verb agreement in texts written by real students. The material is collected from the *Uppsala Learner English Corpus* (ULEC) and consists of 117 essays produced by students at Swedish Gymnasium Year 1 (GY1) and Swedish Gymnasium Year 3 (GY3). To investigate the aims of the study with the scope in mind, the following research questions will be used:

- Which grammatical environments (i.e. common noun, pronoun, existential clause, singular, plural, subject-postmodifier, other distance, no distance, auxiliary verb, regular verb, and irregular verb) produce the highest error rate in terms of subject-verb agreement errors?
- Are there differences across education levels and gender in terms of subject-verb agreement error rate (i.e. do the students improve across level)?

These research questions will help assess how the linguistic and extralinguistic variables

affect the production of subject-verb agreement errors. The linguistic variables will be discussed further in section 3.3.

## 2. Literature and Research Review

This section will be divided into two subsections to go in-depth into the different aspects that are vital to this study and to the analysis of the result. The first section, theoretical basis, will cover previous research done on the subject and SLA. The SLA section will cover some concerns established with contrastive analysis to elaborate on areas of difference between the two languages before connecting the method employed for analysis, namely error analysis, to the study and previous research. The second section will establish a grammatical basis which will be the backbone for the analysis itself.

### 2.2 Theoretical Basis

#### 2.2.1 Previous Research

Previous SLA studies conducted on Swedish students and their proficiency in subject-verb agreement are limited. However, Köhlmyr (2003) has done an extensive investigation regarding grammatical errors produced by 16-year-old students. The material investigated in Köhlmyr's study is randomized and consists of 400 essays produced by students and thus could be seen as a good approximation of what is expected out of this group. One part of her study is dedicated to subject-verb agreement and will thus be used as a point of reference when analyzing the results of this study.

Other research focuses on the difficulty of understanding and learning subject-verb agreement based on the L1 of the learners (Wei et al., 2015; Jackson et al., 2018). Jackson et al. (2018) claim that Swedish students learning English as their L2 should have an easier time understanding and using subject-verb agreement in production as compared to languages without verb inflection. This is because Swedish sentence structure is relatively similar to English and both English and Swedish make use of verb inflection (more on this in section 2.3). Despite this, according to Källkvist & Peterson (2006: 112), subject-verb agreement is the fourth most frequent grammatical

error produced by not only Swedish students from compulsory school, but also Swedish university students. What this means is that regardless of education level, it appears to be a problem despite improved knowledge of the language. While there is no clear consensus on why students struggle with understanding concepts such as third-person singular 's' for English present-tense, data presented in empirical studies like Köhlmyr proves that it is a problem (Köhlmyr, 2003; 174). Researchers speculate that students struggle with third-person singular 's' due to the inconsistencies of the English language (Källkvist & Peterson, 2006: 115). These inconsistencies are things like indefinite pronouns taking third-person singular verb forms even though they sound like they are plural, or the verb 'be' having inflections in both present tense and past tense regarding number (Källkvist & Peterson, 2006: 115).

Research on the last extralinguistic variable, gender, appears to be lacking when it comes to the impact gender has on L2 acquisition. van der Slik, van Hout, and Schepens JJ (2015; 1) explain that “the relative absence of research on the role of gender in L2 acquisition might be that female L2 learners doing better than male learners is regarded as being common knowledge.” Due to this, this study will use the statistics provided by the Swedish National Agency for Education (Skolverket) and the Programme for International Student Assessment (PISA) as examples to demonstrate the gender-gap situation in Sweden. According to PISA, gender is starting to become less of a factor when it comes to the result in overall education (PISA, 2018). This is also suggested by the statistics provided by Skolverket as their records show that there is virtually no grade difference between males and females in English 5 and 6 (Skolverket, 2020). However, PISA (2018) did show that in reading comprehension, although the gap between genders is decreasing, the gap between males and females in Sweden is still significant, in which females perform better than males. While there is no evidence that reading comprehension affects L2 acquisition by a measurable margin, it could be a contributing factor in not understanding concepts like subject-verb agreement. For the sake of this study, the statistics provided by PISA and Skolverket will serve as a guideline for the role of gender in the Swedish education system, while the comment made by van der Slik et al. (2015) will serve as a reminder of the global perspective regarding how gender affects L2 acquisition.

## 2.2.2 Second Language Acquisition

Second language acquisition is defined as acquiring an L2 after having learned the basis for the L1 (Mitchel et al., 2013: 1). It is a rather complex concept when including the many factors that go into acquiring a second language. However, SLA research is of importance because the result may help teachers and students alike in the process of acquiring an L2 (Mitchel et al., 2013: 2). This section will briefly mention the concept of contrastive analysis before focusing on the main method of analysis that will be used for this study, error analysis.

The contrastive analysis was created as a response to the behaviouristic side of SLA and focuses on the dissimilarities between the L1 and L2 (Mitchell et al. 2013: 29). These dissimilarities make a language more or less difficult to acquire and is an important basis of the analysis as it serves as a way to prevent errors in SLA. Lightbown & Spada (2013: 41) state that contrastive analysis is based on assumed transfer from the L1 to the L2. The example they use to illustrate this is that French learners of English 'should' make the opposite errors of English learners of French (Lightbown & Spada, 2013: 42). In French, direct objects that are nouns succeed the verb just like in English, but if it is a pronoun it instead precedes the verb and contrastive analysis would thus predict this to be an issue for L2 acquisition. This is how research like, for example, Jackson et al. (2018) and Wei et al. (2015) utilizes the contrastive analysis to create a hypothesis. For this study, seeing as there is no inflection based on subject number in Swedish, subject-verb agreement as a whole should be problematic for Swedish learners of English and so should uncountable nouns as some are countable in Swedish but not English (e.g. Eng. *Money* Sw. *Pengar*; more on uncountable nouns in section 2.3.1).

Error analysis categorizes and analyses errors produced by L2 learners of English (Lightbown & Spada, 2013: 42). Unlike contrastive analysis, it focuses on the production of students in order to identify flaws in their knowledge (Mitchell et al. 2013: 35). In other words, while L1 interference may be a reason for some of the errors, they do not have to originate from the L1. Research conducted on material produced by students generally utilize this method of analysis to find and categorize errors made by the students (Köhlmyr, 2003: Källkvist & Peterson, 2006). This research further demonstrates that L1 does not have to be the main reason behind the struggle to acquire subject-verb agreement, and therefore illustrates the importance of analyzing the data for errors (Köhlmyr, 2003: 173-181; Källkvist & Peterson, 2006: 121-126). Comparing

the use of error analysis in linguistic research (Köhlmyr, 2003; Källkvist & Peterson, 2006) with the use of contrastive analysis in research (Jackson et al. 2018; Wei et al., 2015), researchers use contrastive analysis to compare the structure of the involved languages to find areas of difference which may cause issues for learners of the target language, while they use error analysis to analyze why the students make certain errors. Contrastive and error analysis can thus be used concurrently in SLA studies to include more perspectives behind linguistic phenomena. If this study analyzed more aspects of grammar, like prepositions, contrastive analysis would be of higher importance to explain and predict more specific errors made by Swedish students due to the differences between their L1 and L2. Error analysis is mainly criticized as a method of analysis for two reasons: (i) it is generally limited to the errors produced by learners and (ii) the selection of data may cause the result to be biased (Khansir, 2012). The first criticism (i) is considered a limitation due to the fact that error analysis primarily focuses on analyzing and categorizing the errors made by learners. This limit could be avoided if the researcher includes the cases in which the student produces the analyzed aspect correctly. The second criticism (ii) is considered a flaw as the data is selected by the researcher and could be selected to generate a biased result. An example of this would be if this study used students in different types of educations, such as comparing GY1 students in a university preparation program to GY3 students in a work preparation program. This would likely show that GY1 is better than GY3 and, if that was the hypothesis, would thus show a biased result. However, what this means for this study is that the data has a significant flaw based on the nature of error analysis studies, namely that it only looks for the errors produced by learners. In order to get a more accurate representation of the issues analyzed with error analysis, the study would have needed to include the cases in which the learner accurately produces the correct forms of subject-verb agreement in order to compare that data to the cases in which they do it inaccurately. More on this issue in section 4.4.

### 2.3 Grammatical Background

Subject-verb agreement is one of the fundamentals of grammatical knowledge as every sentence consists of a subject and a verb. While some subject-verb agreement rules may be simple in theory, such as third-person singular verbs taking the 's' ending in most cases where the subject is singular, there are some rules that are more complex (Harmer,

2007: 23). This section will describe some of the language aspects that may cause the students to produce subject-verb agreement errors, while also describing the concept of subject-verb agreement in detail. Table 1 demonstrates some errors produced by students and serves as a basis for the error analysis in this study.

Table 1. Examples of errors from ULEC

Errors where the subject number or verb form is inaccurate	Errors where the auxiliary verb form is inaccurate
(1) * <b>Car</b> are my dads biggest hobby (2) * <b>But this</b> are just some of the few things i'm really intrested (sic) (3) *He always <b>talk</b> about cars (4) *My snowboard <b>are</b> really ugly (5) *People <b>walks</b> around in houses saying	(6) *I guess if the pepole (sic) who <b>has</b> been told ... (7) *I <b>has</b> been in Egypt

Table 1 shows some of the many ways to produce subject-verb agreement errors. While example (1) may be a typo, it still demonstrates a case in which the student produces a subject-verb agreement error where the context implies that it is plural while the subject is written as singular. Due to the length of this section, it will be divided into three categories, subject, verb and distance, to elaborately explain and exemplify these language aspects and will work as a section of reference for the analysis. The grammatical background section is divided into these sections for structure, as there are things related to these categories which may cause the students to produce the incorrect verb forms.

### 2.3.1 Subject

While some cases of subject-verb agreement errors may be based on simplistic things (e.g. not using third person singular 's' on 'he/she/it'), some errors are going to be based on more complex rules connected to the subject. This subsection will go through the areas of concern when it comes to the realization of the subject number.

(8) She **likes** cars.                      Sw. Hon **gillar** bilar

(9) We like cars. Sw. Vi **gillar** bilar

(10) The car **is** dirty. Sw. Bilen **är** smutsig

(11) The cars **are** dirty. Sw. Bilarna **är** smutsiga

The examples (8)-(9) demonstrate the concept of subject-verb agreement in its simplest form. Example (8) is a singular subject and thus requires the ‘s’ ending on the verb, while example (9) is a plural subject and requires a plural verb form. An exception to this is the verb ‘be’ which has the ‘am/is/are’ inflections in the present tense, as seen in (10)-(11). When it comes to the subject, there are a couple of concepts that are important to understand in order to produce the correct verb form: (i) uncountable nouns, (ii) indefinite pronouns, (iii) existential clauses, (iv) compound subjects and (v) collective nouns.

Uncountable nouns (e.g. *Money is important*) always require the verb to be in the singular form, as in (12) (Estling Vannestål, 2007: 121). The issue here for Swedish learners of English is that the noun is countable in Swedish but is uncountable and thus treated as singular in English (e.g. Eng. *applause*. Sw. *applåd/er*, Eng. *homework*. Sw. *läxa, läxor*) (Estling Vannestål, 2007: 121).

(12) \*Money are important.

Indefinite pronouns (e.g. *Everyone says that they like to work out*) are special and can be problematic for L2 learners of English. The reason why they become problematic is that certain pronouns, such as ‘everyone’, require a verb in singular form despite it appearing to be a word that describes “a plural concept” (Estling Vannestål, 2007: 349). Indefinite pronouns such as ‘someone’, ‘somebody’, ‘everyone’, ‘everybody’, ‘anyone’, ‘anybody’, ‘no one’ and ‘nobody’ all require the singular form of a verb (Estling Vannestål, 2007: 346, 349). One issue is that words such as ‘none of’ before a plural noun can take both singular and plural forms as they are becoming accepted in both forms which may add to the confusion of understanding what form to use (Estling Vannestål, 2007: 345).

In existential clauses (e.g. *There are ghosts in the attic*) the subject comes after ‘there is/are’ and thus may confuse students as illustrated in (13) (Estling

Vannestål, 2007: 321).

(13) \*But there are a few stupid humans that think there **is** ghosts around us.

Compound subjects require specific knowledge of how noun phrases and coordinating conjunctions affect the number realization of a subject. A compound subject is a subject where two or more noun phrases are connected by coordinating conjunctions such as ‘or’, ‘nor’ and ‘and’. When the subject is connected by ‘and’, the subject is generally plural and thus the verb needs to take a plural form as in (14) (Estling Vannestål, 2007: 128). However, that changes when the subject is realized as a singular unit as it would then require the singular form (e.g. *Just a little peace and quiet is all I ask from you.*) (Estling Vannestål, 2007: 128). It is also singular when the ‘each’ or ‘every’ comes before two nouns connected by ‘and’ (e.g. *Every cookie and cake is delicious*) (Estling Vannestål, 2007: 129). When the subject consists of two noun phrases that are connected by ‘or’ and ‘nor’, the verb takes the form of whatever noun phrase is closest to the verb (15)-(17) (Estling Vannestål, 2007: 129).

(14) The boy and the girl **watch** TV every day.

(15) The boy or the girl **watches** TV every day.

(16) Neither the boy nor the girls **watch** TV every day.

(17) Neither the girls nor the boy **watches** TV every day.

Finally, collective nouns, such as audience and family may be complex because they can be realized as either a single unit (singular) or as a group of individuals (plural) (Estling Vannestål, 2007: 126). It is preferred to use the singular verb form no matter how the noun is meant to be conveyed if the writer uses American English, but it is generally seen as a matter of staying consistent (Estling Vannestål, 2007: 126). These issues may lead to the subject realized as something it is not, and thus the wrong verb form is applied despite knowledge of subject-verb agreement.

### 2.3.2 Verb

In cases where the subject is not complex, the verb may be the area of concern when it comes to subject-verb agreement. Examples (18)-(19) illustrate simplistic subjects that require particular verb inflections.

(18) The car **drives** itself.

(19) The cars **drive** themselves.

In cases (18)-(19), the main difference between Swedish and English is that the verb itself has inflection based on subject number in English while in Swedish it does not. This fundamental difference is why contrastive analysis would suggest that this would be an area for concern for Swedish students learning English as their L2. However, as illustrated by (18)-(19), in the present tense it may seem like the verb is not a cause of concern when learning subject-verb agreement as it appears to be as simplistic as the following rule: if singular, add an 's' to the end of the verb. However, there are complex situations when it comes to verbs that need to be addressed: (i) past tense and (ii) auxiliary verbs

(20) The car **was** dirty.

(21) The cars **were** dirty.

(22) The car **drove** fast.

(23) The cars **drove** fast.

Some verbs in the past tense (20)-(23) may be required to take number into account, as shown by (20)-(21) while some do not (22)-(23). What this means is in cases like (22)-(23) there are no differences between Swedish and English sentence structure and inflection. However, the verb 'be' is a unique case that has number inflection in its past tense inflection as seen in (20)-(21). Due to this unique verb inflection attribute, the verb 'be' is predicted to cause issues for learners of English.

(24) The car **has been** dirty.

(25) The cars **have been** dirty.

(26) \*We [are] going to look in al the expansive shops

Auxiliary verbs may be an area for concern when it comes to subject-verb agreement errors as the same rules for inflection is relevant for auxiliary verbs as for regular verbs. In examples (24)-(26), the auxiliary verb 'have' needs to be correctly used in order to match the number of the subject. There are, however, also cases in which the auxiliary verb has been completely forgotten which causes complications for interpretation of the student's produced sentences (26). The reason why this causes complications for interpretation is that it is impossible, based on the context, to determine if the student knows subject-verb agreement or not (see section 3.3).

### 2.3.3 Distance

This final section will go into the concept of distance and how it relates to subject-verb agreement. In sentences that include a noun phrase with a postmodifier, the student will have to remember the number realization of the subject in order to avoid subject-verb agreement errors. These postmodifiers could, for example, be prepositional phrases or relative clauses that add extra distance between the head of the noun phrase and the verb after the postmodifier (e.g. *the woman on the sofa likes my company* or *the woman that sits on the sofa likes my company*). Just like in existential clauses, the student needs to be aware that distance between the head of the noun phrase and the verb may cause issues if they are not careful.

The reason why this may be difficult for students is because there are several places in which the student needs to remember to use the correct verb form. In longer and more complex noun phrases, the student may simply forget the number realization of the subject, especially if there is a long distance between the head of the noun phrase and the verb (e.g. *\*I think that people who become spirits has something unclarified on the earth*). The reason why this is included as a perspective for this study is to determine if distance between the subject and the verb causes students to frequently produce errors in terms of subject-verb agreement. This distance could be as small as one word, such as between a noun phrase and relative clause, or as long as an entire postmodifier. It is predicted that this will be an issue for students as certain postmodifiers could become complex and difficult to follow when in the process of writing, and it requires quite extensive knowledge of specific grammatical rules in order to not produce errors. Errors that are within a postmodifier will be coded as 'other distance' as the distance is generally one word such as 'that' in *\*they take advantage of*

*people that misses there dead relatives and friends*, while errors that have a postmodifier between the verb and the head of the noun phrase will be coded as ‘postmodifier’ (e.g. \**Especially people in Africa is in a need of my help*). More on the coding decisions in section 3.3. The next section will introduce the method and material used to conduct this study, and it will also introduce the validity and reliability concerns that are associated with error analysis as a methodology.

## 3. Method & Material

### 3.1 Primary Sources

The primary source of the study is the *Uppsala Learner English Corpus* (ULEC) database created by Christine Johansson and Christer Geisler at Uppsala University (Johansson & Geisler, 2009). The corpus includes teacher trainees’ submissions of their students’ texts as a part of the teachers’ degree projects (Johansson & Geisler, 2009). The corpus consists of approximately 136,000 words and has variables suitable for linguistic analysis (Johansson & Geisler, 2009). The variables of relevance to this study are the following: (i) education level (e.g. YR\_1 for Swedish Gymnasium Year 1), (ii) gender, and (iii) age. For this study, the education level and course level are synonymous due to the structure of the English courses. English 5 (listed as A in ULEC) is taken in GY1, and English 7 (listed as C in ULEC) is taken in GY3.

When working with corpora, it is important to keep in mind what questions the data will answer. According to Podesva & Sharma (2013: 274), these questions usually revolve around the frequency and contexts of texts. For the purpose of this study, a corpus like ULEC can help to determine the frequency and conditions of subject-verb agreement errors produced by learners. A common issue when searching for data in corpora is that the results are decontextualized (Podesva & Sharma, 2013: 275). Furthermore, due to the nature of spelling mistakes and other common linguistic errors, searching for a specific verb or word would not yield all the results possible. For example, if a student fails to use the auxiliary verb ‘have’ appropriately in a sentence, and if ‘have’ was used as a search term within ULEC to find cases in which it was used as an auxiliary verb, ULEC would not include any cases where the auxiliary verb is forgotten by the student. However, as ULEC provides the full texts in the database, this will not be an issue due to this study being based on manual error analysis of the

students' essays. Due to this, the study will not run into the issue of results being lost or decontextualized due to spelling mistakes and other linguistic errors.

### 3.2 Selection of Data

There are 117 student essays included in the analysis and the metadata for the texts can be seen in Table 2. The raw information regarding word count for each variable is included for the sake of transparency and to know the data behind the normalized frequencies. When it comes to the genre of the essays, 81 of them were descriptive (44 in GY1 and 37 in GY3) while 17 essays in GY1 were fiction and the final 19 in GY3 were argumentative.

Table 2. Metadata from ULEC

	Female		Male		Total	
	Students	Words	Students	Words	Texts	Words
GY1	29	8148	32	9315	61	17463
GY3	28	8264	28	7468	56	15732

The linguistic data had to be related to subject-verb agreement and thus other issues with verbs were omitted due to the scope of this essay. Examples of omitted data are (27) and (28).

(27) \*I want to be one of the eleven players that can **takes** thousands of fans to the sky. (ULEC Tafreshi\_77)

As shown by (27), the student demonstrates knowledge of third-person singular 's' but fails to acknowledge that when they use modal verbs, the infinitive form is required. While this error may stem from subject-verb agreement rules, this type of error is not included in this study. However, some of the errors will be brought up to exemplify certain aspects that are important for subject-verb agreement acquisition when it comes to interpretation of the linguistic data.

(28) \*a couple of cusins (sic) that **living** in south of Rome. (ULEC AGorman\_52)

An error like (28) could be due to any reason, such as not understanding tenses. While it is impossible to say if it is a subject-verb agreement error, because the auxiliary verb is omitted, the error itself shows a lack of knowledge regarding the inflection of verbs. While the analysis will not focus on this issue, it will list the number of cases in which the auxiliary verb was omitted (see section 3.3).

### 3.3 Method of Analysis

The data was divided into the following clusters of variables: (i) 'subject type', (ii) 'subject number', (iii) 'distance' and (iv) 'verb'. This was done to classify what type of error they were. Within each cluster, there are several linguistic variables that allow the errors to be coded as accurately as possible. Each cluster, and their variables, were analyzed independently and exclusively to avoid inconsistencies in the methodology. The variables in each cluster are (i) common noun, pronoun, existential clause; (ii) singular, plural; (iii) subject-postmodifier, other distance, no distance; (iv) auxiliary verb, regular, irregular. This means that, for example, an error such as 'a singular common noun with a distance between the head of the noun phrase and the regular verb' (e.g. \*A *dog that jump a lot likes you*) would have been categorized as 'singular' in the subject number cluster, 'common' in the subject type cluster, 'other distance' in the distance cluster, and 'regular' in the verb cluster. The clusters will be elaborated at the end of this section. Köhlmyr (2003) has, in her research, used more variables to analyze more aspects of language acquisition, such as counting cases of countable contra uncountable nouns and analyzing other types of subjects and clauses. Due to the time constraint of this study, it focuses on larger grammatical environments to draw more general conclusions. While some of these variables are justified simply due to their importance for understanding subject-verb agreement (see section 2.3), they will be further clarified here in order to justify including those variables into the analysis.

However, due to the methodology of using error analysis to categorize the data, the study will be flawed in the sense that only the errors are categorized under the linguistic variables of this section. To understand how much of an issue, for example, the existential clause is for students, the analysis would have required to consider the times the students used subject-verb agreement correctly. This is a flaw of error analysis and will be discussed further in section 4.4. The errors found during the error analysis were coded according to the linguistic variables listed in this section. Due to the poor

level displayed in some of the essays produced by students, there may be cases of subject-verb agreement missing in the analysis. This is another shortcoming of the method and will be discussed in section 4.4.

The process of error analysis, for this study, could be described as a three-part process. The first step would be to read the students' texts and close-read any cases of subject-verb agreement in order to analyze if it is produced correctly or not. Second, when an error has been found, the full sentence is saved in a database so that it can be closely analyzed and categorized correctly. The third and final part is that the data is evaluated and categorized with the help of the linguistic variables and compared to other data in the same category to ensure that it is evaluated correctly. This section will conclude with an elaboration of the following four clusters of linguistic variables: (i) subject number, (ii) subject type, (iii) distance and (iv) verb.

(i) 'Subject number' is a cluster that refers to whether the produced subject is singular or plural, e.g. \*we **was** there the last time. This example would be coded as 'plural' as the subject is plural. This variable is used to determine if students struggle the most with singular or plural subjects.

(ii) 'Subject type' is a cluster that refers to whether the produced subject is a pronoun, a common noun, a proper noun or existential clause, e.g. \*we **was** there the last time. This example would be coded as 'pronoun' as the subject is a pronoun. Existential clause refers to when the error is shown in an existential clause and is coded as 'yes' or 'no', e.g. \**There is still may things...* This example would be coded as 'yes'. See section 2.3.1 for relevance. This variable helps to determine if any particular subject type caused issues for students.

(iii) 'Distance' is a cluster that refers to whether there is a distance between the head of the noun phrase and either the main verb (in which case it will be classified as a 'subject postmodifier') or the verb (in which case it will be classified as 'other distance'). 'Subject postmodifier' has a post modifier that creates a distance between the head of the noun phrase and the verb. This is coded as either 'yes' or 'no', e.g. *I want to be one of the players that during 90 minutes **unite**...* This example would be coded as 'yes' due to the distance between the subject and verb, in this case, 'one' and 'unite', being a post-modifier. This variable is of relevance to see if the distance between the subject and the verb causes students to produce errors, as discussed in section 2.3.1. 'Other distance' refers to if there is anything else that causes a distance between the head of the noun phrase and the verb. It is coded as 'yes' or 'no', e.g. \**Three things*

*i really wants at the moment is*. This example would be coded as ‘yes’ as there is a distance between the subject and verb which is not a postmodifier. As a variable, it is of relevance for the same reason as the subject post-modifier variable, to see if the distance between the subject and the verb causes errors.

(iv) ‘Verb’ is a cluster that deals with data regarding the verb and is classified as either ‘auxiliary verb’ or ‘regular/irregular verb’. ‘Auxiliary verb’ refers to whether the subject-verb agreement error is caused by inaccurate verb inflection of the auxiliary verb. Auxiliary verb is coded as either ‘yes’, ‘no’ or ‘omitted’. It is coded ‘yes’ if there is an auxiliary verb that causes a subject-verb agreement error, e.g. *\*The house I got most pleased with of those I were working on I would keep*. This is an error as the student does not intend to use the subjunctive mood in this context. It is coded ‘no’ if there is no error. Finally, it is coded ‘Omitted’ if the student omits an auxiliary verb and what this means is that the error will be classified as potentially being a subject-verb agreement error. This is caused by not having access to the student to clarify the intended meaning of the sentence, e.g. *\*one of the best things i [have] done in my life*. The same thing is done in cases in which the subject is omitted, e.g. *\*[missing] likes to go party*. Coding these errors is done for the sake of accuracy as the subject simply does not agree with the verb in the way it is written, as the tense has an impact on subject-verb agreement. ‘Regular’ and ‘irregular’ refers to whether the verb is regular or irregular and is coded as ‘yes’ or ‘no’, e.g. *\*Three things I really wants at the moment...* This example would be coded as ‘yes’ and it is of relevance to see if the students produce more errors with regular or irregular verbs.

### 3.4 Validity and Reliability

This section will go through the validity and reliability of the methodology used in this study, namely error analysis. Daniel (2010: 926) claims that the results of a study are externally valid if “inference can be confidently made from the study’s sample, either to a particular target population or across various populations and settings”. Internal validity, on the other hand, is defined as “a factor of the degree to which research participants’ performance on outcome (dependent) variable(s) is a direct result of the treatment(s) received (independent variables) and not on other factors”. (Daniel, 2010: 927). In other words, the study risks losing internal validity if other factors and variables affect the result, and it risks losing external validity if the results cannot be

applied to other groups. Behar-Horenstein & Dix (2010: 738) define reliability as “the consistency and repeatability of a measurement when the testing procedure is repeated on a population of individuals or groups”. In other words, any concerns regarding reliability would be things that affect the consistency and repeatability of the study and its results. To go into the validity and reliability of this study, this section will discuss any concerns regarding internal and external validity as well as reliability, and what has been done to avoid as much of these concerns as possible.

While the process of error analysis is used frequently in research, it is an interpretation of data based on the author’s knowledge regarding the concept of subject-verb agreement and is thus susceptible to concerns regarding not only internal validity but also reliability. An error may be missed due to complex noun-phrases or implicated errors made by the student (e.g. \*Ghost are not real). In an error like this, one could presume that the student means ‘ghosts’ and thus the subject-verb agreement is in place. However, since the student fails to produce the correct form it will be included as an error as the student still demonstrates the lack of either: (i) plural form of nouns or (ii) third-person singular form of ‘be’. However, due to the nature of close reading and error analysis, there are bound to be missed instances of subject-verb agreement errors. These missed instances may affect the reliability of the results, as another researcher may find missed cases of subject-verb agreement. Another question of reliability is the coding of the data itself, as someone may use a different coding system to define the errors and thus reaches a slightly different conclusion. Not only this, but the human factor of making mistakes due to exhaustion, sloppiness et cetera means that the reliability of the study may be affected. Exhaustion, sloppiness et cetera may be considered variables that affect the internal validity of the study. In order to minimize the risk of these reliability and internal validity concerns, the coding of data was divided into multiple sessions with careful notes of how things were coded to ensure as little variance as possible.

When analyzing the data itself, there were issues causing some concern regarding validity and reliability. Firstly, students’ inability to follow the fundamental rule of subject-verb agreement, namely having a subject that is followed by a verb, caused issues when interpreting data and thus may cause reliability concerns. What was done to avoid this was counting the cases in which the subject or verb was omitted instead of attempting to interpret them. An interpretation of data would cause reliability concerns and this way allows for more consistent results. Secondly, in cases where the student produces the wrong verb form, such as using the progressive form without the

appropriate sentence structure, the mistake was omitted from the study as it is impossible to tell if it is a subject-verb agreement error or not. The lack of access to the students means that it is not possible to get clarification regarding their errors which means that uncertain cases will be completely left out in an attempt to maintain the internal validity and reliability of the study. Thirdly, the limited amount of time allocated to the study caused it to be narrower than anticipated, and more essays from other education levels could have been covered if time was not an issue. Thus, for the sake of validity and reliability, fewer texts were analyzed as manual error analysis takes time to do accurately. If more texts were included and consequently rushed, errors may have been coded inaccurately and would cause the study to lose internal validity and reliability.

A concern for external and internal validity is that the collection of data was not randomized in the sense that the 117 essays were chosen at random. Instead, six different teachers' submissions were selected, each with about twenty essays made by students. Four of the teachers' submissions were chosen at random and two were handpicked to balance the gender representation. Due to the sample size being from six teachers in total, the teacher and environmental variables are a part of the study despite not being covered as a variable. What this means is that if one teacher fails to teach a certain language aspect, it could impact a large portion of the sample size and could thus affect the internal validity of the result. How it may affect the study is that even though the focus lies on the type of errors students make (e.g. *\*I likes cats* and *\*She swim in the...*), students may not be required to write about things that generally cause them to make subject-verb errors under normal circumstances. This is described by Khansir (2012) as the avoidance phenomena. For example, if a student normally demonstrates problems with the pronoun 'it' and fails to use third-person singular 's', and they are told to write what their dream vacation would be, they are likely going to avoid including cases of third-person singular as the text will be from their perspective and thus avoids the problem they have with third-person singular verb forms. A completely randomized selection of essays may have been a better approach in order to remove the teacher and environmental variables, and thus improve the internal validity, but it would also improve the external validity by having a more generalized result. For a more accurate result, it could have been better to decide on a sample size per extralinguistic variable and randomly select texts instead of picking bulks of texts from a teacher. However, the consistency, and thus reliability, of the result is not affected by this as the

same methodology was used for all the essays regardless of its origin.

## 4. Results

The results section is divided into four subsections to go in-depth into the data gathered in this study. First is section 4.1, linguistic variables, and it analyses the data across 117 essays with a total of 33,195 words with only the linguistic variables in mind and does not include any data regarding extralinguistic variables. The second is section 4.2, extralinguistic variables, and it looks at the normalized frequencies of the extralinguistic variables gender and level. Third comes section 4.3, the cross-analysis of the variables, and it analyses the errors across the linguistic and extralinguistic variables. The fourth and final is section 4.4, the discussion, in which there will be a discussion regarding what was uncovered in the study, if there is any criticism regarding the methodology and if the study could have been done differently.

### 4.1 Linguistic Variables

This section will cover the raw frequency of the errors that were produced by students without looking at any extralinguistic variables. Each table introduces linguistic variables from section 3.3 and goes into the data associated with those linguistic variables. Table 3 shows that over half (n=75) of the errors made were done with common nouns as the head of the subject, followed by pronouns with 31.5% (n=41). Proper nouns are not included in the table due to zero errors having proper nouns as the head of the subject and therefore cannot be analyzed. Existential clauses such as ‘*there (is/are)*’ were expected to cause issues due to the complexity of the phrase, and 10.8% (n=14) of the errors made by students were linked to these phrases. The subjects ‘it’ and ‘there’ are not included in the pronouns category if they are a part of an existential clause.

Table 3. Data regarding the subject type cluster.

Pronoun	Common	Existential <i>there</i>	Total
41 (31.5%)	75 (57.7%)	14 (10.8%)	130 (100%)

As illustrated in Table 4, 54.6% (n=71) of the errors were produced when the subjects

are in the plural and only 45.4% (n=59) of the errors occurred with singular subjects. This data is collected by looking at the true subject of a sentence and codes it accordingly. For example, ‘people’ is coded as plural, ‘I’ is coded as singular and ‘There is shows on TV’ is coded as plural to list a few. All cases were looked at as the student had produced it, so cases that may have been typos are considered errors for the sake of consistency (e.g. \**Car are my dads biggest hobby*).

Table 4. Data regarding the subject number cluster.

Singular	Plural	Total
59 (45.4%)	71 (54.6%)	130 (100%)

Table 5 shows the data on the different variables regarding distance between the subject and the verb. The reason why these variables are important when studying subject-verb agreement is that distance between the subject and verb forces the student to remember the realization of the subject to produce the correct verb form. Only 6.9% (n=9) of the errors were cases with postmodifiers between the head of the noun phrase and the main verb, but 23.8% (n=38) of the errors came with some sort of distance between the subject and the verb, such as within a postmodifier. This distance is most commonly a single word, as in (32). Subjects that include a noun phrase with a post-modifier causing distance between the subject and the verb, such as in example (31), could be underrepresented due to the complexity of those phrases.

(31) \*People who say they've seen ghosts and stuff like that **is** probably over reacting.

(sic) (ULEC JSennerby\_58)

(32) \*He also **like** to hang out with her friends (ULEC K-Staaf\_88)

Table 5. Data regarding the distance cluster.

Subject-Post Mod	Other Distance	No Distance	Total
9 (6.9%)	31 (23.8%)	90 (69.2%)	130 (100%)

As illustrated by Table 6, Irregular verbs proved to be the most problematic with 59.2% (n=77) of the cases being due to irregular verbs, while 32.3% (n=42) were regular verbs. The remaining 8.5% (n=11) were errors in which the wrong auxiliary verb was produced by the students and caused errors like example (33).

(33) I guess if the pepole (sic) who **has** been told what's gonna happen to their life  
(ULEC JSennerby\_58)

Table 6. Data regarding the verb cluster.

Auxiliary Verb	Regular	Total
11 (8.5%)	Yes: 42 (32.3%), No: 77 (59.2%)	130 (100%)

The *subject omitted* and *auxiliary verb omitted* variables were not originally planned to be included. However, they were included in Table 7 to illustrate the cases in which the student has failed to include the subject in the sentence or auxiliary verb and thus fails to display an understanding of subject-verb agreement as a concept. This may be considered as something outside of the scope of the essay, but it was deemed worth mentioning due to the fundamental concept on which the subject-verb concord rests upon, in which every sentence contains a subject and a verb. More on this dilemma in section 4.4.

Table 7. Data regarding omitted items

Subject Omitted	Aux. Verb Omitted
15	20

## 4.2 Extralinguistic variables

This section is dedicated to raw data and the normalized frequencies of the extralinguistic variables gender and education level. The normalized frequency was calculated with the following formula:  $\frac{\text{number of cases in that category}}{\text{total words in that category}} * 1000$ , so for example for 'female' in the gender subsection it is  $\frac{63}{16412} * 1000 = 3.838654643 \approx 3.839$ . While the focus of this section is to present the raw and normalized frequencies across the extralinguistic variables, it will also briefly mention what was expected out of the analysis contra what was uncovered to be used as a foundation for section 4.3 and 4.4

The data on gender in Table 8 shows the raw and normalized frequencies of the study. The frequencies for male and female students are expected to be even based on the statistics provided by the Swedish National Agency for Education (Skolverket) and the results presented by PISA. If this was an international study, as in

not based on Swedish learners of English, it would have been expected that the female students perform significantly better. The data presented in Table 8 roughly describes what was expected based on previously mentioned statistics from Skolverket. Males and females have almost equal normalized frequencies, but it appears that the advantage females have in reading comprehension does not translate into a much better understanding and production of subject-verb agreement. It also resonates with the general understanding that females are better academically than males.

Table 8. Raw and normalized frequencies across gender

<b>Female</b>		<b>Male</b>		<b>Total</b>	
Raw	Norm.	Raw	Norm.	Raw	Norm.
63	3.839	67	3.992	130	3.916

The raw and normalized frequencies for the educational level are listed in Table 9. It is expected that the students in GY1 produce more errors than the students in GY3 and thus GY1 should have a higher normalized frequency. However, as illustrated in Table 9, the students' ability to accurately produce subject-verb agreements only improves slightly according to the normalized frequencies for both GY1 and GY3. This phenomenon will be further analyzed in a cross-variable analysis in section 4.3 and in 4.4, in which there will be an attempt to identify the reasons why students in GY3 continue to produce subject-verb agreement errors despite having more knowledge compared to the students in GY1.

Table 9. Raw and normalized frequencies across education level

<b>GY1</b>		<b>GY3</b>		<b>Total</b>	
Raw	Norm.	Raw	Norm.	Raw	Norm.
69	3.951	61	3.877	130	3.916

In Table 10, the data is compared across both gender and education level to see how the students perform in GY1 and GY3 based on their gender. This table should show that male and female students produce fewer errors the further they get in their education, meaning that both genders should have a lower normalized frequency of errors in GY1 compared to GY3. However, as we can see in Table 10, females are improving significantly while males worsen to the same degree when looking at the normalized frequencies with level and gender as variables. This will be discussed more in-depth in

section 4.4, but it can be said that it was an unexpected result as male students get significantly worse in GY3 compared to females who get significantly better.

Table 10. Raw and normalized frequencies across gender and education level

	Female		Male		Total	
	Raw	Norm.	Raw	Norm.	Raw	Norm.
GY1	42	5.155	27	2.899	69	3.951
GY3	21	2.541	40	5.356	61	3.877

### 4.3 Cross-analysis of the variables

This section is dedicated to comparing the results across education level and gender and provides examples from the data to illustrate the differences between GY1 and GY3 in the production of subject-verb agreement errors. Each table is based on the gender and education level variables, but they are analyzed based on the binary outcomes ‘male’ and ‘female’ for gender, and ‘GY1’ and ‘GY3’ for education level.

As shown in Table 12, in GY1 the students struggle equally with common nouns (n=31) and pronouns (n=31). Most of these errors are relatively simple ones, such as not understanding third-person singular for *he*, *she* and *it*, and not plural forms for *I*, *you*, *we*, and *they* (33)-(36).

(34) so she can [do] whatever she **like** to do. (ULEC AGorman\_52)

(35) I should buy a dog becuse i **likes** dogs mutch (sic). (ULEC AGorman\_52)

(36) he **know** me pretty well (ULEC CHermmodsson\_66)

The area of difficulty changes for GY3 students, as common nouns are dominantly the issue for them at 70.5% (n=43). Their issues, however, tend to be with common nouns such as ‘people’ (37) and more complicated subjects with coordinating conjunctions (38). Existential *there* (*is/are*) does not appear to improve from GY1 (n=7) to GY3 (n=7), although this may be explained due to the frequency of use and complexity of the phrase.

(37) don't judge people who **believes** in ghosts. (ULEC JSennerby\_58)

(38) ghost and many other things that are impossible to explain **is** something that always will fascinated people. (ULEC JSennerby\_58)

The gender variable suggests that female students struggle equally with pronouns (n=23) as they do with common nouns (n=34), while male students struggle more with common nouns (n=40) than pronouns (n=19). Due to how the data was categorized, there is no information regarding if any particular nouns (e.g. *people* or *everyone*) were more problematic for males or females, which is an oversight of the methodology and could have added another layer to the analysis.

Table 12. Data regarding the subject type cluster across the extralinguistic variables.

	Pronoun	Common	Existential there	Total
GY1	31 (44.9%)	31 (44.9%)	7 (10.1%)	69 (100%)
GY3	11 (18%)	43 (70.5%)	7 (11.5%)	61 (100%)
Female	23 (36.5%)	34 (54%)	6 (9.5%)	63 (100%)
Male	19 (28.4%)	40 (59.7%)	8 (11.9%)	67 (100%)

Table 13 illustrates that there are no real differences between the extralinguistic variables. The percentages for 'other distance' vary only between 23.2% to 26.2%, and 'no distance' vary between 63.9% and 72.5%. In terms of subject-verb agreement, it appears that students in GY1 make fewer errors related to subject-verb agreement (4.3%, n=3) compared to students in GY3 (9.8%, n=6). This could be because of the complexity of postmodifiers and that students in higher education levels attempt to use them more frequently compared to the students with a lower education level. In terms of the gender variable, out of the nine recorded cases of subject-verb agreement errors associated with postmodifiers, female students made seven of them.

Table 13. Data regarding the distance cluster across the extralinguistic variables.

	Subject-Post Mod	Other Distance	No Distance	Total
GY1	3 (4.3%)	16 (23.2%)	50 (72.5%)	69 (100%)
GY3	6 (9.8%)	16 (26.2%)	39 (63.9%)	61 (100%)
Female	7 (11.1%)	15 (23.8%)	41 (65.1%)	63 (100%)
Male	2 (3%)	17 (25.4%)	48 (71.6%)	67 (100%)

Table 14 illustrates the difference in errors based on subject number and the only relevant difference here is that GY3 struggles less with singular subjects at 34.4% (n=21) compared to GY1 that has more of an issue with singular subjects at 55.1% (n=38). Plural subjects appear to be the leading issue for subject number for GY3 at 65.6% (n=40) while 44.9% (n=31) of the errors were plural subjects for GY1. The gender variable shows that the difference between plural and singular subjects is shifted towards plural for male students as they have the largest difference between the two variables with plural subjects at 61.2% (n=41) and singular subjects at 38.8% (n=26). Female students, however, seem to struggle almost equally with plural subjects at 47.6% (n=30) as with singular subjects at 52.4% (n=33).

Table 14. Data regarding the subject number cluster across the extralinguistic variables.

	Singular	Plural	Total
GY1	38 (55.1%)	31 (44.9%)	69 (100%)
GY3	21 (34.4%)	40 (65.6%)	61 (100%)
Female	33 (52.4%)	30 (47.6%)	63 (100%)
Male	26 (38.8%)	41 (61.2%)	67 (100%)

Table 15 shows that there are no significant differences when it comes to auxiliary verb errors, nor the distance between the subject and the verb. There is also a minor shift towards regular verbs being an issue for GY3 students at 34.3% (n=24) compared to GY1 students at 24% (n=18).

Table 15. Data regarding the verb cluster across the extralinguistic variables

	Auxiliary Verb	Regular	Total
GY1	5 (6.7%)	Yes: 18 (24%), No: 52 (69.3%)	69 (100%)
GY3	6 (8.6%)	Yes: 24 (34.3%), No: 40 (57.1%)	61 (100%)
Female	6 (8.2%)	Yes: 16 (21.9%), No: 51 (69.9%)	63 (100%)
Male	5 (6.9%)	Yes: 26 (36.1%), No: 41 (56.9%)	67 (100%)

Table 16 shows that the rate of omitted subjects increases from 6 in GY1 to 9 in GY3, while the omitted auxiliary verbs almost completely disappear in GY3 (n=1) compared to GY1 (n=19). In both cases, females produce more cases of omitted subjects and

auxiliary verbs than their male counterparts.

Table 16. Data regarding omitted items across the extralinguistic variables

	Subject Omitted	Aux. Verbs Omitted
GY1	6	19
GY3	9	1
Female	10	11
Male	5	8

#### 4.4 Discussion

The research questions of this study are focused on looking at linguistic and extralinguistic variables to see in which environments students produce subject-verb agreement errors. The expectations were that students would improve over time due to them having more exposure to the language. That proved to be somewhat accurate. To analyze this thoroughly, this section will be divided into four main sections: (i) extralinguistic variables, (ii) omitted subjects and auxiliary verbs, (iii) postmodifiers and distance and (iv) methodology.

Table 10 shows that level is a smaller factor than perhaps anticipated. The students produced practically equal amounts of subject-verb agreement errors, although different in nature, in GY1 compared to GY3 when looking at the normalized frequencies for those levels of education. GY1 made more mistakes of simplistic nature, such as not knowing the inflection for the personal pronouns (33)-(35). GY3, on the other hand, made a mix of errors, ranging from the same as GY1 (33)-(35) to more complex ones that stem from not understanding, for example, the number realization of specific words and compound subjects (36-37). This concept appears to be true for other linguistic variables as well, such as the omission of auxiliary verbs almost disappearing in GY3 compared to GY1. One plausible explanation is that students in GY3 make fewer errors in simple scenarios, but instead attempt to write more elaborate and complicated sentences which causes them to get exposed to language rules that they have not faced before. One reason behind more simplistic errors, regardless of level, could be that Swedish students learning English as their L2 struggle with defining rules for when to use third-person singular 's' (Källkvist & Peterson, 2006: 125). In a study by Källkvist & Peterson (2006: 125), the students inaccurately describe the rules for

third-person singular as follows: “Get with the singular. Gets with the plural”. In the same study, the students’ descriptions of the different forms of ‘be’ (was/were, is/are) were only inverted like with ‘get/gets’ once. The rate of mistakes reflects upon this as 41-46% of the participants made errors with ‘get’ while the rate of mistakes for ‘was/were’ 88-77% and ‘is/are’ were 76-77% (Källkvist & Peterson, 2006: 122). While the rate of mistakes in Källkvist & Peterson’s study cannot be comparable to this study due to how the data is presented, it does raise an interesting question. Do students simply not understand the concept of subject-verb agreement and how it functions? It could be that they simply learn structures, such as ‘it should be *you are* and *he is*’ but not because of the subject being realized as singular or plural.

Putting level aside for a moment to talk about the other extralinguistic variable, the data on gender provided an interesting difference between male and female students. According to grade statistics from Skolverket, male and female students should have near equal performance when it comes to English, and Table 8 illustrates that it appears to be the case, although there is a small difference in favor of female students. Furthermore, Table 9 shows that female students improve from GY1 to GY3, going from a normalized frequency of 5.155 to 2.541 while male students drastically worsen, going from 2.899 to 5.356. This aspect of language acquisition could be something worth pursuing to see what causes males and females to have such a significantly different trend in terms of becoming more proficient with the language, but for this study, it is enough to note that there is a measurable difference of improvement between males and females.

The inclusion of omitted subjects and auxiliary verbs in the analysis could be considered out of the scope of the essay. There were 15 omitted subjects and 20 omitted auxiliary verbs in total and these errors might not be considered errors of subject-verb agreement. While they may stem from a lack of understanding that a sentence needs a subject and a verb, and that auxiliary verbs need to be conjugated according to tense and number, the sentences are simply incomplete and are thus difficult to analyze appropriately if they are included. The difference between omitted auxiliary verbs and subjects is that every sentence needs a subject, something which should have been taught especially when teaching students about subject-verb agreement, while auxiliary verbs are needed, in this case, to convey tense and number, e.g. *she has been* contra *she have been*. Students may not recognize that ‘have’ needs to be conjugated according to subject-verb agreement rules even when used as an auxiliary

verb. According to Table 12, omitted subjects became more of an issue in GY3 while errors with omitted auxiliary verbs practically disappear in the same education level. This appears to be inconsistent and could be explained as a statistical anomaly as according to the normalized frequencies of errors produced in GY1 and GY3, these errors should both lessen as the students improve.

Postmodifiers and distance were predicted to be issues for students due to them needing to remember the number realization of the true subject. While a lot of errors were within a postmodifier, such as within a relative clause or prepositional phrase, few cases of subject-verb agreement errors may have occurred as a result of a postmodifier being between the head of the noun phrase and the verb. In total, 6.9% of the errors (n=9) were cases in which the verb after a postmodifier was inaccurate according to subject-verb agreement. 23.8% of the errors (n=31) were cases in which there was a distance, not including postmodifiers, between the head of the noun phrase and the verb. In other words, in 30.7% (n=40) of the errors included in the error analysis, distance was a factor when looking at subject-verb agreement errors. The fact that there are only nine cases of subject-verb agreement errors with a postmodifier between the head of the noun phrase and the main verb is surprising. It may be due to the students actively avoiding using postmodifiers due to it being too complicated for them. However, without knowing how many times students used postmodifiers and including the cases in which it was done correctly, it is impossible to reach a clear conclusion on the matter.

To conclude this section, the methodology of this study will be evaluated and discussed regarding how it affected the results of this study and what could have been done differently. Error analysis as a methodology can be considered flawed for studies like this one in the sense that only errors are analyzed and the “avoidance phenomenon” is not included (Khansir, 2012). The avoidance phenomenon is described by Khansir (2012) as a “learner strategy of avoiding what is difficult” and it could explain certain aspects of this study, such as the low frequency of ‘existential there’ phrases. The results in this study could be heavily affected by this, as many students used phrases such as ‘I would go to town and she would come with me’ where *would* removes the requirement for subject-verb agreement. By avoiding what is difficult, it is almost impossible to get a true perspective of what the students know apart from what they do correctly in an essay. That leads to the other criticism that error analysis focuses on errors and not the rest of the information in the corpus (Khansir, 2012). Being unable

to see how often certain variables are correct and when they are incorrect means that it is impossible to understand how much of an issue it is. For example, in the case of ‘existential there’ phrases, what is the actual error rate for those phrases? As illustrated by Table 3, there are 14 cases of ‘existential there’ phrases that are done inaccurately, but how many times did the students produce them with no errors? Is it a big problem, or a small one? This level of analysis is impossible with error analysis as the main methodology as vital information is left out. Furthermore, Khansir also argues that the classification of errors is done improperly and that the material used is selected in a biased way. The classification of errors is the largest concern for a study like this, as despite being as thorough as possible, there may be things that are overlooked or misclassified. However, that is more of a concern for the linguistic variables more so than extralinguistic ones. Furthermore, any mistakes in the categorization of subject-verb agreement errors would only slightly affect the percentages of the linguistic variable categories, meaning that the conclusion of this study should be sound with the possibility that the percentages may vary slightly. If this study was done again, two things would have been done differently, at least in terms of methodology. The first thing is that all cases of subject-verb agreement would have been categorized, correct and incorrect, to include how often the linguistic variables were produced accurately and consequently inaccurately. The second thing that would have been included is more subcategories to analyze things such as the error frequency of countable contra uncountable common nouns to try and explain why grammatical environments such as ‘common nouns’ were as problematic as they were.

## 5. Conclusion

The study aimed to answer the following research questions:

- Which grammatical environments produce the highest error rate in terms of subject-verb agreement errors?
- Are there differences across education levels and gender in terms of subject-verb agreement error rate (i.e. do the students improve across level)?

The study did show that the grammatical environments that caused the highest

frequency of errors for students as a collective were *common nouns* 57.7% (n=75), *plural subjects* 54.6% (n=71), *irregular verbs* 59.2% (n=77) and *distance* 30.8% (n=40). Each of these grammatical environments were the most problematic in their cluster of linguistic variables. The clusters were analyzed independently of one another to avoid inconsistencies in methodology (see section 3.3). The ‘distance’ grammatical environment is interesting as the data suggests that students often struggle whenever there is any distance between the head of the noun phrase and the verb. Out of the 40 cases of distance being a factor, only nine errors occurred when a whole postmodifier was between the subject and the verb. When using the extralinguistic variables to cross-analyze the data, common nouns proved to be a consistent issue across both gender and education level. The ‘subject number’ cluster, however, was not consistent across gender and education level, as plural subjects caused more issues for GY3 (65.6%) and male students (61.2%), while singular subjects caused more issues for GY1 (55.1%) and female students (52.4%). Research on gender as a variable in L2 acquisition suggested that females would do better than males, but the performance difference between female and male students was expected to be rather equal due to it being a study conducted on Swedish learners of English based on the recent statistics provided by Skolverket. The data did show that female students were not only producing fewer errors, but they also improved significantly from GY1 to GY3 while male students produced almost twice as many errors in GY3 compared to GY1. The results across education level as a variable were rather similar. While the normalized frequency suggests that the students in GY1 make 3.951 errors compared to the 3.877 produced by students in GY3, it is such a minor difference that it could simply be down to the data selected. However, close-analyzing the errors produced by the students in these levels, it was observed that students in GY1 made errors that were more simplistic in nature, like not adding the third person singular ‘s’ to ‘he’, ‘she’ and ‘it, while GY3 had issues with the number realization of specific common nouns such as ‘people’.

## References

### Primary References

Student essays collected from ULEC in the following files:

Gy\_1\_S\_AGorman\_52.txt

Gy\_1\_S\_CHermodsson\_66.txt

Gy\_1\_S\_MarBlom\_64.txt  
Gy\_3\_S\_JSennerby\_58.txt  
Gy\_3\_S\_K-Staaf\_88.txt  
Gy\_3\_S\_Tafreshi\_77.txt

## Secondary References

- Behar-Horenstein, L. S., & Dix, A. C. (2010). 'Reliability' in Kridel, C (ed.). *Encyclopedia of Curriculum Studies*. Los Angeles: Sage Publications, Inc. 738-739.
- Daniel, L. G. (2010). 'Validity, External/Internal' in Kridel, C (ed.). *Encyclopedia of Curriculum Studies*. Los Angeles: Sage Publications, Inc. 925-926.
- Estling Vannestål, M. (2007). *A University Grammar of English with a Swedish Perspective*. Lund: Studentlitteratur
- Harmer, J. (2007). *The Practice of English Language Teaching (4. [rev.] ed.)*. Harlow: Longman.
- Jackson, C. N, Morner, E. & Brehm, L. (2018). 'The Production of Subject-Verb Agreement Among Swedish and Chinese Second Language Speakers of English'. *Studies in Second Language Acquisition*; 40(4): 906-921.
- Johansson, C & Geisler, C. (2009). 'The Uppsala Learner English Corpus: A new corpus of Swedish high school students' writing' in Saxena, A., Viberg, Å. (eds.). *Multilingualism: proceedings of the 23<sup>rd</sup> Scandinavian Conference of Linguistics* Uppsala: Acta Universitatis Upsaliensis, 181-190.
- Khansir, A. A. (2012). 'Error analysis and Second Language Acquisition' in *Theory and practice in Language Studies*; 2(5):1027-1032.
- Källkvist, M., Petersson, S. (2006) 'An s, or not an s; that is the question. Swedish teenage learners' explicit knowledge of subject-verb agreement in English' in Einarsson, J., Larsson-Ringqvist, E., and Lindgren, M. (eds.). *Språkforskning på didaktisk grund*. Uppsala: Svenska föreningen för tillämpad språkvetenskap. 112-140.
- Köhlmyr, P. (2003). *To Err Is Human ... An Investigation of Grammatical Errors in Swedish 16-year-old Learners' Written production In English*: doctoral thesis from Göteborg University, Göteborg.
- Lightbown, P.M. & Spada, N. (2013). *How Languages Are Learned*. (4. ed.) Oxford: Oxford University Press.
- Mitchell, R., Myles, F. & Marsden E. (2013). *Second Language Learning Theories (3<sup>rd</sup>*

edition). London & New York: Routledge.

Podesva, R. J. & Sharma, D. (rev.) (2013). *Research Methods in Linguistics*. Cambridge: Cambridge University Press.

Skolverket. (2011).

<https://www.skolverket.se/download/18.4fc05a3f164131a74181056/1535372297288/English-swedish-school.pdf>

Skolverket. (2020). Kursbetyg i Gymnasiegemensamma ämnen läsåret 2018/19 – Engelska.

[https://siris.skolverket.se/siris/sitevision\\_doc.getFile?p\\_id=549431](https://siris.skolverket.se/siris/sitevision_doc.getFile?p_id=549431)

PISA. (2018). <https://www.skolverket.se/publikationsserier/rapporter/2019/pisa-2018.-15-aringars-kunskaper-i-lasforstaelse-matematik-och-naturvetenskap>

van der Slik FWP, van Hout RWNM & Schepens JJ. (2015). *The Gender Gap in Second Language Acquisition: Gender Differences in the Acquisition of Dutch among Immigrants from 88 Countries with 49 Mother Tongues*. PLoS ONE 10(11): e0142056.  
<https://doi.org/10.1371/journal.pone.042056>

Wei, W., Chen, B., Lian, L & Dunlap, S. (2015). 'Native language influence on the distributive effect in producing second language subject–verb agreement' in *Quarterly Journal of Experimental Psychology*; 68(12): 2370-2383.