



FACULTY OF EDUCATION AND BUSINESS STUDIES
Department of Humanities

Subject-Verb Agreement Errors in Swedish 9th and 11th Grade Students' English Written Production

Maiko Tsukanaka

2023

Student thesis, Bachelor degree, 15 HE
English

ENG805 English 61-90

Supervisor: Henrik Kaatari, PhD
Examiner: Kavita Thomas, PhD

Abstract

This study aims to investigate possible factors contributing to subject-verb agreement errors in Swedish junior and senior high school students' English written production. The sample data is collected from the Swedish Learner English Corpus (SLEC), which comprises student texts produced in a classroom setting. The texts are randomly chosen but evenly distributed in terms of binary gender, school year, and type of high school program. In this study, the texts included in the scope are written by students attending a Swedish-speaking school with Swedish as their first language.

Errors are classified as overgeneralization or transfer and further classified in relation to the subject type, the verb type, and the distance between the subject and the verb. The classification of all correct instances of subject-verb agreement is also performed to further investigate possible error explanations. A total of 41 agreement errors were found in 24 texts written by students in the 9th and 11th grades. The results show that overgeneralization is more frequent than transfer errors. Overgeneralization suggests that the students are aware of the third-singular form but do not always apply it correctly, while transfer errors show a potential lack of awareness or attention to the form. In both cases, the errors indicate that these students have not automatized the principle yet. Errors are often related to subject types "a pronoun" or "a noun/noun phrase" and the verb *be*, which is the most frequently used verb. Most of the errors occur when the subject and the verb are in immediate contact, and more than half of them involve a relative pronoun as subject, which indicates that the learners have misinterpreted the grammatical principle or have not fully acquired it. Overuse of the third-person singular form can also be an effect of teaching and explicit learning, which makes learners apply the form whenever it seems possible and relevant.

Keywords: Second Language Acquisition (SLA), Swedish Learner English Corpus (SLEC), Swedish-speaking learners of English, Error analysis, Subject-verb agreement, Transfer, Overgeneralization

Table of Contents

1. Introduction	4
1.1 Aim and Research Questions	5
2. Theoretical Background	5
2.1 Error Analysis in Second Language Acquisition	5
2.2 Cross-Linguistic Transfer	7
2.3 Overgeneralization	9
2.4 Subject-Verb Agreement	9
3. Method	10
3.1 Material	11
3.2 Error Classification	12
3.3 Method for Data Analysis	13
3.4 Validity and Reliability	14
4. Results and Discussion	16
4.1 General Results	17
4.2 Overgeneralization and Transfer Errors	19
4.3 Subject Types	21
4.4 Types of Verbs	24
4.5 The Distance Between the Subject and the Verb	26
5. Conclusion	29
References	30
Primary Sources	30
Secondary Sources	31
Appendices	32
Appendix I	32
Appendix II	33

1. Introduction

In Sweden, English has long been the most important foreign language studied, and it is compulsory in elementary and upper secondary education. English is also widely used in Sweden, both in higher education and at workplaces as the official working language. The younger generation comes into contact with the English language in their daily lives, mostly by using various digital social platforms, which also enable natural contacts among people with different language backgrounds, where the mutual communication language is English.

Ringbom (2007, p. 6) claims that cross-linguistic similarity is relevant for the comprehension of a new language. For example, there are usually cognates across related languages that are similar in form and meaning. The learner will also experience fewer problems when there is grammatical congruence between the native language (NL) and the target language (TL), and an English learner of Swedish will find a great similarity between the noun morphology of Swedish and English (Ringbom, 2007, p. 6). Ringbom (2007, p. 107) argues that Swedish-speaking English learners can make full use of the cross-linguistic similarities between English and Swedish at early stages, and even though one-to-one correspondences in item learning are often oversimplified, they provide a useful start for the learners. After an elementary knowledge of English grammar has been achieved, Swedish-speaking learners can make effective use of implicit learning through language examples (Ringbom, 2007, p. 107), and such similarity should facilitate the learning of English for many native speakers of Swedish. This is also reflected in the general fluency in English among the Swedish population, which places the nation as one of the highest ranked in the assessed knowledge of English as a foreign language internationally (Breene, 2019). However, there are also some significant differences that can lead to learner errors in their language production in the TL, and one of them is the inflection of verbs in the third-person singular. This type of error is claimed to be most common among Swedish learners of English (Köhlmyr, 2003, p. 172), and there are different factors behind this, despite the relative simplicity of the rule of the inflection of verbs in the third person. Errors are a part of language learning, and the aim of this study is to investigate possible background factors for subject-verb agreement errors, which can hopefully provide language teachers with a notion of how learners process input and make use of it.

1.1 Aim and Research Questions

The aim of the study is to examine and better understand possible factors contributing to errors in Swedish junior and senior high school students' English written production. The central research questions of the essay are the following:

1. When classified into intralingual (overgeneralization) or interlingual (transfer) errors, which type of subject-verb agreement errors are most common among the students in their writing production?
2. To what extent are these errors related to the types of subjects, verbs, or the distance between the subject and the verb?

The target group of this study is junior and senior high school students with Swedish as their school language and as their first language (L1). Their written production in English as the TL will be studied with a focus on subject-verb agreement errors.

2. Theoretical Background

In this section, a literature review covering previous research and other relevant background for the study is given. First, previous studies of error analysis within second language acquisition (SLA) research are discussed, followed by studies of cross-linguistic transfer in learner language. Furthermore, the concept of subject-verb agreement is explained.

2.1 Error Analysis in Second Language Acquisition

In the 1950s and early 1960s, second language (L2) learning theory was seen as an accompaniment to the practice of language teaching. In this era, learning was seen as acquiring a set of appropriate speech habits, and it was important to build courses of instruction around a structural syllabus to ensure systematic step-by-step progress (Howatt, 2004, as cited in Mitchell et al., 2019, p. 40). The general learning theory at the time was thus related to mainstream psychology and behaviorism (Mitchell et al., 2019, p. 40). From this point of view, the process of learning a first language (L1) should simply be about learning a set of new habits as learners respond to stimuli in their environment (Mitchell et al., 2019, p. 41). However, this would mean that this set of habits needs to be replaced by a set of new ones when learning an L2, and the old L1 habits can interfere with this process; if the L2 structures are similar to those of the L1, this will facilitate learning, but if the structures are different, they will make learning more difficult (Mitchell et al., 2019, p. 41). These beliefs led to the idea of focusing on areas of difference, and in the 1970s, researchers such as Daniel Slobin (1970, 1985, as cited in Mitchell et al., 2019, p. 43) and Roger Brown (1973, as cited

in Mitchell et al., 2019, p. 43) conducted Contrastive Analysis to compare pairs of languages to identify areas of difference in the learning process. They found similarities in the L1 learning behavior of young children, regardless of the language they were learning, and Brown's morpheme studies on child L1 learning became influential for SLA research.

Morpheme studies within L2 acquisition were based on the work of Brown (1973, as cited in Mitchell et al., 2019, p. 44) in L1 acquisition mentioned above, which traced several children's development of 14 grammatical morphemes in L1 English. With the hypothesis of the similarity between child L1 and SLA development (Gass et al., 2020, p. 114), a series of L2 morpheme studies were conducted in the 1970s to investigate what L2 learners' acquisition of the same grammatical morphemes looked like. Studies by Dulay and Burt (1973, 1974, 1975, as cited in Mitchell et al., 2019, p. 48) analyzed the speech of children with Spanish or Chinese as their L1 acquiring English as their L2, and their acquisition sequences showed a significant similarity to each other despite their different levels of L2 proficiency. Further studies showed that this also applied to adult learners of English and that both child and adult learners of L2 English developed their accuracy of these grammatical morphemes in a similar order, independent of the learning situations (Mitchell et al., 2019, p. 49). Similarities between L1 and L2 can lead to overproduction of L1 function on the L2 form (Gass et al., 2020, p. 141) and affect how quickly learners acquire different features of the L2 (Gass et al., 2020, pp. 137-138). The English word "hello," for example, is a form with a function to greet an acquaintance or answer the phone (Kroeger, 2005, p. 1). The form conveys the speaker's wish to acknowledge the presence of or initiate a conversation with the hearer, and speakers of English know how to use it (Kroeger, 2005, p. 1). Swedish has a form equivalent to the English hello, and when both formal and functional similarities can be established, this makes for positive transfer (Ringbom, 2007, p. 6), but it is not certain that other languages have the same form to achieve the same function; a particular form may be used for different functions, depending on the context (Kroeger, 2005, p. 2).

Similarities can also differentiate the acquisition paths dependent on the learner's L1 (Gass et al., 2020, p. 138), while differences between L1 and L2 can lead to the avoidance of certain linguistic features of L2 (Gass et al., 2020, pp. 135-136). When the structural difference between L1 and L2 is significant, this difference would make learning difficult (Mitchell et al., 2019, p. 40). Contrastive Analysis is based on this hypothesis, which aims to identify and predict possible difficulties for L2 learners through a systematic comparison between the L1 and L2 structures. However, not all errors can be explained by and connected to the learner's L1, and similarities and differences are not always related to difficulties in L2

learning (Mitchell et al., 2019, p. 47). Difficulties that occur in one language do not always occur in the other, and at the same time, areas where the L1 should have prevented errors are not always free from them (Mitchell et al., 2019, p. 47).

Error analysis focuses on the language output produced by L2 learners and the systematic investigation of their errors, while the Contrastive Analysis hypothesis has a focus on the difference between the L1 and target L2 and aims to predict possible learning difficulties (Mitchell et al., 2019, pp. 47, 85-86). When a linguistic element is new for a learner, this new element can lead to an intralingual error, which is not a transfer from the learner's L1 (Gass et al., 2020, p. 87). An intralingual error is attributed to the TL and is not dependent on the learner's NL. The second type of error is of an interlingual character, which can be related to the learner's NL (Gass et al., 2020, p. 87). When a learner does not make the same type of error consistently, it can also be defined as a mistake (Gass et al., 2020, p. 86). A weakness of the error analysis method is the difficulty in determining what type of error it is (Gass et al., 2020, p. 89) and the fact that the L1 features will not always result in transfer errors, which will be discussed in Section 2.2.

2.2 Cross-Linguistic Transfer

Discussions of language transfer began with the work of American linguists in the 1940s and 1950s (Odlin, 1989, p. 6), but its role in SLA has long been a controversial topic; some scholars consider it an important factor in SLA, while others have been skeptical (Odlin, 1989, p. 3). Despite the behaviorist notion of transfer, which often implies the extension of earlier habits, Odlin (1989, p. 25) argues that transfer is not a consequence of habit formation and that the acquisition of a second language does not require this. Traditionally, the term *interference* has been used as a synonym to *transfer*, and the notion of interference can be applicable in the description of some aspects of L2 performance, such as pronunciation inaccuracies (Odlin, 1989, p. 25). However, the negative attitude toward transfer gradually changed in the 1980s (Ringbom, 2007, p. 33), and since much of the influence from one's L1 and other known languages can give an advantage in SLA, Odlin (1989, p. 26) suggests that the terms *positive* and *negative transfer* are to be preferred. Even in more recent research, transference from the L1 can be seen as an enrichment of the language, and the term *positive transfer* or *crossover* is used as opposed to *interference* or *negative interference* (Modiano, 2009, p. 138). Ringbom (2007, p. 56) points out that the term 'crosslinguistic influence' might be more appropriate as nothing is actually transferred; however, in this study, the term transfer is mainly used.

Odlin (1987, p. 36) classifies several different types of effects that cross-linguistic similarities and differences can produce: positive transfer, negative transfer, and differing time required for acquisition. Cross-linguistic similarities between the NL and the TL, such as syntax and phonological similarities, can produce positive transfer in several different ways (Odlin, 1987, p. 36). This can be applied to English learners with Swedish as their L1, but as the effects of such positive transfer can only be determined through comparisons of the success of groups with different native languages (Odlin, 1987, p. 36), they will not be investigated further in this essay. Negative transfer, on the other hand, is relatively easy to identify as it involves divergences and deviations from the norms in the TL (Odlin, 1987, p. 36). Generally, negative transfer is seen as production errors, but they can also be studied from the perspective of how a learner's L2 performance differs from the native speakers' behavior, for instance, underproduction, overproduction, production errors, and misinterpretation (Odlin, 1987, pp. 36-38). Avoidance of structures in TL that are not present in one's NL can be connected to both underproduction and overproduction; the latter occurs when learners avoid a certain English construction and overproduce other constructions instead as a result of avoidance, while substitutions by borrowing forms from the NL, such as direct interpretations and alternations of structures, can be defined as production errors (Odlin, 1987, pp. 36-37).

The Universal Grammar approach (UG) argues that children must have some innate core of abstract knowledge about language form that guides them in L1 acquisition unless they have learning disorders (Mitchell et al., 2019, pp. 12, 83). There are different views on UG and its relevance to L2 acquisition: L2s as not UG-constrained or L2s as UG-constrained with full or partial access to UG principles and parameters (Mitchell et al., 2019, pp. 102-104). With partial access to UG, some parts of UG are not available, and if a certain L2 parameter is not activated in the learner's L1, such as grammatical gender or article, it can be difficult to adjust parameters, which leads to the learner continuing to apply L1 settings (Mitchell et al., 2019, p. 103). UG claims for possibilities of full L1 transfer, with the L1 parameters transferred to construct the initial L2 hypothesis, or no transfer, where the L1 parameters do not contribute to the learner's hypothesis about the L2 parameters (Mitchell et al., 2019, pp. 113-114). Studies show that L2 learners construct grammar and activate parameters that are not available in their L1 (Mitchell et al., 2019, p. 113).

Ringbom (2007, pp. 51-52) claims that the positive transfer effect of having an L1 related to the TL is more significant in practice than the length of the learning period, although the difference between the languages becomes less significant as the proficiency

level gets higher, e.g., at university level. Furthermore, Ringbom (2007, p. 78) points out that transfer does not occur only from L1, but also from other languages known to the learner if their L2 and L3 seem more related than L1 and L3. Even unrelated non-native languages may provide support in the form of positive transfer when they share some linguistic characteristics (Ringbom, 2007, p. 79).

2.3 Overgeneralization

Overgeneralization is a phenomenon in which language learners apply a rule or pattern in a situation where it does not apply in the target language, resulting in an overgeneralization error (McKercher, 2018, p. 1). It can be observed in both L1, L2, or additional language acquisition but should be distinguished from language transfer, where properties of one's L1 give rise to errors in an L2 or more languages. Overgeneralization is most obvious in morphology but can also be found in the domains of argument structure, syntax, and phonology. A typical example of overgeneralization is the use of regular inflections in situations where exceptional, suppletive forms are the conventional words, such as *runned* instead of *ran* and *foots* instead of *feet* (McKercher, 2018, p. 1).

As explained above, overgeneralization is a type of intralingual error where the main cause is not related to learners' L1. In Swedish contexts, learners of English with L1 Swedish are likely to overuse certain grammatical features of English, such as the third-person singular *-s* (e.g., *My brothers and my parents *thinks*) and the progressive form *-ing* that are not available in their L1 Swedish structure (Köhlmyr, 2003, p. 233). Another possible factor of overgeneralization can be 'induced errors', as learners apply a specific rule they have recently learned even in places where it is incorrect, e.g., the third-person singular *-s* (Köhlmyr, 2003, p. 234). Köhlmyr argues further that this is a basic grammatical feature that is frequently practiced and emphasized in school grammars, and thus some errors may be attributed to the teaching and teaching materials.

2.4 Subject-Verb Agreement

Subject-verb agreement mistakes¹ are one of the most typical grammatical mistakes made by learners of English (Estling Vannestål, 2015, p. 88). When the subject is in the third-person singular, one of the following actions must be taken:

- a) Add an *-s* to a verb in the present tense, e.g., *walks*.

¹ When a learner does not make the same type of error consistently, it can be defined as a mistake (Gass et al., 2020, p. 86). In this study, the word "error" will be used when referring to subject-verb agreement errors.

- b) Add *-es* to a verb ending in an /s/-, /z/-, /f/-, /tʃ/- or /dʒ/-sound, e.g., *misses*.
- c) Add *-es* to *do* and *go*, e.g., *does* and *goes*.
- d) Use *is* in the present tense of *be*.
- e) Use *has* in the present tense of *have*.
- f) Use *was* in the past tense of *be*.

(Estling Vannestål, 2015, p. 89)

Köhlmyr (2001, p. 173) explains further the principle of this agreement, where the verb agrees with the subject in number, whether:

- a) the subject or the head of the subject is a noun or a noun phrase (NP), e.g., *The flower grows/The flowers grow*, or
- b) a pronoun, a finite (e.g., *He likes/I like*) or a nonfinite clause (e.g., *How you go there doesn't matter/To go there is dangerous*).

(Köhlmyr, 2003, p. 173)

The subject-verb agreement explained above is the most important type of agreement in the English language (Köhlmyr, 2003, p. 173), and the principle should be easy to follow when there is no postmodifier in the preceding noun phrase (Estling Vannestål, 2015, p. 89). Nevertheless, many learners make errors even in simple cases, and one possible reason presented by Estling Vannestål (2015, p. 89) is that learners might have misunderstood the principle and think that an *-s* should be added to the verb when the noun is in the plural, e.g., *the boys runs. Köhlmyr's study shows that most errors occur with the primary verbs *be*, *have*, and *do* (2003, p. 178), and furthermore, that the error rate increases with coordinated noun phrases, especially when two nouns (e.g., *Football and tennis is my favorites), and a pronoun and a noun (e.g., *N and I spends time together) are coordinated (Köhlmyr, 2003, p. 179).

3. Method

The central research questions of this essay are to study which types of subject-verb agreement errors are most common among the students in their writing production and to what extent these errors are related to the types of subjects, verbs, or the distance between the subject and the verb. In this section, the material for the study will be presented, followed by descriptions of the applied error categorization and the method for data analysis. Finally, questions regarding the validity and reliability of the study will be discussed.

3.1 Material

For this essay, a corpus study using the Swedish Learner English Corpus (SLEC; first release) is conducted. SLEC comprises argumentative texts written by young English learners who attend Swedish junior and senior high schools (Kaatari et al., 2024, p. 1). To ensure inter-text comparability, SLEC currently comprises argumentative texts written on the same topic, and in the first version, the corpus contains 1,098 texts with a total of 481,155 words (Kaatari et al., 2024, p. 2). In the first release, three different versions of SLEC are available: version 1A, which includes all original files and the metadata; version 1B, which contains 300 manually cleaned subsample texts; and version 1C, which comprises 120 texts both assessed and cleaned (Kaatari et al., 2024, p. 4). The majority of the assessed texts in the corpus are graded as B1 or B2 according to the Common European Framework of Reference for Languages (CEFR) scale of grading (Kaatari et al., 2024, p. 8).

This study focuses on the written production of junior and senior high school students in the 9th and 11th grades, the latter attending their second year of high school in both academic and vocational programs. The sample data for the study consists of a total of 24 texts, which are randomly chosen but evenly distributed in terms of binary gender and school year. The representation of 11th grade students in academic and vocational programs is also balanced, as shown in Table 1.

Table 1. Student variables in relation to collected data.

	9 th grade male	9 th grade female	11 th grade academic male	11 th grade academic female	11 th grade vocational male	11 th grade vocational female
No. of students	6	6	3	3	3	3
Tot. no. of words (incl. the headings)	1,933	2,263	1,494	1,418	878	1,289

SLEC contains, as mentioned, 1,098 texts with a total of 481,155 words (Kaatari et al., 2024, p. 2). As the individual normalization of the total numbers of words for each student group would have involved a full manual operation, this has not been done due to a limited timeframe for the study. All the analyzed texts are written by students who attend Swedish-speaking schools and programs with Swedish as their L1, and the texts are produced in a classroom setting in collaboration with their teachers. As the students were allowed to use a

spell checker in their text production (Kaatari et al., 2024, p. 2), ambiguous spelling errors where it is not certain whether they can be more than a simple spelling mistake, e.g., *it's* spelled as *its* without an apostrophe, are excluded from the error analysis in this study. In SLEC, various detailed metadata on the students' background is provided, for example, educational information, language background, and extramural English activities that enable the study of relations between the linguistic properties of student texts and various extralinguistic and learner variables. The corpus also includes further information on the circumstances of the text production in the classrooms, such as the time allowed, whether it was a graded assignment, or whether there were preparatory lessons prior to the production (Kaatari et al., 2024, p. 3). In the process of data selection for this study, the students were categorized into groups according to binary gender, school year, and type of program for high school students. Other variables taken into consideration were the students' and their parents' L1 and the students' experience of attending an English-speaking school. In this study, only students with Swedish as their L1 and Swedish-speaking parents are included, and they all have no experience attending an English-speaking school. This is in an attempt to eliminate transfer factors from other languages than Swedish and thus ensure the validity and reliability of the study, which will be discussed in Section 3.4. Detailed information about the variable settings is available in Appendix I, and information on the chosen texts is in Appendix II, as well as in the section on primary sources.

3.2 Error Classification

As discussed in Sections 2.2 and 2.3, there are several ways to analyze and categorize learner errors, which should occur repeatedly in contrast to mistakes. First, errors need to be categorized as intralingual or interlingual; the former is overgeneralization not related to L1, which involves incorrect use of third-person singular forms, while the latter is L1-related transfer (Gass et al., 2020, p. 87; Köhlmyr, 2003, p. 233). Table 2 below shows some examples of both types of errors.

Table 2. Examples of intralingual and interlingual subject-verb agreement errors.

	Incorrect form	Correct form
Intralingual (overgeneralization)	*I does, *I has, *I is (primary verbs) *We goes (full verb)	I do, I have, I am We go
Interlingual (transfer)	*My father are, *I were (primary verbs) *My mother come (full verb)	My father is, I was My mother comes

Thereafter, both types of errors are to be further classified and analyzed. In this study, the following variables are used:

- a) grade: 9th or 11th; within the latter group, also according to the program attended, either vocational or academic,
- b) gender: binary male or female,
- c) types of the subjects preceding the verbs (classified in the categories below),
- d) types of the verbs: primary verbs (*be, have, do*) or other, and
- e) the distance between the subject and the verb, in terms of number of words.

Further classification of the subject types is made as follows (Köhlmyr, 2003, pp. 173-174, 177, 179-180):

- i) a pronoun: a personal, a relative, an indefinite, or a demonstrative pronoun,
- ii) a noun/NP²: the nouns or the heads in the subject NPs can be either count nouns, noncount nouns, or proper names,
- iii) coordinated NPs: two nouns coordinated with *and*, or a pronoun and a noun combined with *and*,
- iv) existential *there*, or,
- v) a small, mixed group of other types of subjects.

After the primary error analysis, the same texts were investigated again to sample the correct examples of subject-verb agreement in order to study where it occurred most correctly in different groups. As mentioned in Section 3.1, ambiguous spelling and other orthographic errors are excluded from this study, as the students were allowed to use a spell checker in their writing activities (Kaatari et al., 2024, p. 2), and it can affect the validity of the study. On the other hand, among the correct samples of subject-verb agreement, simple errors where it is obvious what the writers intend to mean, e.g., I wass [was] with two s in the end, are included since the intended word is obvious and the agreement is correctly working.

3.3 Method for Data Analysis

For a corpus project, there are two major methodological approaches available: corpus-based or corpus-driven, depending on the research question (Granger, 2012, p. 13). The corpus-based approach is essentially deductive and aims to test a hypothesis or rule against corpus data, where the corpus does not act as the supporting medium to confirm or refute a pre-

² In this study, collective nouns are excluded from the analysis, as they can be handled as both singular and plural in different variants of English. In American English, a singular verb or pronoun form is generally preferred, while a plural verb form is often used in British English (Estling Vannestål, 2015, p. 126).

existing theoretical construct. The corpus-driven approach, on the other hand, is inductive and tries to exploit the full force of the corpus in order to generalize from the observation of data to build up the theory or rule (Granger, 2012, p. 13). The corpus-based approach describes and explains linguistic patterns of variation and use, and its goal is to discover the systematic patterns of use that govern the linguistic features recognized by standard linguistic theory, not to discover new linguistic features (Heine & Narrog, 2015, p. 196). This study thus applies the corpus-based approach since it aims to investigate the variation and use of the recognized grammatical principle of subject-verb agreement.

In this study, the steps of error analysis presented in Gass et al. (2020, p. 87), except the last step of pedagogical intervention, are applied: (1) collect data, (2) identify errors, (3) classify errors, (4) quantify errors, and (5) analyze the source. As mentioned in Section 3.1, the data for this study is sampled from SLEC. The data used in this study consists of full texts from version 1A, which were manually close-read and analyzed for data collection. The AutoCorrect function in Word is partially used to do the first check, but all the texts are manually investigated to detect errors. Microsoft does not provide information on the accuracy of AutoCorrect for Word. Nevertheless, there are accuracy tests on other autocorrection functions, such as Google, performed by private individuals. One such test (Bawaskar, 2020) achieved 87.1% accuracy with a dataset consisting of words that do not repeat themselves, which increases the complexity of the problem and thus reduces accuracy. Bawaskar (2020) claims that state-of-the-art results should lie at 89-90% accuracy, which seems reasonable compared to the error detection results by AutoCorrect in this study.

All the errors found are then listed and sorted according to the classification presented in Section 3.2. After this, the classification of all correct instances of subject-verb agreement is also made in order to examine possible factors and explanations of the errors. The error rate is counted per 100 words, i.e., the raw error frequency count is divided by the number of words in the texts and multiplied by 100, in order to make the results more visible. This calculation formula for the normalized frequency is also applied to the samples of correct subject-verb agreement. Even the relative frequency is calculated by dividing the number of subject-verb agreement errors per student group by the total error instances. This is multiplied by 100 in order to show the error scores in percentage.

3.4 Validity and Reliability

To measure hypothetical constructs, which SLA research often employs, the constructs need to be processed from theory to more concrete empiricism (Révész, 2012, p. 203). Validity

refers to how appropriately and precisely this operationalization matches the theoretical definition of the construct (Révész, 2012, pp. 203-204). Validity is even concerned when coding data, regarding how accurate and meaningful interpretations can be made, although there are both systematic and random sources of error (Révész, 2012, p. 204). Internal validity refers to the accuracy of the results, which must presume that the differences that occurred in the dependent variable were produced by the independent variable rather than by some other factor (Roever, 1977, p. 341). External validity, on the other hand, asks the question of generalizability, e.g., if the effect can be generalized and to what populations, settings, treatment variables, and measurement variables (Roever, 1977, pp. 341-342).

This study employs an analysis of subject-verb agreement errors produced by Swedish junior and senior high school students in written English. In terms of internal validity, it is required to mention that certain error types, such as misspellings, have been excluded from the analysis, as mentioned in Section 3.1. Furthermore, the analyzed texts do not represent all Swedish junior and senior high school students who study English; this study includes a total of 24 texts, which is a small number in terms of validity. However, due to the limited time frame, it is necessary to limit the number of texts to be examined, and the aim of this study is to investigate common types of subject-verb argument errors within the target group. In an attempt to increase validity, the students' texts are sampled randomly with regard to a balanced representation of gender, grade, and other variables, as stated in Section 3.1, and detailed information on the applied variables is attached in Appendix I. Despite these measures, it is necessary to mention potential problems regarding validity and reliability. First, the time allowed for the writing activities differs for the different grades: 55 minutes for 9th grade students, 50 minutes for vocational program students, and 80 minutes for academic program students. The main reason for these options was that this setting generated most student texts, which made random selection of the texts possible. Another variable that might play a role in terms of validity and reliability is the grading of the texts by the teacher. In this study, no grading was chosen for the same reason as the writing time, but it should be mentioned that the grading might affect the students' motivation. The selected data is approached with the utmost care, as is the chosen methodology, to achieve as much validity as possible. However, as mentioned in Section 2.4, error analysis has some disadvantages since it can be difficult to make a clear distinction between an error and a mistake, especially when it is not possible to contact the students who produced the texts. Furthermore, the analysis can also be biased by the researcher's knowledge of the English language, which may affect the validity of the results.

Reliability refers to the consistency of the measure, which means that the researcher needs to ensure that anyone coding the data will get similar results each time (Polia, 2012, p. 147). There is a risk of failing to find significant differences when the measure used has low reliability (Polia, 2012, p. 147), which one has to be aware of. To ensure as much reliability as possible, the data material is investigated and classified according to the objective criteria presented in Section 3.2. As pointed out earlier, the study involves manual control of the chosen texts, and this can affect the reliability of the analysis.

Finally, Granger (2012, p. 9) claims that learner corpora provide researchers with a wealth of occurrences to study linguistic phenomena, although it is often necessary to select a representative sample when the whole data cannot be studied. Granger also points out that the optimal size of a learner corpus is dependent on what type of linguistic phenomenon is targeted; highly frequent components such as articles can be studied well in a small corpus, while lexical words often require a larger corpus. One should also not focus only on the total number of words in a corpus but also on the number of learners that produced data, since representativeness measured in terms of the number of learners has no apparent relation to the corpus size measured in terms of the number of words (Granger, 2012, p. 9). As mentioned in Section 3.1, this study uses SLEC, which is a relatively small corpus. However, since the study focuses on subject-verb agreement which occurs with a high frequency in English sentences, the corpus size should be sufficient in this case. The corpus also comprises only argumentative texts on the same topic written by young learners in a school setting, which ensures high intertextuality and facilitates comparisons between different student groups.

4. Results and Discussion

In this section, the results of the corpus study will be presented. The research questions of the study are to investigate which types of subject-verb agreement errors are most common among the target student groups in their writing production and to what extent the errors are related to the types of subjects, verbs, or the distance between the subject and the verb. First, the general results of the study and the results of each separate group in relation to grade and binary gender will be presented in Section 4.1. The overall error rate in this case pertains to every unique error identified in each text. After this, the results classified either as overgeneralization errors or transfer errors will be presented in Section 4.2. Thereafter, the errors will be further analyzed in relation to the subject types (Section 4.3), the verb types (Section 4.4), and finally, the distance between the subject and the verb (Section 4.5). Along

with the error examples, the correct instances will also be presented to further investigate possible background factors.

4.1 General Results

In this study, 41 errors were identified in 24 texts written by students in the 9th and 11th grades. The distribution between the grades and the binary gender, the number of errors per group, and the error rates are summarized in Table 3. In the table, the total number of correct examples of subject-verb agreement is also presented, along with the normalized frequency as mentioned in Section 3.3. The error rate in percentage is calculated from the total instances of subject-verb agreement for each student group.

Table 3. General results of each student group.

	9 th grade male	9 th grade female	11 th grade academic male	11 th grade academic female	11 th grade vocational male	11 th grade vocational female	Total
No. of students	6	6	3	3	3	3	24
Tot. no. of words (incl. the headings)	1,933	2,263	1,494	1,418	878	1,289	9,275
Average no. of words per text	322	377	498	473	293	430	386
No. of errors (relative frequency)	6 (14.6%)	5 (12.2%)	5 (12.2%)	12 (29.3%)	5 (12.2%)	8 (19.5%)	41 (100%)
Error rate (per 100 words)	0.310	0.221	0.335	0.846	0.569	0.621	0.442
Tot. no. of correct subject-verb agreements	231	290	156	153	104	154	1,088
Normalized frequency (per 100 words)	11.950	12.815	10.442	10.790	11.845	11.947	11.730
Errors per tot. instances of subject-verb agreement (%)	2.53	1.69	3.11	7.27	4.59	4.94	3.63

The error rate is generally higher among senior high school students in the 11th grade than in the 9th grade. Regarding gender differences, female students in the 11th grade show a higher error rate than male students, while it is the opposite in the 9th grade. The lowest error rate is achieved by female students in 9th grade, with 0.221. At the same time, they have the least number of words per sentence on average among all groups, with 17 words. The error rate in Table 3 takes word count into account and is not intrinsically related to text length; nevertheless, they might still be related. At first glance, it seems that there is a correlation between the average number of words per sentence and error rates, as 9th grade female students have the lowest error rate (0.221) and the fewest words (average 17 words) per sentence, while 11th grade academic female students have the highest error rate (0.846) and average 21 words per sentence. However, this is not consistent since vocational male students have a lower error rate of 0.569 with 24 words per sentence on average, and 11th grade academic male students have the second lowest error rate of 0.335, although they produced the most words per sentence on average with 25 words.

Female students have generally been expected to achieve higher results than male students in the Swedish school system (Sveriges Kommuner och Landsting, SKL, 2018, p. 4), and at the junior high school level, male 9th grade students achieved better results than female students in only two subjects at the end of the spring term 2018: PE and sign language, which is not taught at all schools (SKL, 2018, p. 8). The Swedish National Agency for Education, Skolverket, also concludes (Skolverket, 2022b, p. 4) that the proportion of female 9th grade students eligible for applying to a national high school program has been higher than male students since the Agency started the observation in spring 1998. However, this is not mirrored in the error rates above, where female students have higher achievement than male students only in the 9th grade. Even at the high school level, the National Agency for Education states that female students achieve higher study results than male students attending the national programs (Skolverket, 2022a, p. 30). Study results cannot be related only to writing ability and not least to a certain grammatical rule, but in this case, it is interesting that both male groups in 11th grade showed a lower error rate than their co-female students. All students share the same variables regarding their L1, their parents' L1, the language they speak at home, and their lack of experience attending an English-speaking school. There may be some other variables or personal factors that affected the outcome in that case, e.g., extramural English activities. The texts written by 11th grade students examined in this study were produced in English course 6, which is not mandatory for high school students in vocational programs. This means that these students made an active choice

to study the course, and thus motivation may also be one of the factors for the lower error rate these students show compared to female 11th grade students in academic programs where English 6 is mandatory.

At the same time, it should be mentioned that the students succeed with subject-verb agreement much more often than they make errors. As seen in Table 3, the normalized frequency per 100 words for correct subject-verb agreement in the analyzed texts is significantly higher than the error rate. When the correct agreement is most likely to occur, e.g., in relation to the subject type and the verb type, will be discussed later in this section.

4.2 Overgeneralization and Transfer Errors

As mentioned in Section 4.1, a total of 41 errors were identified in 24 texts written by students in the 9th and 11th grades. These errors were classified either as overgeneralization or transfer according to the criteria presented in Section 3.2, which can be seen in Figure 1.

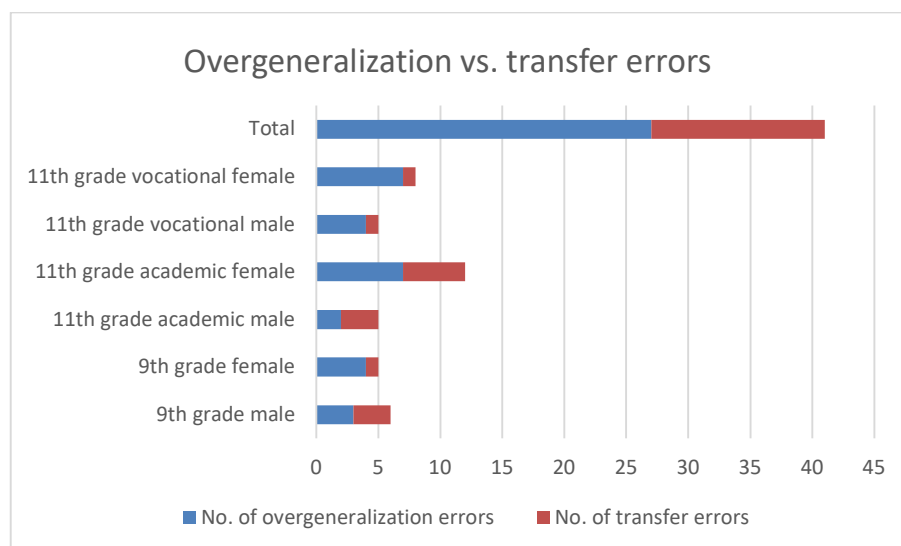


Figure 1. Error types classified as overgeneralization or transfer.

Among the total of 41 errors, 27 (66%) are identified as overgeneralization errors and 14 (34%) as transfer errors. In this study, subject-verb agreement errors of the type where the verb in unmarked form ($V\emptyset$) is incorrectly used with 3rd person subjects are regarded as transfer errors (Köhlmyr, 2003, p. 259). Köhlmyr claims that problems relating to this type of subject-verb agreement are very frequent, which can also be observed in the result of this study. One of the most likely reasons for this type of transfer error is, as Köhlmyr explains, the fact that Swedish verbs do not inflect for person or number, unlike English verbs. Agreement errors with the verb in marked singular form (V_s), as in *we sings, are here treated

as overgeneralization because there is no 3rd person verb inflection in Swedish (Köhlmyr, 2003, p. 259).

Below are some examples of overgeneralization errors:

- (1) ...you maybe want to buy a product that you does not even need. [do]
- (2) Social media also creates an addiction and a bad habit that is hard to break. [are]
- (3) Different kinds of relationships brings safety... [bring]

As can be seen in Figure 1, overgeneralization errors play a large part among the 9th grade female students and the 11th grade vocational students, both male and female. In total, 11 of the 27 overgeneralization errors are related to the primary verb *be*, where the marked 3rd singular form is incorrectly applied. Other frequent verbs related to overgeneralization are the lexical verb *make* (n = 3), and the primary verb *do* (n = 3). A visible tendency is that the proportion of overgeneralization errors is significantly or slightly higher than transfer errors among groups with a higher error rate: 11th grade female students in academic programs and 11th grade vocational students, with an exception for 9th grade female students, who have the lowest error rate but also more frequent overgeneralization errors. The result does not corroborate the general assumption of decreased transfer with increasing language proficiency (Köhlmyr, 2003, p. 320), but it indicates that the students who produce overgeneralization errors are aware of the inflection system of the English verbs but might not have fully acquired the agreement rules or might not manage to apply the rules when required.

Regarding the transfer, half (n = 7) of the identified 14 transfer errors are related to the primary verb *be*, where students used the unmarked form *are* incorrectly, as in (4) and (6):

- (4) The physical health are mostly affected by your diet... [is]
- (5) That is also a thing that help your life both physically and mentally. [helps]
- (6) Secondly eating healthy after individualized diet are also something... [is]

Köhlmyr (2003, p. 275) claims that transfer errors with the auxiliary *be* are due to phonological and orthographic similarity between Swedish *är* and English *are*, which results in grammatical transfer errors. This explanation is reasonable; however, this study also shows that in some cases, the distance between the subject and the verb seems to make it difficult for the learners to recognize if the subject is singular or plural and apply the correct verb form. This can also affect the auxiliary *be*, like it does with other verbs, which will be discussed in Section 4.5.

Third-person singular -s was one of the last morphemes in child L1 acquisition order according to Brown's morpheme studies (1973, as cited in Gass et al., 2020, p. 110), and this

also turned out to be applicable in an L2 context (Gass et al., 2020, pp. 111, 116-117; Mitchell et al., 2019, p. 49). This implies that subject-verb agreement overgeneralization errors are supposed to appear first when a learner is familiar with other basic grammatical principles, and moreover, the learner cannot produce this type of error without knowing the principle of third-person singular -s. It is also possible that it is related to U-shaped learning, as learners first copy and conform to TL norms, deviate from them after a while, and then go back to the correct form again (Gass et al., 2020, pp. 307-308). In the second stage of the U-shaped learning process, learners can overproduce marked forms in TL, which, in this case, results in the overuse of third singular -s where it is not necessary.

Regarding the transfer errors, it is not certain if such an error depends on a lack of awareness, attention to the form, or a simple production mistake. However, in both overgeneralization and transfer errors, there may be a lack of automaticity, which makes it possible for the learner to have control over linguistic knowledge and produce the form effortlessly (Gass et al., 2020, p. 304). If the rule is more automated, it should be possible to devote more attentional resources to other parts of language learning (Gass et al., 2020, p. 304), and in this case, explanations may also be found in students' choices of words and sentence constructions, which will be discussed in the coming sections.

4.3 Subject Types

The identified errors and correct instances were further classified in relation to the subject types, according to the description in Section 3.2. The result is shown in Table 4. For the subject category pronouns, the instances with a relative pronoun are presented in parenthesis.

Table 4. Incorrect and correct instances of subject-verb agreement in relation to subject types.

	Subject type	9 th grade male	9 th grade female	11 th grade academic male	11 th grade academic female	11 th grade vocational male	11 th grade vocational female	Total	%
Subject-verb agreement errors	A pronoun (whereof a relative pronoun)	2 (1)	2 (2)	1 (1)	6 (3)	2 (2)	2 (2)	15 (11)	36.6
	A noun/NP	3	1	0	4	2	5	15	36.6
	Coordinated nouns/NPs	0	0	1	1	0	1	3	7.3
	Existential <i>there</i>	0	0	0	0	0	0	0	0
	Other	1	2	3	1	1	0	8	19.5
	Total	6	3	4	6	3	6	41	100

Correct subject-verb agreement	A pronoun (whereof a relative pronoun)	157 (16)	207 (18)	94 (14)	96 (17)	71 (5)	126 (2)	751 (72)	69
	A noun/NP	58	55	45	48	22	24	252	23.2
	Coordinated nouns/NPs	4	4	2	6	3	4	23	2.1
	Existential <i>there</i>	6	7	3	2	6	0	24	2.2
	Other	6	17	12	1	2	0	38	3.5
	Total	231	290	156	153	104	154	1,088	100

The most common subject types related to the errors were the categories “a noun or a noun phrase (NP)” and “a pronoun,” including personal, relative, indefinite, and demonstrative pronouns, with 15 errors each. In third place comes the category “other subjects” (n = 8), followed by “coordinated NPs,” with 3 errors. No errors were found with the existential *there*.

Shown below are examples of errors with a pronoun as the subject:

- (7) No matter how cliché it might sound, or how many times you’ve heard it and still doesn’t believe it... [do] (overgeneralization)
- (8) everyone don’t mind others much and start focusing on themselves... [does] (transfer)
- (9) I love thinking about things that makes my life a good life, even though I know that it can’t always be like it is right now. [make] (overgeneralization)
- (10) when it is we who is in the stable every days on the weekend. [are] (overgeneralization)

There were only a small number of errors with personal, indefinite, and demonstrative pronouns in this category. Most of the errors were related to relative pronouns, such as *that*, *which*, or *who*, as the subject in subordinate clauses. When using such relative pronouns, it is the noun these relative pronouns are referring to that determines if the subject is singular or plural. These errors in the study suggest that many students fail to pay attention to this principle or have not acquired it fully enough to apply it in their writing production.

Errors with a noun/NP as the subject were also visible:

- (11) The mental health are mostly affected by social life and traumatic experiences (PTSD). [is] (transfer)
- (12) My friends means a lot for me... [mean] (overgeneralization)

In cases with a plural noun/NP subject, the errors could be attributable to “perseveration” of the -s (Thagg Fisher, 1985, p. 45), i.e., both the third-person singular ending -s and the plural noun ending -s are marked, as in (12): *My friends means a lot for me... However, Köhlmyr (2003, p. 276) argues that errors with plural pronouns as subjects, where there is no -s to influence the verb, cannot be explained this way. In such cases, Köhlmyr claims that it is more likely that the marked verb form is triggered by the “one -s principle” (Thagg Fisher, 1985, p. 45), which is the learners' interlanguage hypothesis that at least one, but not more than one, of the agreeing elements should be marked with the -s morpheme. Another difficulty is the use of nouns that are considered single and uncountable, such as *news* in the following example:

- (13) ...the amount of depressing news that are a constant presence in most people's daily life. [is] (transfer)

In this case, the subject is a relative pronoun, but whether the verb should be singular or plural is determined by the word the pronoun is referring to. Problems with non-countable nouns occur often, as in (13) since these types of nouns can be classified differently in English and Swedish (Thagg Fisher, 1985, p. 74). In the above example, *news* is uncountable and treated as singular in English, and this error most likely comes from the Swedish word “nyheter,” which is plural and also treated as plural. Thagg Fisher (1985, p. 74) also claims that Swedish-speaking English learners can make errors due to “the one -s principle,” when the final -s of the singular noun becomes misinterpreted as the plural morpheme. Swedish learners of English often have problems with irregular and unmarked plural nouns, and especially when plural nouns lack the regular -s morpheme, learners with L1 Swedish tend to produce agreement errors. In the error example below (14), the writer might have focused on the word *happiness* instead of *pursuit*, which should be the head word, and misinterpreted the -s ending in *happiness* as plural:

- (14) ...our pursuit of happiness are mainly why we keep digging for something more... [is] (transfer)

In the category “other types of subjects,” there were 8 errors, mostly with a gerund phrase as the subject, as in (15), while there were only three errors with coordinated nouns/NPs as the subject, and all of them appeared to be overgeneralization errors, as in (16):

- (15) ...like how training and eating good, makes you feel good. [make]
(overgeneralization)

- (16) I would argue that my social life and health is the most important thing... [are]
(overgeneralization)

In the above cases, contiguity, i.e., whether the head of the subject phrase and the verb are in immediate contact or not (Thagg Fisher, 1985, p. 11), may have played a role. Thagg Fisher (1985, p. 11) claims that both in contiguous and non-contiguous constructions, the subject must be stored in short-term memory, and when the verb comes in immediate connection to the subject, it facilitates the retrieval. This should suggest that the error rate increases in relation to the distance between the subject and the verb, which will be discussed in Section 4.5.

As stated at the beginning of this section, it also needs to be mentioned that most students succeed with subject-verb agreements in the majority of the sentences they produce. Even most of the correct examples are related to the subject types “a pronoun” and “a noun/NP,” as can be seen in Table 4. Most of the pronouns used as the subject in correct instances are personal pronouns, such as *you*, *we*, *they*, or *I*. The result also shows that female students tend to use a pronoun as the subject more often than their male co-students. This can be due to the topic of their writing activity, which is to discuss a good life. Especially, 11th grade female students in vocational programs refer to personal details in their texts, which has resulted in a frequent use of personal pronouns. There were no errors with existential there, as presented earlier, while the majority of groups used the construction *there is/there are* correctly (Table 4). This can suggest that it is automated (Gass et al., 2020, p. 304) for many of them, but at the same time, it can also be related to contiguity, as the verb comes directly after the subject *there*, which probably facilitates the correct inflection of the verb (Thagg Fisher, 1985, p. 11). Another remarkable tendency is that the two groups with the lowest error rates, 9th grade female and 11th grade academic male students, used other types of subjects correctly, mostly gerund phrases, more often than the other groups.

4.4 Types of Verbs

Regarding the types of verbs, errors in combination with primary verbs (*be*, *have*, *do*) are more frequent (56.1%, n = 23) than with other verbs (43.9%, n = 18), as expected due to the more frequent use of primary verbs in general. Table 5 shows the results of incorrect and correct instances of subject-verb agreement classified in relation to the verb types.

Table 5. Incorrect and correct instances of subject-verb agreement in relation to verb types.

	Verb type	9 th grade male	9 th grade female	11 th grade academic male	11 th grade academic female	11 th grade vocational male	11 th grade vocational female	Total	%
Subject-verb agreement errors	Primary verbs	5	3	3	6	2	4	23	56.1
	Other verbs	1	2	2	6	3	4	18	43.9
	Total	6	5	5	12	5	8	41	100
Correct subject-verb agreement	Primary verbs	90	114	52	67	47	71	441	40.5
	Other verbs	141	176	104	86	57	83	647	59.5
	Total	231	290	156	153	104	154	1,088	100

Among the 23 errors with the three primary verbs *be*, *have*, and *do*, as many as 18 errors (78%) are related to the verb *be*, while there are only four errors with *have* (n = 1) and *do* (n = 4). This means that nearly half of all errors in the study are related to the verb *be* (43.9%, n = 18). The most common lexical verb used in relation to the errors was *make*, which appeared four times in total.

When looking at the correct examples in relation to the verb type in Table 5, primary verbs are used in 40.5% (n = 441), while they are more frequent in error examples (56.1%, n = 23). In the course of analysis, it has turned out that 9th grade female students, who have achieved the lowest error rate in this study, show a large number of correct agreement examples with modal verbs (n = 78) that do not require inflection to match the third-person singular. Nearly one-third of all modal verbs in correct examples (n = 249) belong to this group, and this can partly explain their low error rate. Their texts often contain expressions with different degrees of possibility by using modal verbs such as *can*, *may*, and *might*, although it is not clear what background factor can be related to this phenomenon. At the same time, this can also be an example of overproduction due to avoidance (Odlin, 1987, pp. 36-37), as mentioned in Section 2.2.

As discussed in Section 4.1, overgeneralization errors are more frequent than transfer errors. When this classification is applied to the verb types (Figure 2), the result shows that overgeneralization errors occur with primary and other verbs at the same frequency, while transfer is more frequent with primary verbs. Normalized by the total number of primary verbs or other verbs, the error frequency is 4.96% for primary verbs and 2.71% for other verbs.

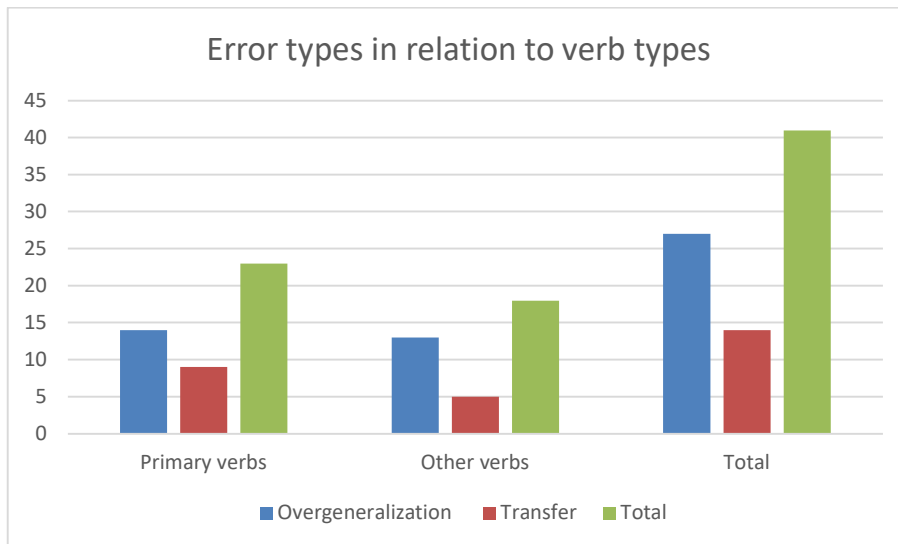


Figure 2. Subject-verb agreement error types in relation to verb types.

Köhlmyr (2003, p. 276) argues that a possible reason for overgeneralization errors with the third-person singular -s is that the learner knows the -s form and sees it as "typically English" and thus uses it too generously, as a maximization of special L2 forms and structures. It is a reasonable explanation, especially for verbs that are regular. According to Köhlmyr, this is the most frequent type of subject-verb agreement error, both with pronominal and noun/NP subjects, which was also discussed in Section 4.3. On the other hand, transfer errors are more frequent with primary verbs ($n = 9$) compared to transfer errors with other verbs ($n = 5$). Primary verbs *be*, *have*, and *do* have more complex inflection patterns than other verbs, where a simple -s ending is required for third-person singular, and this may result in more frequent use of the unmarked form ($V\emptyset$) and its replacement by the same form, *are*.

4.5 The Distance Between the Subject and the Verb

Another important factor related to subject-verb agreement errors is subject-verb contiguity, i.e., the distance between the subject and the verb. The results classified by gender, grade, and program are summarized in Table 6. The distance is measured by the number of words between the subject and the verb. In the case of an NP and a coordinated NP, the distance is counted from the head word of the NP, and in a gerund phrase and an infinitive phrase, the first word is counted as 0. Thus, the zero distance in Table 6 means that the subject and the verb are placed next to each other in the sentence.

Table 6. Results for each student group in relation to distance between the subject and the verb.

	Distance between subject and verb	9 th grade male	9 th grade female	11 th grade academic male	11 th grade academic female	11 th grade vocational male	11 th grade vocational female	Total	%
Subject-verb agreement errors	0	5	2	1	8	3	5	24	58.54
	1-3	0	1	1	2	1	3	8	19.51
	4-6	1	2	1	2	0	0	6	14.63
	7+	0	0	2	0	1	0	3	7.32
	Total	6	5	5	12	5	8	41	100
Correct subject-verb agreement	0	187	247	111	125	84	129	883	81.2
	1-3	27	25	29	17	15	19	132	12.1
	4-6	13	13	9	9	2	5	51	4.7
	7+	4	5	7	2	3	1	22	2.0
	Total	231	290	156	153	104	154	1,088	100

Generally, non-contiguity is regarded as one of the factors that results in subject-verb agreement errors. Looking at studies conducted in the Swedish context, Thagg Fisher (1985, p. 93) found that 45% of the subject-verb agreement errors occurred in non-contiguous cases, where the subject and the verb were not placed next to each other, and 55% in contiguous cases. Köhlmyr's (2003, p. 181) study, on the other hand, showed that 86% of the error cases were contiguous, i.e., the subject and the verb were placed next to each other. As seen in Table 6, this study shows that more than half of the errors (59%, n = 24) occur when the subject and the verb are in immediate contact, contrary to general understanding.

Errors where the subject and the verb are placed next to each other are as follows:

- (17) What make a good life? [makes] (transfer)
- (18) ...I feel how my body feel better and then my mental health is better. [feels] (transfer)
- (19) Your relationships matters more than you sometimes want to admit... [matter] (overgeneralization)
- (20) There are also multiple psychological studies which shows that it is easier to deal with sadness by talking to trusted people about it. [show] (overgeneralization)

As discussed in previous sections and as mentioned at the beginning of this section, errors that involve a relative pronoun such as (20) are evident in this category; 11 of the 24 errors (46%) with zero distance are related to relative pronouns such as *that*, *which*, and *who*. There can be several explanations for this type of error; one of them can be a misinterpretation of the grammatical rule, where relative pronouns are associated with the

third-person singular for some reason or the plural ending -s on the preceding head triggers the "persistence of -s" (Thagg Fisher, 1985, p. 45), which leads to a repetition of the third-person singular ending -s when there is the plural noun ending -s. The other possibility is that the learners' focus is on the relative pronoun and the structure of the relative clauses they create, resulting in less attention to the preceding head word, which determines the correct verb form.

A closer look at the result shows that the groups with the lowest error rates, 9th grade female students and 11th grade academic male students, made few mistakes with a single noun/NP as subject: 4 of 5 errors made by 9th grade female students are related to a relative pronoun or a gerund as subject, such as (21), while 3 of 6 errors made by 9th grade male students are connected to a single noun or a personal pronoun (e.g., you does, health are).

(21) We've also talked about how being introverted and being all alone isn't nearly the same thing. [are] (overgeneralization)

Among 11th grade male academic students, all errors were either related to coordinated nouns/NPs as in (22), relative pronouns, or a gerund phrase as in (21) above:

(22) Where training, food, and health goes hand in hand for a better life. [go] (overgeneralization)

The above errors have a longer distance, three words or more, between the subject/head word and the verb, which may have led to the wrong verb form. More complex sentence structures can thus be connected to the errors, but as the study includes a small sample of texts, it is difficult to show a clearer relationship. It can also be possible that the students in the study did not produce many examples with a longer distance between the subject and the verb, which also appears to be the case when examining the correct instances in relation to the distance between the subject and the verb (Table 6).

Of all the correct examples (n = 1,088), the majority of them occur in contiguous structures (81.2%, n = 883), followed by a distance with 1-3 words (12.1%, n = 132). Only 22 (2%) of the correct examples had more than seven words between the subject and the verb, which confirms the foregoing hypothesis. Thagg Fisher (1985, p. 11) claims that there seems to be a human desire among English learners to place the subject and the verb near each other, which can be an explanation for the result in Table 6. Even though this study shows that errors occur mainly in contiguous structures, as Thagg Fisher (1985, p. 260) also concludes, the massive dominance of the contiguous structures among the correct examples cannot be ignored. Nevertheless, it does not mean that the distance does not affect the error occurrence,

and further studies are required to examine what impact the distance between the subject and the verb has.

5. Conclusion

At the beginning of this essay, two central research questions were presented. The first question is which type of subject-verb agreement errors are most common among the students in their writing production when classified into intralingual (overgeneralization) or interlingual (transfer) errors, and the second question is to what extent these errors are related to the types of subjects, verbs, or the distance between the subject and the verb.

In this study, a total of 41 agreement errors were found in 24 texts written by both male and female students in 9th grade and 11th grade, the latter from both vocational and academic high school programs. Regarding the first research question, the study result shows that overgeneralization (66%, n = 27) is more frequent than transfer errors (34%, n = 14) in all student groups, except for 11th grade male students in academic programs, who also achieved the second lowest error rate in all groups. 11 of 27 overgeneralization errors and half of 14 transfer errors are related to the primary verb *be*, which is also the most frequent verb (Köhlmyr, 2003, p. 260). Overgeneralization suggests that the students are aware of the third-singular -s but do not always apply it correctly, while transfer errors show a potential lack of awareness or attention to the form. In both cases, the errors indicate that these students have not automated the principle yet or are not capable of applying it without effort. At the same time, the third-singular -s is proven to be one of the last morphemes to acquire in SLA (Gass et al., 2020, pp. 116-117), and since the informants of this study are junior and senior high school students learning English as a foreign language, the result is not surprising.

Regarding the second research question, the result of this study shows that most errors (73.2%, n = 30) are related to subject types “a pronoun” or “a noun/NP.” In particular, relative pronouns such as *that* are involved in the agreement errors. Also, as mentioned above, the primary verbs are more related to errors than the other verbs. The verb *be* is involved in 18 of 41 errors (43.9%), which was expected as it is the most frequently used verb (Köhlmyr, 2003, p. 260). A closer look at the distance between the subject and the verb in relation to the errors indicates that there is a possible connection between the distance and the error rate in some of the student groups, but most of the agreement errors in the study occurred when the subject and the verb were in immediate contact (58.54%, n = 24). However, 11 of these 24 errors (46%) involve a relative pronoun as subject, which indicates that the learners have misinterpreted the grammatical principle or have not fully acquired it to be able to process it

in writing activities. Overuse of the third-person singular -s can also be an effect of teaching and explicit learning, which makes learners apply the form whenever it seems possible and relevant.

This study employed error analysis with a small number of written texts, with a focus on subject-verb agreement errors. There were a variety of other error types found during the analysis process, which also showed the complexity of error analysis and the not always clear boundary between intralingual and interlingual errors. In the context of teaching, most language teachers probably know that grammar rules that are easy to explain are not always easy to learn. The aim of the study was to investigate subject-verb agreement errors and background factors, and the results showed some possible answers to the research questions. At the same time, the answers also indicate that this is an area of research that needs to be explored further.

References

Primary Sources

Student essays collected from SLEC:

H_9_M_22_199.txt

H_9_M_22_203.txt

H_9_M_22_204.txt

H_9_M_22_206.txt

H_9_M_22_211.txt

H_9_M_22_212.txt

H_9_F_22_198.txt

H_9_F_22_205.txt

H_9_F_22_207.txt

H_9_F_22_208.txt

H_9_F_22_210.txt

H_9_F_22_215.txt

G_2_S_M_22_95.txt

G_2_S_M_22_104.txt

G_2_S_M_22_109.txt

G_2_S_F_22_114.txt

G_2_S_F_22_120.txt

G_2_S_F_22_134.txt

G_2_Y_M_21_33.txt
G_2_Y_M_21_29.txt
G_2_Y_M_21_32.txt
G_2_Y_F_21_31.txt
G_2_Y_F_21_28.txt
G_2_Y_F_21_34.txt

Secondary Sources

- Bawaskar, T. (2020, April 27). *Deep Learning: Autocorrect, Spell check for short words*. Towards Data Science. <https://towardsdatascience.com/deep-learning-autocorrect-product-and-technical-overview-1c219cee0698>
- Breene, K. (2019, November 15). *Which countries are best at English as a second language?* World Economic Forum. <https://www.weforum.org/agenda/2019/11/countries-that-speak-english-as-a-second-language/>
- Estling Vannestål, M. (2015). *A University Grammar of English with a Swedish Perspective*. Lund: Studentlitteratur.
- Heine, B., & Narrog, H. (2015). *The Oxford Handbook of Linguistic Analysis* (2nd ed., Oxford Handbooks in Linguistics). Oxford: Oxford University Press.
- Gass, S. M., Behney, J., & Plonsky, L. (2020). *Second Language Acquisition: An Introduction Course*. Fifth Edition. New York: Routledge.
- Granger, S. (2012). How to Use Foreign and Second Language Learner Corpora. In Mackey, A. & Gass, S. (Eds.) *Research methods in second language acquisition: A practical guide* (pp. 7-29, 1st ed.). Chichester, West Sussex; Malden, Mass.: Wiley-Blackwell.
- Kaatari, H., Wang, Y., & Larsson, T. (2024). Introducing the Swedish Learner English Corpus: A corpus that enables investigations of the impact of extramural activities on L2 writing. *Corpora*, 19(1).
- Kroeger, P. R. (2005). *Analyzing Grammar: An Introduction*. Cambridge University Press.
- Köhlmyr, P. (2003). *"To err is human-" an investigation of grammatical errors in Swedish 16-year-old learners' written production in English* (New ed., Gothenburg studies in English, 83) [Doctoral dissertation, University of Gothenburg].
- McKercher, D. (2018). Overgeneralization. *The TESOL Encyclopedia of English Language Teaching*, 1-6.
- Mitchell, R., Myles, F., & Marsden, E. (2019). *Second Language Learning Theories*. Fourth Edition. Routledge. Web.
- Modiano, M. (2009). *Language learning in the multicultural classroom: English in a European and global perspective* (1st ed.). Lund: Studentlitteratur.

- Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge University Press.
- Polio, C. (2012). How to Research Second Language Writing. In Mackey, A. & Gass, S. (Eds.) *Research methods in second language acquisition: A practical guide* (pp. 139-157, 1st ed.). Chichester, West Sussex; Malden, Mass.: Wiley-Blackwell.
- Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning* (Vol. 21, Second language acquisition). Clevedon: Multilingual Matters.
- Roever, J.E. (1977). The Depolarization of Laboratory and Field Research: A Matter of Internal and External Validity. *ETC: A Review of General Semantics*, Vol. 34, No. 3, pp. 339-347.
- Skolverket. (2022a). *Betyg och studieresultat i gymnasieskolan 2022*.
<https://www.skolverket.se/getFile?file=10522>
- Skolverket. (2022b). *Slutbetyg i grundskolan våren 2022*.
<https://www.skolverket.se/download/18.2cb9d0c18340c7dae395d/1664351539747/pdf10334.pdf>
- Sveriges Kommuner och Landsting. (2018). *Könsskillnader i skolresultat*.
<https://skr.se/download/18.45167e4317e2b341b24e8e99/1643114468911/7585-724-4.pdf>
- Thagg Fisher, U. (1985). *The sweet sound of concord: A study of Swedish learners' concord problems in English* (Lund studies in English, 73) [Doctoral dissertation, Lund University].

Appendices

Appendix I

Variable settings in the corpus:

1. No setting for age was made, as the focus was on the grades, either 9th or 11th.
2. Gender: binary male or female
3. There is no setting for county; ten different counties are represented in the corpus.
4. English-speaking school? No, only those who attend Swedish-speaking schools are in scope.
5. Previous attendance at an English-speaking school; No, as it is not certain how many years these students attended an English-speaking school, the decision was made not to include them in the study to ensure its validity.
6. School year 9th or 11th.
7. The type of program is only applicable for high school students: academic and vocational.
8. There is no setting for individual programs among high school students.

9. English course 6. Some students from vocational programs study English 5 in 11th grade, but they are not included in the study.
10. Swedish course, only applicable for high school students: Swedish (not as L2). This information is not as important as Swedish as L1, as students with another L1 can choose the subject Swedish at high school level, while assessments at school should be made at junior high school level.
11. L1: only students with L1 Swedish are in the scope of the study. Bilingual Swedish speakers are excluded.
12. Parents' L1: Swedish only
13. Birth country: Sweden
14. Language spoken at home: Swedish only
15. Other languages studied: no settings, since most junior and high school students study either German, French, or Spanish as the second foreign language in the Swedish school system.
16. There are no settings for EE.
17. Preparatory lesson: No
18. Writing time: 55 minutes for 9th grade students, 50 minutes for vocational program students, and 80 minutes for academic program students. These options generated most student texts, so random choice was still possible.
19. Graded by teacher: no setting as it is not crucial for the study.

Appendix II

Information about the chosen texts (detailed information in the section Primary Sources)

1. 9th grade male students (6 texts randomly sampled)
2. 9th grade female students (6 texts randomly sampled)
3. 11th grade male students attending academic programs (3 texts randomly sampled)
4. 11th grade female students attending academic programs (3 texts randomly sampled)
5. 11th grade male students attending vocational programs (3 texts randomly sampled)
6. 11th grade female students attending vocational programs (3 texts randomly sampled)