I JUST DOESN'T KNOW

Agreement errors in English texts by Norwegian L2 learners: Causes and remedies

Lenka Garshol

I JUST DOESN'T KNOW

Agreement errors in English texts by Norwegian L2 learners:

Causes and remedies

Doctoral Dissertation

University of Agder

Faculty of Humanities and Education

2019

Acknowledgements

Writing a Ph.D. dissertation can be an exhausting process and I could never have finished without the help and support of my family, friends, and colleagues. First and foremost, I would like to thank my supervisor, Dagmar Haumann, who has supported me throughout the whole process from my first fuzzy idea of a project to the finished dissertation. I would also like to thank my co-supervisor, Barbara Gawronska, for her practical advice and a view from outside of my own research bubble, and Linda Wheeldon for her advice and suggestions.

I would also like to thank Karoline Aksnes, my doctoral twin, with whom I could share all the ups and downs during this journey, and, of course, the rest of the morning coffee gang for creating a safe place to procrastinate. I also want to thank Mike Prince for being my friend, moral support, and tea-mate from the very first day we met in the university bookstore many years ago, and to Susan Erdmann for always having her door open for me and for the countless hours of discussions, both professional and personal, we have had over the years.

A special thank you goes to my boys who have filled my free time with highly needed non-thesis-related activities, and, most importantly, to my husband who has taken the kids for vacations when I need to focus the most. Thank you for the everyday acts of kindness which always light up my day.

Lenka Garshol
Kristiansand, October 2018

Abstract

This thesis discusses subject-verb agreement errors in written English production of Norwegian high school students. The first part of the project is devoted to an exploration of a corpus of L2 English texts written by young Norwegian learners with a focus on the description of subject-verb agreement errors produced by these learners. The theoretical framework used for the analysis is the Minimalist Program (Chomsky, 1995). The second part of the thesis describes a didactic intervention aimed at increasing the agreement marking accuracy by increasing the metalinguistic knowledge of the students using the Inverted Classroom Method. The results of the intervention are evaluated as a development in agreement marking accuracy between the intervention and the comparison group.

Subject-verb agreement is notoriously difficult for L2 learners, and morphological variability in L2 production often extends well into the advanced stages for many learners. However, when there are problems with morphology in L2 production, it is the lack of inflection that is observed, i.e. the unmarked version of the verb is used. When the inflections are supplied they are appropriate and faulty agreement is reported to be very rare (White, 2003). However, the error patterns found in the Norwegian data deviate from what has been commonly reported among learners with other L1s. Agreement errors from the corpus are divided into suppletive and affixal errors depending on the type of finite verb involved (*BE* or other) on the type of subject (based on the head of the subject phrase: noun, personal pronoun, indefinite pronoun, demonstrative pronoun, expletive, or clause as subject). The analysis shows that young Norwegian learners of English both overproduce the 3rd person singular marker -*s* and produce incorrect suppletive agreement.

The general accuracy in agreement marking is high in this population of learners (on average around 90%), but when they produce agreement errors, these are as often illicit agreement as lack of agreement. If the agreement errors were caused by performance issues, it would be expected that the learners would do significantly better in grammaticality judgment test than in production. However, Jensen, Westergaard, & Slabakova (2017) report that Norwegian learners accept illicitly marked verbs in up to 40% of the tested sentences. Since the general accuracy of agreement marking in this learner population is high, it is also unlikely that the errors could be caused by a lack of acquisition of the agreement features in English. In addition, similar error pattern,

although less pronounced, is also found in data from Swedish learners (Thagg Fisher, 1985). It is thus proposed that a potential reason for the described error pattern could be L1 influence. It is argued that based on a superficial similarity between the L1 and L2, the Norwegian learners might have misanalyzed their English input creating an incorrect analogy with their L1. This result is compatible both with the Fundamental Differences Hypothesis (Bley-Vroman, 1990) and the Features Reassembly Hypothesis (Lardiere, 2008).

Since the data are semi-longitudinal, a developmental perspective is also considered. The texts of each student are divided into three measurement points (Fall, Midterm, and Spring) and the error rates are compared across these measurement points. Even though some students improve their agreement marking accuracy during the collection period, the differences are not statistically significant on the group level. The agreement error rates from the students who contributed to the corpus are used as the backdrop against which the intervention data are evaluated. Based on the data from the corpus, I selected 12 grammar and structure/discourse related topics which seem to be problematic for many Norwegian high school students. These topics were then incorporated into a 12-module Inverted Classroom course in English grammar on the Campus Inkrement platform. The Inverted Classroom Method (ICM) is a didactic method which 'flips' the traditional Presentation-Practice-Production model. The ICM moves the presentation of new topics and controlled practice into the homework domain via a short video lecture with integrated exercises. This enables the teachers to use more time for production (written or oral), guidance, and feedback in the classroom. The purpose of the Inverted Classroom course is to give teachers a tool to introduce additional explicit grammar instruction without using classroom time for teacher-centered activities. The grammar instruction can be adjusted to the level of the learners, and the individual modules can be used repeatedly for revision. Classroom time can then be used for learner-centered communicative activities.

High school students from the same schools as in the corpus compilation process were enrolled in the intervention group. These students were given access to the 12 module Inverted Classroom course in English grammar. The students from the two groups were matched for schools, study programs, and their English teachers. Error scores in the three measurement points were compared to see whether the students exposed to additional explicit grammar instruction make fewer agreement errors than the students

from the comparison group. There were no significant differences in the overall number of words produced or the error scores between the comparison group and the intervention group. There were also no significant differences found in the development of these two groups regarding agreement marking accuracy during the school year. However, a post hoc analysis of log-in data revealed that a majority of the students did not use the ICM material as intended, which could explain why there was no effect found. In order to test the effect of the Inverted Classroom Method, stricter control of the actual use of the material would have to be implemented in further research.

Table of contents

List of tables

List of figures

Introduction and overview of the thesis

The objectives of studying second language acquisition (SLA) with the use of learner corpora are among others an increased understanding of the principles of interlanguage of learners of specific L1 backgrounds and a possibility to develop targeted didactic tools and methods which could focus on the problematic areas (Granger, 2012, p. 8). Both of these objectives are represented in this project. Part I of the thesis focuses on the interlanguage of intermediate to advanced Norwegian learners of English with a specific focus on subject-verb agreement errors. Part II discusses a didactic intervention designed to enhance the explicit grammar knowledge and metalinguistic awareness of these learners, again with a specific focus on agreement marking accuracy.

The first part of the thesis reports on an analysis of subject-verb agreement errors in a corpus of L2 English texts collected from a group of Norwegian high school students. The agreement errors are analyzed within the framework of the Minimalist Program and corresponding language acquisition theories. The analysis of the agreement errors in the corpus is conducted with the aim to identify the structures that are especially difficult for the learners. The corpus data are screened for subject-verb agreement errors which are further sorted and analyzed according to the type of subject (pronominal, nominal, clauses as subjects, etc.) and the type of verb (*BE*, i.e. suppletive agreement, or other verbs, i.e. affixal agreement). The results of the analysis are then compared to published and unpublished data from other learner groups and discussed in the light of different acquisition theories.

The second part of the thesis describes an intervention study conducted using the Inverted Classroom Method to teach English grammar. The learner corpus from the first part is used to develop an Inverted Classroom English grammar course for Norwegian high school students which is tested on an intervention group of students. The students who contributed to the corpus in the first part of the project serve as a comparison group in the intervention. The goal of the intervention is to develop and test a linguistically informed didactic tool which could be used in the English instruction in Norwegian schools to help learners to increase their accuracy in the L2. Based on the learner corpus, 12 grammar and discourse related topics which are problematic for many Norwegian high school students were selected. These topics were then incorporated into a 12-module Inverted Classroom English grammar course on the Campus Inkrement platform. The Inverted Classroom Method (ICM) is a didactic method which 'flips' the traditional

Presentation-Practice-Production model (PPP). The ICM moves the presentation of new topics and controlled practice into the homework domain via a short video lecture with integrated exercises. This enables the teachers to use more time for production (written or oral), guidance, and feedback in the classroom. The course is available online and thus usable by any teachers (and students) regardless their location.

The thesis is structured as follows. Part I provides the theoretical background for the morphosyntactic analysis of the corpus data, including a review of previous studies of agreement marking acquisition in L2 English. The corpus compilation methodology is also included in this part. The results of the analysis of the erroneous structures in the learner data are then compared to the reviewed theories of language acquisition. Part II includes an overview of the current trends and theories in the field of instructed second language acquisition (ISLA), including an overview of previous studies using the Inverted Classroom Methodology. The design and development of the Inverted Classroom course in English grammar is discussed in detail before the data from the two groups are compared and statistically tested. This part also includes an evaluation of the intervention procedure. A detailed overview of the chapters of the thesis is included in the introductions to the respective parts.

Part I

1 Introduction

The first part of this thesis is devoted to an exploratory study of a learner corpus. The corpus has been collected and compiled in cooperation with high school teachers from several schools in the southern part of Norway. Subject-verb agreement errors have been chosen as the main focus for the analysis. The first part of the thesis aims at a qualitative and quantitative description of the most common types of non-standard subject-verb agreement marking in the written production of young Norwegian L2 learners. The error types are divided into categories based on the morphosyntactic qualities of the involved clause elements. The frequencies of the errors are then assessed within each category and across categories. The final part of the analysis also includes a discussion of potential sources of the errors other than faulty or incomplete acquisition of agreement. The Norwegian data are also compared to data from other published and unpublished sources, namely data from Swedish and Austrian-German learners of English.

Based on the analysis and the comparison of non-standard subject-verb agreement marking in the learner Englishes under consideration, several hypotheses regarding L2 development are evaluated. As the analysis of the data is based on an exploration of a corpus, it is not possible to address any issues on an individual level. The tendencies which are identified on the group level can be summarized into two research questions. The first question of interest is whether the observed inaccuracy in agreement marking can be ascribed to performance errors or whether the agreement acquisition in this population can be considered faulty or incomplete. The second question, in case the agreement acquisition is indeed faulty or incomplete, is which hypothesis of L2 development could best explain this outcome. As it is not possible, on the basis of the corpus observation alone, to conclude with certainty which hypothesis is adequate, some questions for further corpus and experimental research are discussed in the end of the first part of the thesis.

Chapter 2 provides a brief introduction to the Minimalist Program framework and its implications for language acquisition. The patterns of verb movement and subject-verb agreement marking in English and Norwegian are briefly described here as well. Different hypothesis regarding second language acquisition in general and predictions for the acquisition of agreement in particular are also evaluated. The chapter concludes with a

brief description of the role of learner corpora in the second language acquisition research and an overview of the corpora used in the analysis.

Chapter 3 reviews some previous studies of agreement marking in L2 English to provide a backdrop against which the Norwegian data are analyzed. Studies of learners from different L1 backgrounds are included to addresses the potential of crosslinguistic influence in some of the cases of non-standard agreement marking in the Norwegian data. Even though the Norwegian learners are young, their proficiency in English is fairly high. Both adult and child L2 studies are therefore included in this chapter. Most of the reviewed studies are based on corpus data, but also one grammaticality judgement study is included to serve in the discussion of competence vs. performance errors in learner language.

Chapters 4 and 5 focus on the corpus data. Chapter 4 describes the collection and compilation procedures used in the corpus. This chapter also provides some basic descriptions of the subjects and text types included. Chapter 5 focuses on the analysis of the agreement data within the subject and verb categories outlined above. After the analysis of the Norwegian data, a comparison with data from two other learner groups (Swedish and German L1s) is conducted.

Chapter 6 provides a discussion of the findings from the corpus analysis in the light of the theories discussed in chapter 2. Several hypotheses are outlined and two are considered plausible after the discussion. However, more research is necessary to provide any firm conclusions. Chapter 7 summarizes the findings and suggests topics for further research.

2 Theoretical framework

This chapter is devoted to the theoretical background of the corpus analysis performed in chapters 5 and 6. Corpus analyses can be performed within several theoretical frameworks, but the corpus as such is seen here rather as a tool for data systematization than as a framework connected to a specific theory of language. The language data extracted from the corpus are analyzed syntactically within the Minimalist program (see Chomsky, 1995). However, since this study does not focus on the language production on the individual level but rather tries to generalize on the level of a group of speakers, the frequencies of the observed phenomena in the corpus must also be taken into consideration. Such a combination of approaches is seen as beneficial for a potential generalization of the findings for wider populations of second language learners.

Sections 2.1 and 2.2 introduce basic assumptions regarding syntactic structure in the Minimalist Program in general and the syntax of subject-verb agreement in particular. Agreement marking and verb movement in English and Norwegian is briefly contrasted as well. Sections 2.3 and 2.4 deal with the generative theories of second language acquisition (SLA) and the role of Universal Grammar (UG) in the process of L2 acquisition. Finally, sections 2.5 and 2.6 of this chapter outline the role of learner corpora in SLA research and briefly describe the features of the corpora used in this study.

2.1 Syntactic structure in the Minimalist Program

In the Minimalist Program (Chomsky, 1995), syntactic structure is seen as the result of successive applications of the operations MERGE, AGREE, and MOVE (see also Franck, Frauenfelder, & Rizzi, 2007, p. 174). These operations apply to basic syntactic elements creating larger structures. Under the current understanding, morphosyntactic features are considered the smallest 'atoms' of language (Adger, 2003, p. 22) and lexical items are then considered to be 'bundles of features' (Adger, 2003, p. 62; Adger & Svenonius, 2011) which enter into syntactic derivations at specific places based on the features they possess. Features can be interpretable or uninterpretable depending on whether they have an effect on the semantic representation or not.

The three basic syntactic operations may apply cyclically, and their various combinations create all possible structures of natural languages. MERGE joins any two syntactic elements together creating a hierarchical structure (Adger, 2003, pp. 62, 69) as represented in *Figure 1*:
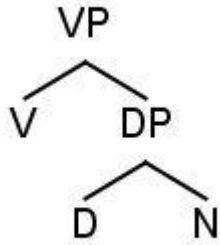
Figure 1. Basic hierarchical tree structure.

MOVE displaces elements already present in the structure leaving a trace copy in the original position (Radford, 2004, pp. 152, 190), *Figure 2* (adapted from Radford, 2004):
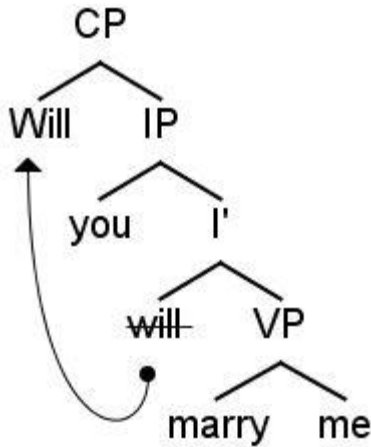
Figure 2. Movement of the finite non-lexical verb in question formation.

AGREE in general is a feature checking operation. A probe with unvalued features searches for a goal within its local domain of c-command. When the appropriate goal with matching features is found, the features are 'checked by valuing.' An unvalued uninterpretable feature will then get a value from a matching interpretable feature and be checked at the same time (see Adger, 2003, pp. 167-169; Radford, 2004, pp. 282-287).

Features relevant for the discussion at hand are φ-features, which traditionally include person, number, and gender features (Adger, 2003, pp. 24, 45). Φ-features are considered uninterpretable on the head of IP (Chomsky, 2000) and must thus be checked during the syntactic derivation. In order for the semantic component to be able to process the syntactic structure, all the features must be interpretable (Full Interpretation

Constraint, see also Radford, 2004, pp. 288-289). Thus, any uninterpretable features must be checked and deleted before the structure enters the semantics interface (Checking Requirement, Adger, 2003, p. 85; Radford, 2004, p. 143). Another feature property relevant for the description of subject-verb agreement is feature strength. Features are either strong or weak: strong features must be checked locally, while weak features may be checked at a distance (Adger, 2003, p. 179). When a strong uninterpretable feature, e.g. tense on a verb, requires checking, the constituent with the feature must move to the node which contains the matching interpretable feature, in the case of tense I. The difference in feature strength between different languages influences the head movement patterns in different structures in these languages (see below for verb movement in English and Norwegian).

2.2 The syntax of subject-verb agreement

In the case of subject-verb agreement, AGREE operates on the relation between the functional node I, which hosts uninterpretable φ-features, and the subject still in the VP. The interpretable φ-features of the subject are copied onto I (AgrS in Franck et al., 2007).[1] In some languages (e.g. French), the verb then moves from V to I, while in other languages (e.g. English), the verb remains in situ and the subject moves out of the VP to SpecIP (Franck et al., 2007; Pollock, 1989). Norwegian exhibits verb movement only in main clauses, while the finite verb stays in-situ in subordinate clauses (referred to as asymmetric V2 language, see Westergård, 2009, pp. 16-18). In languages where the finite lexical verb stays in situ the agreement is instantiated via 'affix lowering' or 'affix hopping' from I to V (Radford, 2004, p. 118), illustrated in *Figure 3* (adapted from Radford, 2004):

---

[1] Some researchers propose that Agreement and Tense features are located in IP (or TP), while others propose separate functional nodes: AgrP and TP (see Pollock, 1989) or AgrS, TP, AgrO (see Franck et al., 2007). Since neither English nor standard Norwegian mark agreement with objects or participles, which would need to be checked in AgrO, this distinction will not be discussed further here. Both agreement and tense features are henceforth assumed to be located in IP-domain without the assumption of any particular ordering or subdivision into other heads within this domain (see also Julien, 2002).
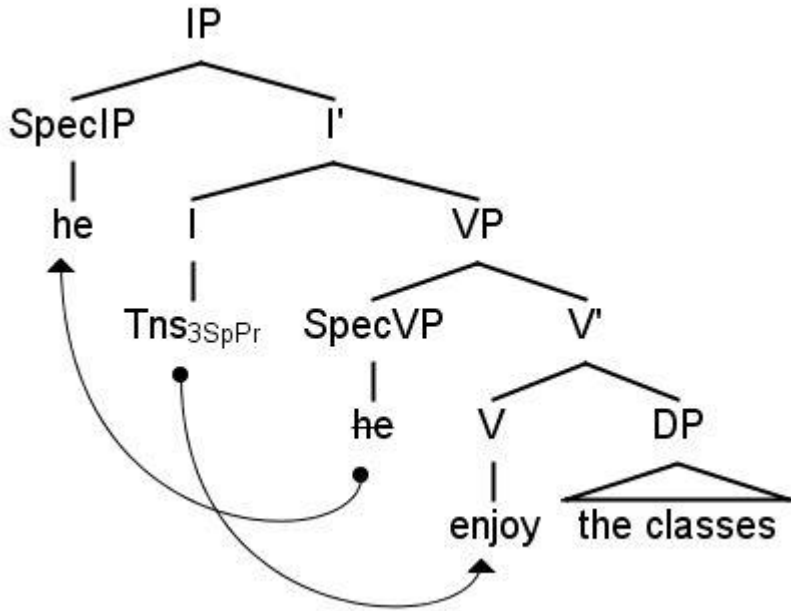
Figure 3. Affix lowering.

Since agreement marking is related to head movement in some languages, it is necessary to briefly describe both the verb movement patterns and agreement marking principles in English and Norwegian.

2.2.1 Verb movement and agreement marking in English

In English, the I node carries an interpretable tense feature and an uninterpretable φ-feature (in addition to other features such as Case and EPP). The subject, still in VP, carries an interpretable φ-feature (number) which needs to be copied onto I via AGREE in order to check the uninterpretable φ-feature in I (Adger, 2003, pp. 220-221). The finite verb then receives its tense and number features from I. In English, only finite non-lexical verbs (modal and non-modal auxiliaries) carry a strong tense feature, which means that they raise to I for the tense feature to be checked.[2] Conversely, tense on lexical verbs is checked by affix lowering, with the verb remaining in situ (Adger, 2003, p. 182).

Present-day English only marks agreement on lexical verbs and non-modal auxiliaries in the present tense, and only the 3rd person singular of lexical verbs receives

---

[2] Auxiliaries in English are generated above VP. There are different assumptions in the literature as to where and whether this depends on the type of auxiliary and its scope (see Adger, 2003, pp. 177-181; Radford, 2004, pp. 166-170; Roberts, 1998). Finite auxiliaries raise to I while non-finite remain in situ.

an overt marker -*s*. The current conjugation pattern is a remnant of a more complex one in Old English which was gradually lost in the subsequent varieties (Barber, Beal, & Shaw, 2009, pp. 123-126). Modal auxiliaries, which lost their inflectional properties in the course of grammaticalization (Roberts, 1993), do not mark agreement overtly in present-day English. The relevant forms for lexical verbs are reviewed in Table 1:

Table 1

Conjugation pattern of English lexical verbs in the present tense

	Singular	Plural
1st person	I give	we give
2nd person	you give	you give
3rd person	he/she/it give**s**	they give

In addition, the verb *BE* marks overt agreement both in the present tense and in the past tense with suppletive forms:

Table 2

Conjugation patters of the verb BE in the present and the past tense

	Present singular	Present plural	Past singular	Past plural
1st person	I **am**	we **are**	I **was**	we **were**
2nd person	you **are**	you **are**	you **were**	you **were**
3rd person	he/she/it **is**	they **are**	he/she/it **was**	they **were**

Because of the different raising patterns for English finite auxiliaries and lexical verbs it has been argued that agreement marking on auxiliaries is 'easier' and thus acquired earlier than agreement on lexical verbs (White, 1992). Ionin and Wexler (2002, p. 116) subscribe to this interpretation and propose that L2 learners initially associate morphological agreement with overt movement to I. This proposal implies that L2 learners should make more errors in agreement on lexical verbs than on auxiliaries. This is further discussed in chapters 3.3 and 5.4.2.

2.2.2 Verb movement and agreement marking in Norwegian

In Norwegian, verbs have only weak tense features: neither finite non-lexical verbs nor finite lexical verbs raise to I to check the tense (see Adger, 2003, p. 184 for Swedish). However, Norwegian is also a V2-language, which means that finite verbs in main

clauses always raise to C. Due to the Locality Principle, which states that all grammatical operations are local (Radford, 2004, p. 15), this movement must happen in steps as a V-to-I-to-C movement (see among others Westergård, 2009). As the verb passes via I it checks the tense feature which is generated there (Åfarli & Eide, 2003, p. 80).

Unlike English, present-day Norwegian (and other Mainland Scandinavian languages) does not mark subject-verb agreement overtly. Although Old Norse (like Old English) had rich verbal agreement (Faarlund, 2004, pp. 200-204), agreement morphology gradually disappeared and is absent in present-day Norwegian. There are thus no uninterpretable φ-features in I which would need checking, and the finite verb only needs to check the tense features in I.

Another contrast between English and Norwegian lies in tense marking. Norwegian has an obligatory present tense marker: a bound morpheme *-(e)r*. Finite verbs are thus always overtly marked for tense. Due to the lack of agreement, the present tense conjugation pattern is fully regular:

Table 3

Conjugation pattern of Norwegian lexical verbs in the present tense exemplified for the verb GI 'give'

	Singular	Plural
1st person	jeg gi**r**	vi gi**r**
2nd person	du gi**r**	dere gi**r**
3rd person	han/hun/det gi**r**	de gi**r**

Finite forms of the verb *BE* are suppletive, but the conjugation itself is fully regular, both in the present and in the past tense:

Table 4

Conjugation pattern of the Norwegian verb VÆRE 'be' in the present and the past tense

	Present singular	Present plural	Past singular	Past plural
1st person	jeg **er**	vi **er**	jeg **var**	vi **var**
2nd person	du **er**	dere **er**	du **var**	dere **var**
3rd person	han/hun/det **er**	de **er**	han/hun/det **var**	de **var**

Modal verbs in Norwegian have suppletive forms in both the present tense and the past tense, but as they do not mark agreement, these forms are identical for all person-number combinations, i.e. 'er' in the present tense and 'var' in the past tense.

Ionin & Wexler's proposal (2002) regarding an initial association of agreement with movement could have an influence on the assumptions Norwegian learners of English make in their L2.[3] English and Norwegian share many syntactic properties and it is conceivable that learners may base their assumptions about English on their first language. The Norwegian V-to-I-to-C movement could superficially mimic the English auxiliary movement to I. Thus, if the learners would assume, as Ionin & Wexler propose, that verb raising is associated with agreement, they could also assume that the Norwegian present tense marker and the English agreement marker have the same function. However, Norwegian finite verbs do not move in subordinate clauses, which complicates the picture for the learners.

2.3   The generative approach to language acquisition

The generative approach to language acquisition is based on 'Plato's problem' or the 'poverty of the stimulus argument', i.e. the fact that humans are able to acquire extremely fine-grained and abstract facts about the language(s) they are exposed to with only limited evidence provided in the input (Chomsky, 1980, 1986). Virtually all children exposed to any given natural language are able to acquire this language and eventually become competent native speakers whose judgments about their native language are close to identical. Given that children might receive input in different quality and quantity, there must be some mechanism in the human cognitive system which facilitates language acquisition and helps to bridge the gap between the imperfect input and the final complete language system. This mechanism is often referred to as language faculty and it is assumed to be innate.

The language faculty enables children to acquire any language(s) they are exposed to quickly and successfully despite limited and imperfect input. A theory of Universal Grammar (UG) describes the language faculty as a set of general principles which are valid in all natural languages and a set of parameters which can differ from language to language. When children are exposed to language, they set the parameters to the

---

[3]   By 'assumptions' I mean unconscious processes involved in language production and processing.

appropriate values based on their input (Radford, 2004, p. 21). Current research shows that also children exposed to several languages simultaneously are able to acquire all the languages they are exposed to, given they receive sufficient input in each of them (De Houwer, 2009; Genesee, 2005; Paradis & Genesee, 1996). In other words, the system of principles and parameters within UG does not limit the number of distinct languages one mind can acquire. However, whether UG stays available throughout the lifetime or whether its resources are limited to first language acquisition is still debated.

2.3.1 Universal Grammar in second language acquisition

One of the key questions in generative second language acquisition research is whether UG is available for L2 learners and if so, to what degree they can access it. The second, but related, question concerns what the starting point of the learners is or, in other words, the initial state of their L2 competence. White (2003) convincingly argues that L2 acquisition suffers from the same poverty of the stimulus problem as L1 acquisition: L2 speakers seem to have implicit knowledge of subtle abstract linguistic phenomena which are not present in their L1, not easily available in the input and not taught in L2 classrooms (White, 2003, pp. 22-56). The generative explanation of this logical problem is that UG must be available also in SLA, at least to some degree.

2.3.1.1 The initial state of the L2 competence

As for the initial state of the L2 competence, learners must either start with their L1 knowledge, or they start with a blank slate and the options provided by UG. White (2003) argues for the Full Transfer Full Access Hypothesis of second language acquisition, i.e. that all L1 knowledge is transferred initially, while all of UG stays available for the L2 learner and interacts with the L2 input. The task of the L2 learner is then to evaluate whether the parametric options instantiated in L1 are compatible with the L2 input and, if necessary, reset the parameters of the developing L2 grammar. This implies that interlanguage grammars at all stages should be constrained by UG and thus constitute natural languages (but not necessarily conform to the L2, or even the L1 of the learners). The Full Transfer Full Access Hypothesis makes several predictions about both the initial state of the L2 competence and the progress towards ultimate attainment. If learners initially transfer all of their L1 competence, then learners of different L1s should have problems in different areas of the same L2. Furthermore, as soon as a parameter mismatch

is detected, and the parameter is reset in the developing L2 grammar, all related phenomena should fall into place almost instantaneously. Consequently, L2 competence should converge, over time, more and more to the competence of native speakers.

There is compelling evidence of initial L1 transfer from both longitudinal case studies (Haznedar, 1997) and several experimental studies (see White, 2003, pp. 61-66). The studies indicate that at least some transfer of the L1 knowledge happens early on in the porcess of L2 acquisition since learners with different L1s take different paths in their acquisition of the same L2 (Schwartz & Sprouse, 1996; Slabakova, 2000). However, the evidence for the instantaneous resetting of parameters is not that compelling. White (2003, pp. 148-149) evaluates several studies which look at the L2 development past the initial state and concludes that, even though many learners apparently do reset the parameters in their L2 grammars, some may arrive at a setting that reflects neither their L1 nor the L2 they are learning. Learners also often show prolonged periods in which the properties of L1 and L2 co-occur, and sometimes the properties associated with one parameter do not cluster in the L2 as they do in L1 (White, 1992). However, the main problem of the Full Transfer Full Access Hypothesis is that it is potentially unfalsifiable. If experimental results show that learners of different L1s perform similarly on properties of L2 which should differ based on their L1, it can always be claimed that they are already past their initial state. Thus, L1 effects are not visible (White, 2003, p. 67). Similarly, if experimental results show that learners who are past the initial state do not reset the L2 values correctly, or reset some but not others, it can always be argued that they have not yet reached the necessary amount of exposure to the L2, or that performance problems blur the results.

Lardiere (1998a, 1998b, 2008, 2009) suggests that the problem with parameter resetting in L2, as evidenced by the prolonged variability and the potential setting which does not reflect either of the involved languages, might be rooted in the concept of 'parameter' as such. Parameters fail to appropriately capture some of the problems in second language acquisition. As mentioned above, a parameter resetting approach would predict that, as soon as a specific parameter value is reset, all features associated with this parameter should fall in place almost instantaneously and the production should be close to error-free. However, this is not the case as L2 learners often go through lengthy periods of variability and optionality in production (Lardiere, 2008, p. 108). Following Chomsky (1995), Lardiere suggests that a more accurate description of parametric variation is a

description of features represented by morphological properties of functional categories. These features are bundled differently in different languages, which then results in what was earlier described as a parametric variation (see also Lardiere, 1998a, 1998b, 2009). The task of the L2 learner is to evaluate whether the configuration of features in their L1 corresponds to that in L2 and if not, to re-assemble the features (Lardiere, 2009). As Lardiere (2008) points out, this approach goes beyond simple feature-selection from the UG options as the learner must also evaluate which features are bundled together and on which lexical items they are realized. If there are discrepancies in this regard between the L1 and L2, periods of L1-like feature selection and bundling in the L2 are expected. The feature reassembly account would thus predict periods when some features associated with a particular parameter approximate the setting in L2, while others are still set to L1. It would also be conceivable to predict a fossilization of such a state, i.e. some L2 learners might never bundle the features correctly in their L2. Such learners would use the L1-assembly of the features in their L2 despite the fact that they have correctly reset the values of said features, and even acquired the appropriate morphology, as in White's (1992) study of verb raising in French L2 English learners.

A different view of the logical problem of L2 acquisition has been put forward by Bley-Vroman (1990). He proposes the Fundamental Difference Hypothesis which states that L1 and L2 acquisition do not operate on the same premises. Since children do not have any language to resort to when they are born, they rely solely on the principles and parameters of UG and some set of 'domain-specific learning procedures.' Adult L2 learners, on the other hand, have already acquired a language and can thus use this knowledge as their starting point. In addition, adults have developed 'general problem-solving systems,' which are not available to young children (Bley-Vroman, 1990, pp. 13-14). Bley-Vroman argues that UG is only indirectly accessible to adults, namely through their first language. The most compelling argument in favor of the Fundamental Difference Hypothesis is the clear difference in the outcome: while virtually all healthy children who are provided with linguistic input acquire their L1 perfectly, native-like competence is not common in adult L2 learners. Bley-Vroman further argues that what is often interpreted as 'UG effects' in adult L2 learners can, in fact, be mimicked by other factors available to the adult learners, such as native language analogy, availability of rich data (e.g. organized presentation of data, explicit statements of rules, etc.), or learning of UG consequences in a piece-meal fashion (Bley-Vroman, 1990, pp. 39-41). The

Fundamental Difference Hypothesis would thus also predict that adult L2 learners could go through prolonged periods of variability and perhaps even fossilize in the state when some parameters (or individual features) have been reset while others have not.

It is difficult to judge whether the apparent UG effects (reported among others by White, 1990, 2003) are in fact caused by direct or indirect (via L1) access to UG. Clearly, many L2 learners never achieve a perfect native-like competence, which would support Bley-Vroman's argument of unavailability of UG. However, there is also evidence that many subtle syntactic features are very resistant to classroom instruction with explicit statements of rules (VanPatten, 2011). This would contradict the argument that adults can compensate the lack of UG access by access to rich data and general problem-solving systems.

2.3.1.2   Interlanguage development

There are various conflicting hypotheses regarding the interlanguage development after the learners move past the initial state as well. There is a wide consensus within the generative community that only primary linguistic data (PLD) can serve as cues in language acquisition. These cues can interact with the internal language system and trigger the restructuring of the system, i.e. parameter resetting. Conversely, negative data (be it corrections or explicit information about grammaticality of the lack thereof in a given language) cannot serve as cues in SLA (Schwartz, 1993, p. 153; White, 2003, p. 164). However, some experimental studies show exactly the opposite effect. White's (1991) study of the effect of different instruction types on adverb placement by French students of English showed that explicit negative evidence was the only intervention type which successfully changed the students' linguistic behavior. These results were interpreted by Schwartz (1993) and later adopted also by White (2003) as changes to 'learned linguistic knowledge' and not true parameter resetting (see also the discussion in chapter 9.1). However, regardless of what changed in the learners' language system, the result was a change in production. If L2 users are conceived of as a subclass or variety of L1 users of the given language, then it can be argued that explicit negative evidence should not have any effect on their language system because it does not affect young children learning their L1. On the other hand, if L2 users are considered to be a different type of language users, as under the Fundamental Differences Hypothesis, their language system might be able to incorporate also other types of evidence than just PLD. Even at

stages when this learned linguistic knowledge is not fully integrated and automatized, it may serve as a monitor in careful production, such as in written tasks or untimed grammaticality judgment tasks. White (1991), contrary to Bley-Vroman (1990), argues that the incorporation of negative evidence into the L2 system does not contradict the full availability of UG parameters in L2 acquisition. L1 users do not use negative evidence because they do not need it as there is no previously acquired language in their system which would influence their interpretation of the input. The L2 learners, on the other hand, may benefit form negative evidence which can serve to point out the discrepancies between the interlanguage and the target.

An alternative option to parameter resetting in L2, a 'weak UG' view is proposed by Clahsen and Hong (1995). They propose that only the principles of UG are available in L2 acquisition, the array of UG parameters is not. Instead of instantaneous parameter setting based on cues in the input, the surface properties associated with the parameters would have to be learned separately by 'pattern matching' (see also White, 2003, p. 101), i.e. the learners compare the surface structures of the L2 and the representations they have of their L1 and adjust the patterns which are set differently for L1 and L2. Such a view would be compatible with Bley-Vroman's (1990) Fundamental Difference Hypothesis and it would also explain why White's (1991) French students exposed to explicit instruction stopped using fully grammatical post-verbal adverbs, which only superficially resemble the illicit verb-adverb-object word order in English. They may have overgeneralized the rule they have been taught based on pattern matching. A similar effect could also be ascribed to the feature reassembly approach to SLA. If it is assumed that the L2 learners need to reassemble the feature bundles from the L1 to the L2 composition, it is conceivable that some features might be assigned incorrectly based on the input. This could result in native-like behavior in some cases while some non-native patterns would continue to emerge in production for lengthy periods or perhaps become fossilized.

All these approaches expect full L1 transfer. Given the amount of experimental evidence of L1 effects in the L2 acquisition, it seems unlikely that L2 learners would somehow refrain from using their L1 knowledge when they first approach their L2. In addition, since all L2 learners have already successfully acquired at least one language, it is also expected that they have some general knowledge of the properties of human languages which can mediate the process of L2 acquisition. I will thus assume full L1

transfer in the following. However, the question of the degree of UG accessibility in L2 seems to be less clear, and I will return to this question in the discussion in chapter 6.

2.3.1.3   Morphological variability in L2

The discussion of L2 competence often relates to the knowledge of abstract properties of the L2 syntax, such as verb movement, null subjects, etc. There is usually less focus on L2 morphology in discussions of L2 proficiency because, as White (2003, p. 178) states, "it is well known that L2 learners exhibit optionality or variability in their use of verbal and nominal inflections and associated lexical items." She further stresses that missing or optional overt morphology should not be taken as evidence of a lack of acquisition of the associated abstract features. She supports this claim both on a theoretical level – an abstract feature does not always manifest with an overt form (e.g. some past tense forms in English are identical to the present tense forms) – and with several case studies of L2 acquisition which show that there is "dissociation between verbal inflection and various syntactic phenomena in production data [of the L2 speakers]" (White, 2003, pp. 180, 188).

However, White repeatedly states that, when there are problems with morphology in L2 production, it is the lack of inflection that is observed rather than superfluous inflection. When the inflections are supplied they are appropriate and faulty agreement is very rare (White, 2003, p. 183). In addition, White states, "suppletive forms are used extensively and are used accurately. In other words, the issue is the *degree* to which the learners supply agreement rather than the *accuracy* of agreement" (White, 2003, p. 196, italics in the original). Most of the published research on subject-verb agreement in L2 production confirms White's predictions (see chapter 3). When there are instances of inaccurate agreement in the L2 data, they are usually explained as performance errors by the authors and only a few studies include suppletive agreement at all. However, the picture might be more complex and the dismissal of incorrect agreement as performance errors can be premature. Comparison of production and grammaticality judgment data from the same or similar populations of L2 users could help distinguish between incomplete or inaccurate acquisition of the agreement features and production issues. Jensen, Westergård & Slabakova (2017) report unusually high numbers of misjudged agreement errors among young Norwegian learners of English. These learners accepted both sentences with missing agreement markers and with incorrect overtly marked

17

agreement. Their findings support the data reported in this study (see chapters 3.3 and 6.3).

2.4   Learner corpus research

Research approaches to second/foreign language learning have been developing rapidly since the use of digital technologies became common in the early 1990s. Especially the option of storing and analyzing large collections of learner data with automatized tools opened the doors to a more systematic research of the interlanguage of learners of various language backgrounds and L2 combinations (Hasselgård & Johansson, 2011; Tono & Díez-Bedmar, 2014). There are currently several large-scale multi-L1 learner corpora available, most notably the International Corpus of Learner English (hereafter ICLE; Granger, Dagneaux, Meunier, & Paquot, 2009), which consists of argumentative essays written by university students of English with 16 different L1 backgrounds. Multi-L1 corpora allow not only a comparison of the interlanguage of selected L1 groups to the language of native speakers, but also a comparison of different L1 groups to each other. Such an approach can help identify features of the interlanguage which are related to transfer from the L1 of the learners (Tono & Díez-Bedmar, 2014, p. 164). Recent research shows that a learner's L1 has greater influence on several aspects of the L2 interlanguage than has previously been assumed (Nesselhauf, 2004).

Learner corpus research has until recently concentrated mainly on advanced adult learners, partially because of the ease of data collection (Tono & Díez-Bedmar, 2014, p. 166). The International Corpus of Crosslinguistic Interlanguage (hereafter ICCI; Tono & Díez-Bedmar, 2014) is a notable exception as it includes data from young learners (third to twelfth grade) from eight language backgrounds. Most learner corpora are cross-sectional (such as ICLE), i.e. the data are collected from each learner only once, or quasi-longitudinal (such as ICCI), i.e. the data are collected from different learners at different stages of their development (Granger, 2012, p. 14). True longitudinal corpora are rare and usually limited to few contributing speakers. The corpus collected for this study aims at partially bridging this gap (see section 0 below). Longitudinal and quasi-longitudinal learner corpora can be a useful tool in tracking L2 learners' development over time, both for specific L1-L2 combinations and for L2 development in general. Focus on specific L1-L2 combinations offers a particularly useful insight for teachers and teaching materials developers (see chapter 10.6). Tracking the general L2 development in

longitudinal learner corpora may also help answer some conceptual questions about the L2 interlanguage development after the initial state.

2.4.1 Learner corpus-informed language instruction

Learner corpora can be used in language instruction both directly and indirectly. Recent development in Data-Driven Learning (DDL), i.e. the use of corpus data directly by the learners in an inductive learning setting, shows promising results (Boulton & Cobb, 2017). Such approaches highlight learner autonomy and bottom-up processes in the L2 instruction which are often recommended in grammar instruction (see among others Doughty & Williams, 1998). Most DDL studies involve the use of L1 corpora, but there is certainly a potential in including also learner corpora in the exploratory language tasks in the L2 classroom, especially with a focus on the differences between the L1 norm and the learner interlanguage.

However, the more typical use of learner corpora in L2 instruction is indirect, e.g. as a measurement of L2 proficiency, as a tool to track L2 development, or as a source of material for didactic tools. The use of learner corpora in a SLA setting has often been criticized for 'comparative fallacy,' i.e. comparing the interlanguage to the native norm (Bley-Vroman, 1983; see also Granger, 2012, p. 26). However, Granger points out that most other methods used in SLA research are also based on an underlying native speaker norm, e.g. proficiency tests, or grammaticality judgments (2012, p. 27). In addition, if the goal of the study is to suggest didactic tools to help the learners approach the (near) native norm, any other comparison seems superfluous. Nesselhauf (2004, p. 126) argues that the use of learner corpora to identify the areas which need specific didactic attention is particularly valuable. It has also been pointed out by the critics of the Contrastive Analysis Hypothesis that the comparison of the learner's L1 to the target L2 is not as informative regarding the common L2 problems as comparing the L2 produced by learners to the production of the native speakers of the same language (Nesselhauf, 2004, pp. 125-126). Learner corpora thus seem to be the perfect tool for interlanguage analysis.

In addition, an undeniable advantage of learner corpus data as compared to small-scale data collections and case studies is the increased generalizability of the results. As Gass & Selinker point out, "[i]t is difficult to know with any degree of certainty whether the results obtained [from case studies] are applicable only to the one or two learners studied, or whether they are indeed characteristic of a wide range of subjects" (Gass &

Selinker, 2001, p. 31). This problem is minimized by obtaining data from many learners and focusing on the frequencies of the studied features in the data instead of studying only a few particular instances produced by a small group of learners in a limited experimental setting (Hasselgård & Johansson, 2011, p. 37). Granger (2012, p. 17) suggests that "a corpus-driven approach exploits the full force of the corpus […] progressively generalizing from observation of data to building up the theory or rule." Large data collections from specific populations of learners should thus provide an opportunity to identify patterns of interlanguage which are common for many members of that population, which can then contribute to theory building both for specific L1-L2 combinations and for L2 interlanguage processes in general.

2.4.2 Limitations of learner corpora

Learner corpus research is, of course, not the answer for everything and there are several limitations to this approach. Individual variation among learners is mostly lost in a learner corpus. The learners contributing to the collection are seen as a more or less homogenous population and any individual differences are smoothened by the magnitude of the data. Individual differences in language production and interlanguage development can, of course, be an important factor in a general theory of L2 development and they are also one of the crucial factors in any didactic intervention. The questions on an individual level which cannot be answered in a corpus study must be addressed in follow-up studies.

Although measures have been taken to ensure as homogenous a population as possible for the collection (e.g. controlling age, school types, study tracks, language background, existence of learning related diagnoses, etc.), it is possible that some of the learners perform in specific ways for reasons which are not controlled for. It is also possible that the learners who agreed to contribute to the corpus are in fact not a representative sample, as some learners could have opted out from participating because they did not feel confident enough in their English. However, this is a general problem with any data collection from subjects because the participation is always voluntary.

2.4.3 Corpora used in this study

The corpus described in this thesis is a written mono-L1 semi-longitudinal corpus of young learner data. The contributors are 15-16-years-old Norwegian students attending their first year of high school, which is their 11th year of English instruction and schooling

in total. The corpus can be considered semi-longitudinal as the collection period is one year and most learners contributed with more than one text (usually three to four texts). Some background information is collected from the participants (see chapters 4.1 and 10). The collected material consists mainly of argumentative essays on given topics written at school with a bilingual dictionary as the only aid and a time restriction (usually 2-3 hours). A minority of the text contributions are analytical or descriptive texts (e.g. movie/book reviews) written at home. The collection and error extraction procedures are described in chapter 4.

The agreement error patterns of the Norwegian students are compared to a published study of learner corpus data of Swedish university students collected by Thagg Fisher (1985). Swedish is closely related to Norwegian so there is a potential for similar L1 influence on the L2 English among the speakers of these two languages. If Norwegian and Swedish L2 English learners produce similar non-standard patterns and such patterns are not attested among learners with other L1s, these could be a result of transfer (Jarvis & Pavlenko, 2008). Therefore, a sub-corpus of the ICCI (Tono & Díez-Bedmar, 2014) is used as a comparison sample in order to evaluate the possibility of L1 influence on the non-standard production of agreement among the Norwegian learners. This sub-corpus consists of data from Austrian learners matched for age (11th grade).[4] 68 students contributed to this sample (11,342 words). Even though German and Norwegian are related languages, unlike Norwegian, (Austrian) German marks agreement in all persons. It is thus less likely that the German speaking learners would experience the same L1 transfer as the Norwegian learners. If the German learners produce different types of errors than the Norwegian learners, this would be further evidence of a potential L1 cause of the non-standard patterns in the L2 English of Norwegian learners. However, the German data sample is very small and can only give an indication which would need to be confirmed by further research.

---

[4] I would like to thank Thomas Rankin for kindly sharing these data with me.

3 Previous studies of agreement in L2 English

Most of the studies of agreement in L2 English focus on adult learners, which is partially due to the availability of data. Until recently, most learner corpora also included only adult learners such as university students (e.g. ICLE, Granger et al., 2009) or other specific groups of adult L2 users (see Breiteneder, 2005; Dröschel, 2011). Corpora of younger learners are rare. Some recently compiled exceptions are the International Corpus of Crosslinguistic Interlanguage (ICCI, written data, Tono & Díez-Bedmar, 2014), the Corpus of Young Learner Interlanguage (CYLIL, spoken data, Housen, 2002), and the Corpus of Young Learner Language (CORYL, written data, Hasselgreen & Sundet, 2017) which are all quasi-longitudinal corpora of young learners (3rd to 12th grade, 9-18-year-olds). Even though there are many studies which discuss agreement in learner language, most of them do not analyze agreement errors in detail but rather compare agreement errors to other error types (Neff et al., 2007; Thewissen, 2015), focus on acquisition order (Abe & Tono, 2005; Izumi & Isahara, 2004), or on native language identification (Wong & Dras, 2009).

There are only few corpus-based studies which investigate agreement marking in L2 English in detail, and with the exception of Housen (2002), they all focus on adult learners. To my knowledge, there have not been any agreement studies conducted based on data from ICCI or CORYL yet. This chapter provides a brief overview of some published corpus-based studies of agreement in L2 English which could serve as a comparison to the Norwegian data discussed in chapter 5. The first section of the chapter focuses on speakers who use English as a lingua franca. The second section includes studies conducted on adult L2 learners and the last section discusses studies of young L2 learners. In addition to learner corpus studies, one spontaneous production study (Ionin & Wexler, 2002) and one grammaticality judgement study (Jensen et al., 2017) of agreement in L2 English are included in the third section.

3.1 English as a lingua franca

Breiteneder (2005) and Dröschel (2011) study spoken data collected from adult professionals who use English as lingua franca in multilingual environments. Both corpora have only limited size (50,000 and 167,086 words respectively) and include speakers from various L1 backgrounds. The L2 users included in these two studies are not necessarily advanced users of English, but it can be argued that they have reached their

ultimate attainment as they no longer receive any L2 instruction and use English as lingua franca in multilingual situations when no common L1 is available.

Breiteneder (2005) studies a spoken corpus of English compiled from working group discussions with representatives from national agencies of higher education within the EU. The corpus contains about 50,000 words and includes samples from speakers of 21 different languages (including five speakers of Danish, two speakers of Norwegian, and one speaker of Swedish). Breiteneder focused only on verbs other than *BE* and found that out of 141 instances of verbs in present tense which should mark 3rd person singular agreement the speakers omitted the marker in 29 cases. She also found 15 cases of 'superfluous *-s* marking' (Breiteneder, 2005, pp. 8-9). In other words, there are twice as many omission errors as overgeneralization errors in her data. Examples (1) and (2) illustrate the missing and superfluous agreement in the speech of a Portuguese and Spanish speaker, respectively (Breiteneder, 2005, pp. 9, 16):

(1)   I suppose it's possible **the thing** <u>function</u> in both possibilities.
(2)   Of course, **the students who apply for the master** <u>needs</u> to know this.

Breiteneder ascribes most of the recorded errors to language phenomena such as proximity and notional agreement. In addition, she stresses that since the 3rd person marker is communicatively redundant in English (Trudgill, 2002), it is understandable that the ELF speakers do not focus on this feature (Breiteneder, 2005, p. 5). However, she points out (following Widdowson, 1994) that such redundant grammatical features often take on other functions in the discourse, e.g. as markers of social identity or prestige (Breiteneder, 2005, p. 5).

Dröschel (2011) focuses on English as lingua franca in Switzerland. The corpus used in this study is the Swiss English Database which consists of transcribed recordings of various oral interactions and some emails exchanges among adult (20-59) non-native speakers of English (94 informants, L1s: Swiss German, French, and Italian). The corpus consists of 167,086 words. The study aims at characterizing Swiss English as a non-native variety of English in its own right claiming that high level of language contact in multilingual country like Switzerland creates a situation which resembles pidgin formation. Dröschel includes many aspects of morphosyntax in her study, but as Breiteneder, she focuses only on the present tense in her account of agreement (i.e. she does not describe the use of *BE* in the past). The number of present tense verbs in the corpus is approximated to be 15,666 (based on a sample of 400 sentences). However, this

approximation does not take into consideration modal verbs, which do not mark agreement overtly, so the amount of agreement marking occasions must be somewhat lower than that. Despite the high number of agreement marking occasions, Dröschel reports only 74 agreement errors: 44 omissions and 30 overgeneralizations (or hypercorrections, as she labels them). In other words, the Swiss speakers are very accurate in marking agreement in the present tense and if they make an error, they are more prone to omissions than to overgeneralizations. Examples (3) and (4) illustrate the omission and hypercorrection errors produced by speakers of French (Dröschel, 2011, p. 218):

(3)   **She** <u>bring</u> a strong leadership.
(4)   **Our customers** are actually English or <u>speaks</u> English.

Since speakers come from three different L1 communities, Dröschel dismisses possible L1 influence and attempts to explain the agreement inaccuracies as "a result of adult non-native learning" (Dröschel, 2011, p. 217). The omission errors are compared to the agreement patterns in New Englishes (typically colonial varieties of English, such as Philippine English or Indian English) where the absence of the 3rd person marker is a very common feature (Dröschel, 2011, p. 215).

The overgeneralization errors are described in more detail as these are not common in New Englishes. There are 20 errors which contain 3rd person plural subjects, nine errors which contain 1st person singular subjects, and one errors with 2nd person plural. However, eight of the nine overgeneralizations with 1st person singular subjects were produced by the same speaker, which suggests that this might be an idiolectic feature. Out of the 20 errors with 3rd person plural, five involve the noun *people* as a subject. As discussed in chapter 5.5.1, this noun is problematic for second language learners of English due to its lack of overt plural marking. Thus, when the eight potentially idiolectic errors and errors involving the noun *people* are removed, there are only 17 overgeneralization errors left in Dröschel's data as compared to 44 omission errors. Following Simo Bobda (1998), Dröschel (2011, pp. 215-216) further argues that the errors with 3rd person plural subjects could be caused by the lack of number agreement, that is to say that the speakers disregard the number distinction and assign overt agreement marking to all verbs with 3rd person subjects. However, this does not seem to be a prevalent strategy of the Swiss speakers of English since there are only 15 instances of 3rd person plural subjects (other than *people*) followed by 3rd person singular verbs

found in the corpus. The remaining two cases of overgeneralizations, namely one case of 1st person singular and one case of 2nd person plural, are dismissed as irrelevant by Dröschel. In conclusion, she returns to her hypothesis that English produced by Swiss speakers displays properties similar to New Englishes with regard to agreement, i.e. the use of omission as a simplification and regularization strategy. For the comparison with the data from the Norwegian learners, I suggest removing only the eight probably idiolectic errors from the dataset because all L2 learners tend to have problems with irregular plural nouns such as people and such errors are relevant for the discussion. Dröschel's data thus contain 44 omission errors and 22 overgeneralization errors.

Both Breiteneder (2005) and Dröschel (2011) report approximately twice as many omission errors than overgeneralization errors in their data. They assume mostly a sociolinguistic approach to their findings and explain the prevalence of omission errors by factors such as communicative redundancy and simplification processes typical for language contact situations. With such low error counts (and generally small corpora) it is difficult to make any claims about any potential patterns in the overgeneralization errors category and these errors might indeed seem idiosyncratic. Neither of the authors investigates suppletive agreement marking so it is not possible to evaluate (following White, 2003) whether their L2 users have mastered the usually early-learned suppletive agreement.

3.2   Adult L2 learners

Thagg Fisher's (1985) study of Swedish learners' agreement production in English is one of the first corpus studies which looked systematically into learner English. The study includes both subject-verb agreement errors and some smaller agreement categories, such as agreement between an anaphoric pronoun and its antecedent. The data consist of oral (retelling of a story and picture descriptions) and written production (translation and argumentative compositions) by Swedish university students collected in the 1960s. The oral data comprises about 15 hours of recorded speech (Thagg Fisher, 1985, p. 37). The written data includes 1052 translation tasks (120-300 words each) and 474 argumentative texts (300-580 words each) (Thagg Fisher, 1985, p. 67). Thagg Fisher does not specify the total number of words included in her corpus, but given the approximate numbers quoted above, the written corpus should comprise about 475,000 words. It is thus comparable in size to the corpus of Norwegian L2 English collected for my project. Even though this

study is a bit dated, it provides relevant comparison data from a student population with typologically closely related L1.

Thagg Fisher looks at the agreement errors from two perspectives. First, she considers the type of subject and verb elements involved in the agreement, i.e. the types of subject phrases (nominal vs. pronominal heads) and the types of verbs (full vs. primary, see chapter 5.7.1 for details). Then she turns to the complexity of the subject-verb construction, i.e. whether the subject and the verb are contiguous (Thagg Fisher, 1985, p. 9). These perspectives, with some modifications, are mirrored also in my analysis, as described in chapter 5. Unsurprisingly, she finds most errors in the spoken production (336), followed by the argumentative texts (267), and least in the translation tasks (118). This clearly reflects that the writers have more time to plan and review their production than the speakers, which contributes to higher accuracy. Thagg Fisher provides a detailed analysis of the agreement errors but not the frequencies of the various error types, which complicates the comparison with the Norwegian data. She concludes that many of the agreement errors produces by the Swedish L2 learners can be explained by three factors (alone or in combination):

1. Difficulties inherent in the English construction, e.g. collective nouns (example (5), p. 78), notional agreement (example (6), p. 81), proximity agreement (example (7), p. 94), and coordination within the subject (example (8), p. 90).

   (5) In 1961 **the population of the island** <u>were</u> 264 inhabitants.
   (6) I don't think **anybody** <u>disagree</u> with me.
   (7) **The dark lines of the coast road** <u>winds</u> across the light brown sand.
   (8) I think that **traveling abroad and the close contact with foreign people** <u>develops</u> one's character.

2. Contrastive problems, e.g. irregular plurals (example (9), p. 75), and mismatch of number or countability of specific nouns between the L1 and the L2 (example (10), p. 81).

(9)   **People** always <u>tries</u> to escape from the monotonous everyday rhythm.

(10)   They form a little society and **everybody** <u>work</u> together.[5]

3. Performance problems, e.g. using the unmarked form as a simplification strategy (example (11), p. 74), '-*s* preservation' (example (12), p. 85), overgeneralization of the 'one -*s* principle' (example (13), p. 188), etc.

(11)   **Christmas** <u>make</u> people feel terribly sad and guilty too.

(12)   **The adventures** <u>begins</u> at once.

(13)   **Two small children** <u>appears</u>.[6]

The errors from the first category are quite common also among native speakers of English, especially in unplanned spoken production. However, Quirk et al. (1985, p. 757) point out that such errors are rarely left uncorrected in written native speaker production. Errors from the second and third categories are specific to L2 English. With the exception of L1 influence (i.e. the mismatch of number or countability), these errors should thus be present in the L2 English of speakers of various L1s. Errors with irregular plurals (specifically with the noun *people*) are indeed reported by Dröschel (2011). However, all the other examples quoted by Dröschel fall into the first category of difficulties inherent in the English construction, such as coordination within subject and proximity agreement (see Dröschel, 2011, pp. 214-218). None of the examples given by Dröschel can be labeled as -*s* preservation.

Johansson (2008) uses the Norwegian part (NICLE) of the ICLE corpus (Granger et al., 2009) for his analysis of learner language. The corpus contains argumentative texts written by Norwegian university students. 317 students contributed to the collection which comprises 211,725 words. Johansson provides a broad qualitative description of learner English produced by Norwegian speakers, but there are no quanatitative data. In his analysis, he recognizes two main categories of non-standrad agreement: errors triggered by the distance between the head of the subject noun phrase and the finite verb, as in (14), and errors triggered by the semantics of the subject, as in (15) (Johansson, 2008, pp. 139-140):

---

[5]   *Everybody* in Swedish (*alla*) and in Norwegian (*alle*) are syntactically plural.
[6]   Overgeneralization of the 'one -*s* principle' concerns mainly irregular plurals and could thus also be categorized under point number 2.

(14)   The chain continues with **effects on the weather** which <u>becomes</u> more marked.

(15)   In Norway **the police** <u>investigates</u> all reported criminal actions.

Johansson does not mention any other error types, but comments generally that "some errors appear to be mere slips which could easily by corrected by the learner" (2008, p. 139). However, it is questionable whether these errors are in fact slips when the data are written, untimed essays with the option of planning, proof-reading, and using reference tools such as dictionaries. In addition, agreement errors seem to be problematic for Norwegian learners also in grammaticality judgment tests (Jensen et al., 2017, see also the discussion in chapter 6).

Neff et al. (2007) report the results of an error tagging project performed on a Spanish subsection (SPICLE) of the ICLE corpus. This study does not focus on agreement in particular, but the authors comment on a surprisingly high number of illicit agreement marking. Out of the detected agreement errors, the Spanish EFL learners omitted the 3rd person singular marker in obligatory contexts 78% of the time while they used it in illicitly 22% of the time (28 occurrences). Most of the overgeneralization errors are reported to occur in relative clauses, typically with additional intervening elements. Even though an overgeneralization rate of 22% is lower than the rates reported by Breiteneder (2005) and Dröschel (2011), the authors comment on it as unusual and call for further analysis.

Of the three studies reviewed above, Thagg Fisher's (1985) data provide the best match for the Norwegian data analyzed in chapter 5. Even though the Swedish students were already adults when the data were collected, English in the 1960s was not as heavily present in the Scandinavian context as it is now and English instruction in school started significantly later than now (see chapter 9.2). Her subjects were thus exposed to English instruction for a shorter time than Johansson's subjects which could have resulted in lower proficiency in English. In addition, neither Johansson (2008) nor Neff et al. (2007) provide enough details about the errors they identified in their data to allow for reanalysis. I believe, therefore, that from these three studies, only Thagg Fisher's findings provide a comparable sample to the data from Norwegian high school students. I return to this comparison in chapter 5.7 where I reanalyze Thagg Fisher's data and compare them to the data from the corpus of young Norwegian learners.

3.3   Young L2 learners

Housen (2002) explores a cross-sectional (quasi-longitudinal) spoken corpus of 46 young Dutch- and French-speaking learners of English (CYLIL corpus). He examines the development of the verb system in English with specific focus on progressive aspect, present tense, and past tense. The corpus consists of 230,000 words and includes data form 46 learners aged nine to 17 (3rd to 11th grade). However, these learners are divided only according to their proficiency scores and not according to their age in the results, which makes it difficult to select a group which would be comparable to the students who contributed to the corpus described in my project. I assume that Housen's higher intermediate and high proficiency groups should correspond to the level of the Norwegian high school students. In addition, Housen's corpus is spoken, which suggests less controled production and higher potential for performance errors than in the written corpora discussed above.

The use of 3rd person singular marker (*V-s* in Housen's terminology) is said to be "either infrequent or erratic" in the low and lower intermediate groups (Housen, 2002, p. 95). Housen gives examples of both underuse and overuse produced by speakers of French. These are reproduced below as examples (16) to (19) (Housen, 2002, pp. 91, 95). He also points out that many of the marked (*V-s*) forms stem from only few informants from the lower intermediate group. However, he does not specify whether this concerns only the overused forms or also correctly marked verbs:

(16)   **mister Neil** <u>have</u> give me the \<la lettre\>

(17)   and **he** <u>go</u> to her grandmother

(18)   there's **some children** who <u>goes</u> to missus Bristow

(19)   **we** <u>speaks</u> french

Housen describes also other verbal inflections, e.g. the appearance of progressive and past tense forms, and concludes that at the early stages of development, the respective morphs (e.g. *V-ϕ* and *V-s*) "behave like allomorphs as they appear first in random variation and then in complementary distribution" (Housen, 2002, p. 97). Specifically for the 3rd person marker, he reports both overextension and underextension at the low and lower intermediate levels but an abrupt change in the accuracy levels between the higher intermediate and high proficiency groups (see *Figure 4* below which is reproduced from Housen's figures 1a and 1b, p. 99).
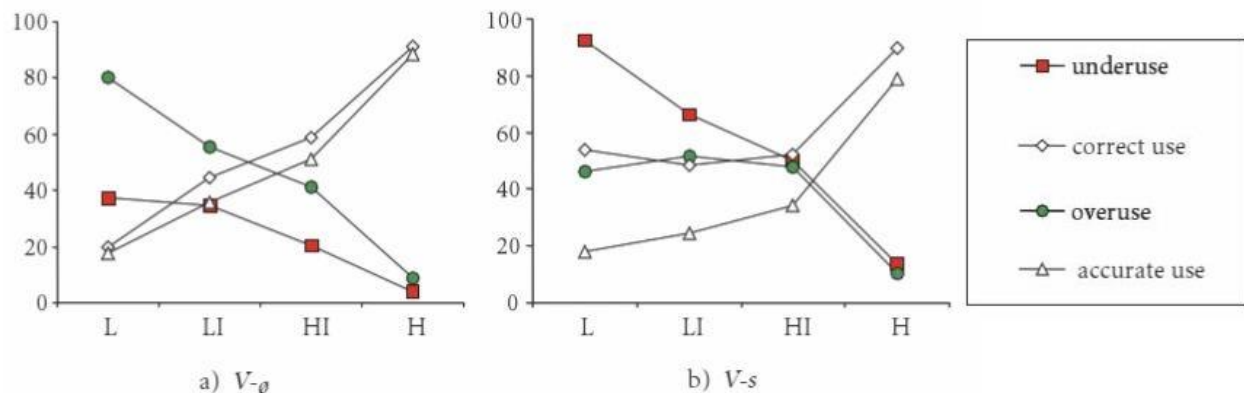
Figure 4. Patterns of over-, under-, correct, and accurate use (in percentage). Reproduced from Housen (2002).

It is difficult to interpret Housen's results regarding agreement marking accuracy because he treats the omission errors and overgeneralization errors separately with overuse and underuse rates in each category. The overuse of the unmarked form and the underuse of the marked form should refer to the same phenomenon, i.e. missing -*s*, but it is not clear from the graphs that this is the case. However, it is evident from *Figure 4* that the accuracy of use for the unmarked form (*V-ø*) increases steadily through all the developmental stages and reaches approximately 90% in the high proficiency group. In the early stages of development, the students overuse the unmarked form, i.e. produce omission errors, but become more and more accurate as their general proficiency grows. The omission errors are in fact the dominant error type in all proficiency groups. In other words, the students produce unmarked forms (*V-ø*) where native speakers would produce marked forms (*V-s*).

On the other hand, the accuracy of the marked form (*V-s*) increases only slightly between the low and higher intermediate stages before a radical increase in accuracy from higher intermediate to high proficiency group when the students reach approximately 80% accuracy. The students mainly underuse the marked form when they produce incorrect forms, i.e. omission errors, until the higher intermediate stage. In other words, they produce unmarked forms (*V-ø*) where native speakers would use marked forms (*V-s*). It seems that at the higher intermediate stage the under- and overuse of the marked forms are approximately on the same level, i.e. as if the students used them interchangeably in the obligatory contexts. However, Housen does not provide any details or examples from

this developmental stage and it is not clear how many errors of each type there are in the material.

All in all, it seems that only the higher intermediate students in Housen's corpus might behave similarly to the Norwegian students described in in this thesis with respect to inaccurate and variable under- and overuse of the 3rd person singular marker. Housen does not address this observation of agreement marking variability specifically for each group beyond what has been quoted above. It is thus difficult to evaluate whether the proficiency of the higher intermediate students is comparable to the Norwegian students. In addition, Housen's data are from oral interactions while the data form the Norwegian students are written. Written language offers the learners more control of the production, including the possibility of planning and revisions. It is thus difficult to compare the error rates between these two production modes directly.

Ionin & Wexler (2002) examine the spoken production of 20 young Russian learners of English (age 3;9-13;10) with short exposure to the L2 (1-3 years). Their study is not corpus based, but the transcribed data of the learners can be seen as a small learner corpus of this particular population. The learners participating in Ionin & Wexler's study are on a lower proficiency level than the Norwegian students examined in my project, but Ionin & Wexler (2002) is to my knowledge the only study of agreement in L2 English which includes also suppletive forms.

Ionin & Wexler report omissions of the agreement marker on lexical verbs as the most common error among their participants: 18 of the 20 children omit the agreement marker in 40-100% of the obligatory contexts (Ionin & Wexler, 2002, p. 108). There are only four instances of inappropriate 3rd person -s use in their material compared to 250 instances of omissions of the marker (78% of all obligatory contexts). This level of accuracy is comparable to Housen's (2002) low proficiency group. However, Ionin & Wexler do not provide the number of verbs in the present tense, only the number of obligatory contexts for the agreement marker, i.e. the number of verbs following 3rd person singular subjects. It is thus not possible to calculate the overgeneralization error rate for this learner population beyond the fact that it is low.

The young L2 learners are more accurate with suppletive forms, i.e. the forms of the verb *BE*. Suppletive agreement is omitted in only about 25% of the obligatory contexts (calculated based on the data in Table 1 in Ionin & Wexler (2002, p. 106)). Even though the inappropriate use of the suppletive agreement is nominally higher than the

31

inappropriate affixal agreement (54 cases of incorrect *BE* forms), due to the higher frequency of *BE* as compared to lexical verbs in the material, the overall rate of suppletive agreement errors is low: only 6% of the instances of *BE* are incorrect (calculated based on the data in Tables 1 and 2 in Ionin & Wexler (2002, pp. 106-107)). This confirms White's (2003) claim that agreement markers, when supplied, are mostly correct. Ionin & Wexler explain the differences in agreement marking accuracy between the affixal and suppletive forms by verb raising patterns in English. They argue that learners might "initially associate morphological agreement with overt movement to Tense" (Ionin & Wexler, 2002, p. 116). Since lexical verbs never raise in English, they need to be marked for agreement by affix lowering, which is a less economical operation and thus acquired later (White, 1992).

Jensen, Slabakova & Westergård (2017), based on Jensen (2016), report an untimed grammaticality judgement study conducted with young Norwegian learners of English (60 learners, 12-19-year-olds). Prior to the experiment, the participants' proficiency in English was tested and they were divided into four groups from low intermediate to advanced. They were then asked to evaluate the grammaticality of the test sentences which focused on subject-verb agreement and incorrect V2 word order in English. Jensen et al. (2017) show that Norwegian learners have more problems with functional morphology (agreement) than with purely syntactic operations (word order). The significance of this study lies in the fact that it does not examine production but rather the internal competence of the learners. Jensen et al.'s (2017) findings show that the agreement issues reported from the corpus material are unlikely to be caused by performance alone. The overall acceptability rate for ungrammatical agreement in their study is 0.68. If only the advanced group was considered, the acceptability rate is 0.47 (Jensen, 2016, p. 65). In other words, even the advanced Norwegian learners accepted almost half of the sentences with ungrammatical agreement.

Jensen et al. (2017) further test the differences between local and long-distance agreement marking and between omission and overgeneralization errors. Based on Ocampo (2013) they predict that long-distance agreement should render more incorrect judgements, especially when the subject is singular and the intervening noun is plural (20). This prediction was not fully borne out as the learners accept more ungrammatical sentences with long-distance agreement with plural subjects and singular intervening

nouns, as in (21), than in any of the other conditions (Jensen et al., 2017, p. 342, examples p. 338):

(20)   * The teacher with black shoes walk to work every day.

(21)   * The kids with the red bike plays in the garden.

In testing omission vs. overgeneralization errors in local agreement settings they predict, based on Dröschel (2011), that the learners should be more prone to accept omission errors (i.e. singular subject with unmarked verb) than overgeneralization errors (i.e. plural subject with superfluously marked verb). This prediction is not borne out. The learners reject grammatical sentences with plural subjects (22) significantly more often than grammatical sentences with singular subjects, as in (23). They also accept ungrammatical sentences with plural subjects (24) significantly more often than ungrammatical sentences with singular subjects, as in (25) (Jensen et al., 2017, p. 342, examples p. 338):

(22)   The teachers give their students a lot of homework.

(23)   The brown dog plays with the yellow football.

(24)   *The teachers gives their students a lot of homework.

(25)   *The brown dog play with the yellow football.

Based on the grammaticality judgement test reported by Jensen et al. (2017) it seems that the young Norwegian learners prefer the marked 3rd person singular verb form over the unmarked plural form. Ungrammatical sentences with plural subjects and illicitly marked verbs (example (24)) are accepted more often than ungrammatical sentences with singular subjects and unmarked verbs (example (25)). The students also accept erroneous long-distance agreement with plural subjects and singular intervening nouns (example (21)) more often than sentences with singular subjects and plural intervening nouns (example (20)). In addition, they reject grammatical sentences with plural subjects and unmarked verbs (example (22)) more often than grammatical sentences with singular subjects and marked verbs (example (23)) (Jensen, 2016, p. 74). Jensen et al. (2017) did not test suppletive agreement.

The three studies conducted with young learners of English summarized above test several predictions about the acquisition is agreement in L2 English which are relevant for the discussion of the Norwegian data:

1. Learners of English are expected to overuse the unmarked form of the verb, i.e. produce omission errors (Housen, 2002; Ionin & Wexler, 2002; Jensen et al., 2017).
2. Learners of English are expected to be more accurate with suppletive agreement than with affixal agreement (Ionin & Wexler, 2002).
3. Learners of English are expected to be more accurate with local agreement than with long-distance agreement (Jensen et al., 2017).

Predictions 2 and 3 are both confirmed in the respective studies, but they were only tested in one study each and it is not possible to compare the results of various L2 groups. Prediction 1 was tested by all three studies and there are clear differences in the results. Young Dutch, French, and Russian learners of English produce more omission errors than overgeneralization errors when they fail to mark agreement correctly, as predicted. Young Norwegian learners were not tested in production, but they accept ungrammatical sentences with superfluously marked verbs more often than ungrammatical sentences lacking the obligatory agreement marking, which goes against prediction 1.

The difference in the results among the L1 groups suggests that the overgeneralization errors of the young Norwegian learners may be caused by L1 influence. On the other hand, there could also be other factors at play, such as the age and proficiency level of the participants. Ionin & Wexler's (2002) participants were significantly younger and less proficient than the Norwegian learners reported on by Jensen et al. (2017). However, it is difficult to imagine that older and more proficient learners should suddenly start overproducing the marked verb forms. The younger and less proficient groups in Housen (2002) produced both omission and overgeneralization errors but the superfluous agreement marking dropped sharply between the high intermediate and high proficiency groups. The young Norwegian learners also improved their accuracy between the high intermediate and advanced groups, but the advanced learners still approached chance judgement of erroneous agreement (46% of ungrammatical sentences were accepted). The type of material collected could also play a role, but in that case an untimed grammaticality judgement test should yield higher accuracy than spontaneous oral production, which was not the case.

4 Data collection

Even though this project is divided into two parts with different aims, the data collection procedure was the same for both parts of the project. The first phase of the project is exploratory and aims to describe and explain the patterns of subject-verb agreement errors in L2 English of Norwegian high school students. The second phase of the project focuses on the development and testing of an English grammar course using the Inverted Classroom Method. The data collected for the first phase of the project are used to identify the areas of English grammar which are problematic for the students and which need to be covered in the teaching materials for the course. The error counts of the students from the first phase also serve as a standard of comparison for the testing of the course in the second phase of the project. This chapter describes the procedures connected to data collection, subject selection, subject background information, corpus compilation and screening for errors for further analysis. The methodology related to the intervention, including a more detailed description of the subject backgrounds, is discussed in chapter 10.

4.1 Subjects

The subjects for the exploratory study were recruited via their English teachers. The subjects attended three different schools in the Agder region of Norway, and the group included students from several study programs offered at these schools. Only first-year students were recruited because English is only compulsory in the first grade of high school across all study programs (vocational study programs divide the compulsory instruction into two years). Since English is taught in Norway from the first grade, all students included in this study have already received at least 10 years of English instruction prior to data collection.

In total, there were 10 teachers involved in teaching 199 students in 17 classes: 6 general studies groups and 11 vocational studies groups. However, as the vocational groups tend to be smaller, the corpus includes texts of 123 general studies students and only 76 vocational students. Among the subjects, there are 113 females, 95 males, and 2 with unassigned gender (these are the cases where the student did not fill in gender information).

Each student involved in the collection was asked to fill in a background questionnaire. The goal of this questionnaire is to control for some factors which can

influence the performance in English (e.g. bilingualism, learning difficulties, etc.) as well as explore the attitudes of the students towards foreign languages in general and formal correctness in the foreign language in particular. All questionnaires are given in Norwegian to ensure that the students fully understand the questions. A sample questionnaire is included as Appendix 1.

For the purposes of the current analysis, only subjects who are native or near-native users of Norwegian were selected. A subject is considered a native or near-native speaker of Norwegian (regardless of the home languages) if he or she has attended school in Norway from the first grade. A subject is considered bi-/multilingual if he or she specified a home language different from Norwegian. There are 14 bilingual subjects in the dataset. There are also 14 subjects in the dataset who indicated a learning difficulty in their background questionnaire (two of these are bilingual). The learning difficulties are not specified in detail. These subjects are excluded from the exploratory analysis but included in the intervention.

4.2 Texts

Each recruited student contributed on average with 2265 words (median value) to the corpus, which gives a total of over 430 000 words for the explorative analysis. The majority of the texts were written exams produced at school. The only allowed aid in such a setting is a bilingual dictionary. Some texts were produced at home and delivered as graded assignments. The aids used in such cases are unclear and cannot be controlled for. The majority of the texts were graded by the teacher; thus the students should have an incentive to do their best. However, the grades were not collected for the current study because this analysis focuses only on one specific linguistic feature of the texts, agreement accuracy, while there are many linguistic and non-linguistic factors which can influence a grade (see also chapter 10.5 for the discussion of alternative methodology options).

4.3 Corpus compilation and error analysis

All the collected texts together with the metadata from the background questionnaires are to be included in an extended version of the CORYL corpus (Hasselgreen & Sundet, 2017) and made available for further research. For the purposes of this exploratory study, only the students who were considered native or near-native speakers of Norwegian were

selected. Based on an unpublished pilot from 2013-2014 (Garshol, 2014), subject-verb agreement errors were selected as the focus of the investigation because these errors occur in the texts of the majority of the subjects throughout the school year.

Errors in learner data can generally be divided between errors caused by a lack of competence, i.e. the L2 interlanguage differs from the standard version of the target language, and mistakes or slips, which are considered to be a performance problem. The students who contributed to this study have generally high level of proficiency in English and produce target-like agreement most of the time (over 90% correct agreement if all students are considered together). However, the frequency of agreement errors does not change during the one-year collection period (see chapters 11.2 and 11.3) and Norwegian students tend to accept non-standard agreement also in grammaticality judgement tests (Jensen et al., 2017). Therefore, it does not seem that the non-standard agreement in the texts of Norwegian high school students can be labeled as an occasional mistake or a performance problem. I have thus chosen to label all cases of non-standard agreement as errors with the exception of the cases discussed in chapter 5.1.

4.3.1   Error detection and categorization

The corpus is not error tagged, so the texts were manually screened for subject-verb agreement errors, and the relevant instances were extracted. Each error was recorded separately as a single line in the extracted material, i.e. one sentence can be represented several times if it contains several agreement errors. The errors in the resulting concordance list were commented on if necessary, e.g. whether the teacher corrected the error or not, whether the error could be a typographical one, whether it could be accepted in some dialects, etc. All and only clear instances of a lack of knowledge or attention resulting in a subject-verb agreement error were taken into the extracted material.

The complete list of the extracted clauses was further divided into categories according to the type of subject, the presence of elements between the subject and the verb, and the type and tense of the verb. Combination of these factors were analyzed and compared in order to isolate the most problematic combinations, i.e. those combinations which most often lead to errors. This approach aims to disclose any recurring patterns in the L2 English of the subjects which differs from the syntax of Standard English. The main reason for choosing a strictly quantitative approach is that the corpus material can easily be misleading if one does not take into consideration the frequency of the observed

element in the total volume of the sample. An element which might be interesting from the descriptive point of view is irrelevant for generalizations if it is infrequent (McEnery & Wilson, 1996, pp. 62-63). This approach makes the project positivist in nature as contrasted with the interpretivist approaches which often focus on anomalies or atypical cases (Della Porta & Keating, 2008, p. 29).

4.3.2 Hypotheses testing

The results of my pilot project show that young Norwegian learners produce many subject-verb agreement errors despite their otherwise high proficiency in English. In addition, the errors are not typical of L2 English. L2 English learners typically underuse the 3rd person singular -*s*, while the Norwegian learners tend to overuse the 3rd person singular -*s* in all other persons. The goal of the first part of the project is to investigate which combinations of syntactic factors most often result in agreement errors in Norwegian L2 English and whether the error patterns are distinct from the errors typically found in L2 English of students of other language backgrounds.

As discussed in chapter 2.4 above, a comparison of groups of speakers of different L1s learning the same L2 can help discover L1 influence in SLA. A better understanding of this influence can inform both theories of L2 acquisition and didactic interventions which can be made designed to be L1 specific based on the common problems learners of that L1 encounter. The data from this study are therefore compared to data from German-speaking learners of English extracted from ICCI (see chapter 5.8) in order to see whether the typical errors found in the Norwegian learners' production are also found in the production of German-speaking learners. In case the dominant error types from the texts written by Norwegian learners are not as prominent in the texts of other learners, this would suggest that the errors can be caused by a negative transfer from the first language. In case the errors are similar in the texts of learners of other language backgrounds, transfer from L1 would be less likely and the high incidence of these errors in the texts of Norwegian upper-intermediate and advanced learners would have to be explained by some other factors.

5 Corpus data

As already mentioned, the data were collected in three schools in the Agder region in Norway both from general studies students (henceforth GS) and vocational students (henceforth VS). The texts in the corpus are of varying length, genre, and topic. The text types include short reflections on current socio-political topics, historical events, or texts read in class, as well as longer essays and written exams. The accuracy of agreement marking varies both across the students and between the different text types and topics. Detailed description of the error scores calcuation and the accuracy development during the collection period is provided in chapter 11.

The first two sections of this chapter introduce the error detection and inclusion in more detail as well as provide an overview of the error distribution among the contributing subjects. The third section is devoted to the analysis of the different error types based the finite verb used in the erroneous clause and the fourth section extends the analysis to the different subject types within each verb category. The fifth section dicusses possible sources for some of the errors before a summary of the Norwegian data in the sixth section. Finally, the last two sections compare the data from the Norwegian students to data from Swedish and Austrian students of English, which then leads into the discussion of the results in chapter 6.

5.1 Error types and criteria for inclusion in the data set

There were in total 3,132 subject-verb agreement errors detected in the corpus. After all questionable cases were removed, there were 2,941 clear agreement errors left in the data set. Examples of excluded questionable cases of errors, i.e. cases where there is a reasonable amount of doubt whether the construction is an agreement error or an acceptable variation in a standard dialect of English, or where the apparent error might be caused by other factors, such as typographical errors, are listed in (26) to (31):

(26)   Number-ambiguous nouns:
       …the government <u>were</u> racist. (15HOK08_May16)[7]

(27)   Indefinite pronouns and quantifiers which Biber et al. (1999, p. 184) classify as combining with both plural and singular verb forms:
       …none of them <u>are</u> positive. (15STV22_Apr16)

(28)   Percentages and similar quantifiers with unclear referents (in the example below people/inhabitants or population) (Biber et al., 1999, p. 185):
       75% <u>has</u> not left their country. (13STK04_Apr14)

(29)   Apparent error in the number of the subject:
       The word <u>were</u> her weapon. (14STA19_Nov14)

(30)   Uninflected verb form in a past context:
       …he laid down and <u>shoot</u> the other ones. (13STK03_Nov13)

(31)   Verb with an *-s* morph in a past context (the letters S and D are next to each other on a QWERTY keyboard):
       The British came with many European diseases and as in America, **many** <u>dies</u>. (15HOK02_May16)

All errors included in the data set were classified according to the type of verb (affixal errors on lexical verbs and auxiliaries *have* and *do,* or suppletive errors on auxiliary and copula *BE*) and the type of subject (noun phrases, personal pronouns, indefinite pronouns, demonstrative pronouns, expletive pronouns, clausal subjects). The erroneous clauses were further coded for the following: the number of the verb, whether the subject phrase is coordinated, post-modified, or whether there are any other intervening elements between the subject and the finite verb. Clauses with noun phrase subjects were also coded for the regularity of plural formation (irregular plurals, non-count nouns, nouns that only occur in singular/plural, etc. were all labeled as irregular).

Affixal agreement errors, i.e. errors with lexical verbs and non-modal auxiliaries except *BE*, were divided into two main categories based on the number of the verb. Errors which involve a 3rd person singular subject and a plural unmarked verb were labeled *omission errors*. Errors which involve a non-3rd person singular subject and a singular

---

[7]   Each erroneous sentence extracted from the corpus is marked with the school year of collection (in this case starting in the fall of 2015), abbreviation for the type of class (HO (helse og oppvekst) – Healthcare, Childhood and Youth Development), code for the school it was collected from (K), random number assigned to each student (08), and the month and year when the text was written (May16).

verb marked with the 3rd person morph -*s* were labeled *overgeneralization errors*. Overgeneralization errors are also labeled as superfluous agreement marking, overuse of the agreement marker, or hypercorrection in the literature (see chapter 3). I have chosen to refer to this type of error as an overgeneralization of the overt agreement marking in contexts where agreement is not overtly marked in Standard English. Errors which involve the verb *BE* were treated separately as *BE* marks agreement with suppletive morphs and it is the only verb in English which marks agreement also in the past tense.

5.2 Data distribution among the contributing subjects

The English curriculum for the vocational studies is divided into two school years, which means that the VS students included in this study only received half of the instruction of the GS students. As a result, they also had fewer written tests and assignments during the year, and they have higher overall error scores in their texts than the GS students. The corpus contains 433,844 words where the GS students contributed 370,903 words (85.49%) and the VS students contributed 62,941 words (14.51%). The GS students wrote on average 2,871 words each during the school year (median value) while the VS students wrote on average only 613.5 words each during the school year. There were altogether 2,941 subject-verb agreement errors detected in the corpus: 2,371 (80.62%) in the texts of the GS students and 570 (19.38%) in the texts of the VS students. The VS students thus stand for a higher proportion of errors than their word contribution would suggest, which can be interpreted as lower accuracy in subject-verb agreement marking in this student group than in the GS group. The median error score for the GS group is 7.77%, while the median error score for the VS groups is 9.90%. The median error score for the entire student population is 8.02%.[8] Table 5 gives an overview of the distribution of errors in three categories based on the overt agreement marking on the verb for these two groups:

Table 5

Overview of the proportion of incorrect agreement marking with different types of verbs by the type of study program the students follow

	Omission errors	Overgeneralization errors	Suppletive errors
GS students	726 (79.52%)	811 (82.50%)	834 (79.81%)
VS students	187 (20.48%)	172 (17.50%)	211 (20.19%)
Total	913 (100%)	983 (100%)	1045 (100%)

---

[8] Error scores are calculated as the number of incorrect agreement marking instances divided by the amount of potential occasions to mark agreement in each text. See also chapter 11.

Despite the nominal differences in the error counts, there are no significant differences in the distribution of the error types between these two groups, neither in the proportion of omission and overgeneralization errors, nor in the proportion of the errors with *BE* and errors with other verbs (chi-square test (Field, Miles, & Field, 2012, pp. 812-828) for omission vs. overgeneralization, $\chi^2(1) = 2.75$, $p = 0.097$; for suppletive vs. affixal, $\chi^2(1) = 0.68$, $p = 0.409$). Since the distribution of the error types is similar in these two groups, all students are in the following considered as one group regardless their study program.

5.3 Verb types

Agreement marking in English can occur with finite lexical and non-modal auxiliary verbs in the present tense as well as with the copula and auxiliary *BE* in both present and past. As there is no difference in the morphology of agreement for lexical verbs and non-modal auxiliaries except for *BE*, in the following, the only distinction is made between the verb *BE* and other verbs (other verbs thus include all verbs that mark agreement with an overt suffix: lexical verbs and the auxiliaries *have* and *do*). The verb *BE* is treated separately for two reasons. First, it requires agreement in several contexts, not only in the third person singular. Second, the agreement paradigm contains several suppletive morphs which are distinct from the affixal 3rd person singular *-s* morph which occurs with other verbs.

Based on previous studies, it is predicted that the students should be more accurate with suppletive agreement than with affixal agreement (Ionin & Wexler, 2002; White, 2003). Of the 2,941 errors identified in the data, 1,896 (64.47%) are in clauses with other verbs, and 1,045 (35.53%) are in clauses with *BE* as the finite verb (22.58% in the present and 12.95% in the past) as seen in the chart in *Figure 5*:

Total number of errors
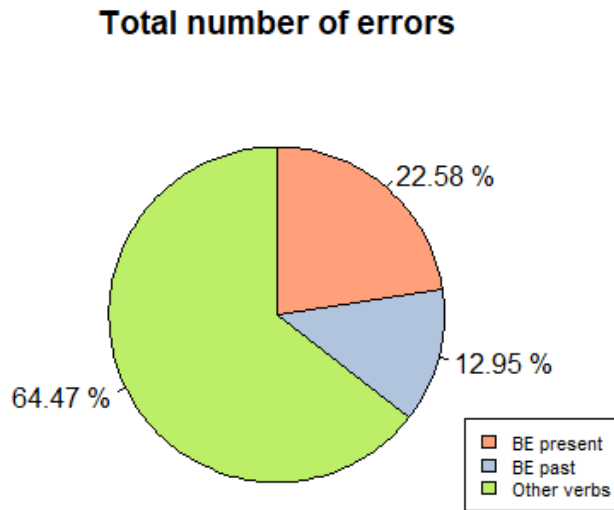
Figure 5. Proportion of errors in each verbal category.

Despite the more complicated agreement paradigm of the verb *BE* as compared to other verbs (and the increased number of contexts to mark agreement due to the marking in the past tense), errors with *BE* are clearly in the minority. This can be due to earlier acquisition of suppletive agreement as compared to affixal agreement, as suggested by among others Ionin & Wexler (2002). However, a contributing factor could also be the high overall frequency of *BE* in English (the second most frequent word in the Corpus of Contemporary American English). The students' input should also contain many examples of different forms of *BE*. In the following sections, I first discuss the errors which occur with other verbs, i.e. the affixal agreement errors, and then I turn to the errors involving *BE*, i.e. the suppletive agreement errors.

5.3.1 Affixal agreement errors

The discussion in the chapter 2.3.1.3 indicates that the most frequent problem L2 learners of English have with the subject-verb agreement is the omission of the obligatory marking of the 3rd person singular as illustrated in (32) to (34):

(32)   …**he** <u>go</u>… (Thagg Fisher, 1985, p. 189)

(33)   …**he** <u>get</u> the diploma. (Breiteneder, 2005, p. 9)

(34)   **Dancing school** <u>start</u> late in the evening. (ICCI, subject 41121)

Agreement errors in which the 3rd person singular marker is supplied in contexts where it is not allowed (overgeneralization) are rarely mentioned in the literature discussing L2 English. If such errors are mentioned, it is pointed out that these are in the minority as compared to the omission errors (Breiteneder, 2005; Dröschel, 2011). Nevertheless, Norwegian students produce as many overgeneralization errors as omission errors. Out of the 1896 errors in clauses with verbs other than *BE,* 983 (51.85%) are overgeneralization errors, as in (35) and (36), and 913 (48,15%) are omission errors, as in (37) and (38):

(35)   **All students** <u>starts</u> at the same level. (13STK02_Feb14)

(36)   **I** <u>does</u> this all the time. (15STV48_Dec15)

(37)   **This quote** <u>stand</u> for the American Dream. (15STV20_Feb16)

(38)   **He** <u>try</u> to play with his daughter. (15STV47_Dec15)

However, it is important to note that the proportions of omission and overgeneralization errors differ depending on the type of subject used in the clauses, especially between clauses with pronominal and NP subjects (see below in section 5.4).

5.3.2   Suppletive agreement errors

Agreement errors in clauses with the verb *BE* are not usually treated separately in the literature, but there are good reasons to do so. The labeling as omission or overgeneralization errors cannot be used for errors with *BE* because the agreement morphology is suppletive. In addition, *BE* has two different agreement patterns – one in the present tense and one in the past tense. There are more errors detected in the present tense than in the past tense, but this is probably due to the higher frequency of present tense in the corpus as such (there are 12,538 instances of *BE* in the present tense in the corpus as compared to 5,054 instances of *BE* in the past tense). There are no instances of erroneous use of *am* in the corpus. The proportions of the errors with *is*, *are*, *was*, and *were* are exemplified in *Figure 6*:
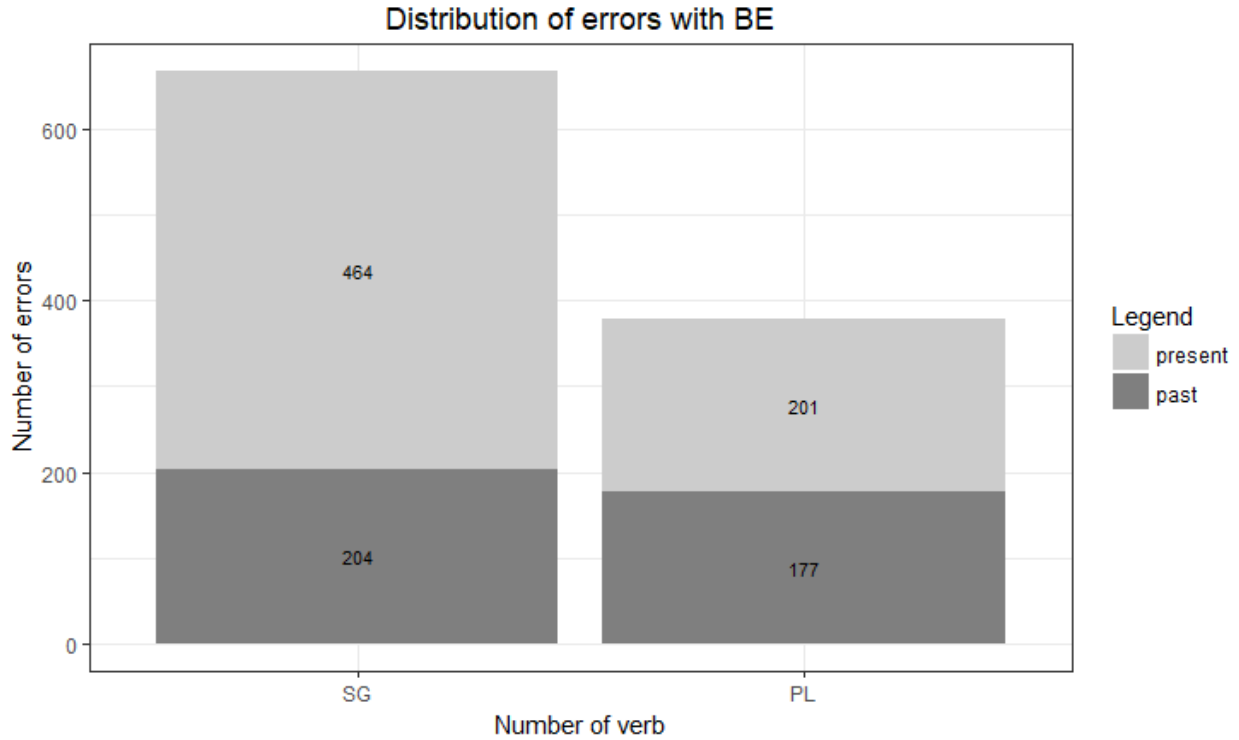
Figure 6. Distribution of errors in clauses with BE.

There is also a contrastive aspect to consider related to the use of different forms of the verb *BE* by Norwegian users of English. Norwegian and English are relatively closely related, and some forms of the *BE* paradigm are phonologically similar in these two languages as illustrated in Table 6:

Table 6

Overview of some forms of the BE paradigm in English and Norwegian with phonological transcriptions

Present	English		Norwegian	
Singular	is	/ɪz/	er	/ær/
Plural	are	/ɑr, ər/		
Past	English		Norwegian	
Singular	was	/wʌz, wɒz, wəz/	var	/var/
Plural	were	/wɜr, wər/		

Norwegian does not mark number or person on verbs, so *er* and *var* have the same form in both singular and plural for all persons. The pronunciation of the present tense *er*

45

differs in different dialects of Norwegian (the transcription above is the standard version of Bokmål). However, the speakers of Norwegian dialects whose pronunciation is close to Bokmål might perceive the Norwegian *er* as sounding similar to the English *are*. The pronunciation of *var* is very similar in most Norwegian dialects and since Norwegian users of English often have problems distinguishing between /v/ and /w/ in English (Graedler, 2002), it is conceivable that they may perceive the Norwegian *var* as sounding similar to the English *were*. Based on these similarities, the Norwegian L2 learners would be expected to overuse *are* and *were* in their English when they make errors in suppletive agreement with the verb *BE*. This is not confirmed by the data as illustrated by the graph in *Figure 6* above.

It is clear that the most erroneously used form of *BE* is *is*. 44.36% of all errors in clauses with *BE* are erroneous uses of *is*. However, when the overall frequencies of the different forms of *BE* in the corpus are compared to the erroneous uses of these forms (*Figure 7* below), a pattern confirming the prediction based on phonological similarity emerges. *Is* is used very frequently in the corpus (9,037 instances), followed by *was* (3,885 instances), *are* (3,102 instances), and *were* (1,160 instances). So, Norwegian learners of English use *were* erroneously in 15.26% of the recorded instances of *were* in the corpus, while they use *was* erroneously only in 5.25% of the instances. The pattern is not as clear for *are* (error rate of 6.48%) as compared to *is* (error rate of 5.13%), which could be related to the dialectal variation in pronunciation mentioned above or the high frequency of these forms in the language. However, both differences are significant (chi-square test (Field et al., 2012, pp. 812-828) for *is* vs. *are*, $\chi^2(1) = 8.07$, $p = 0.004$; for *was* vs *were*, $\chi^2(1) = 128.14$, $p < 0.001$):
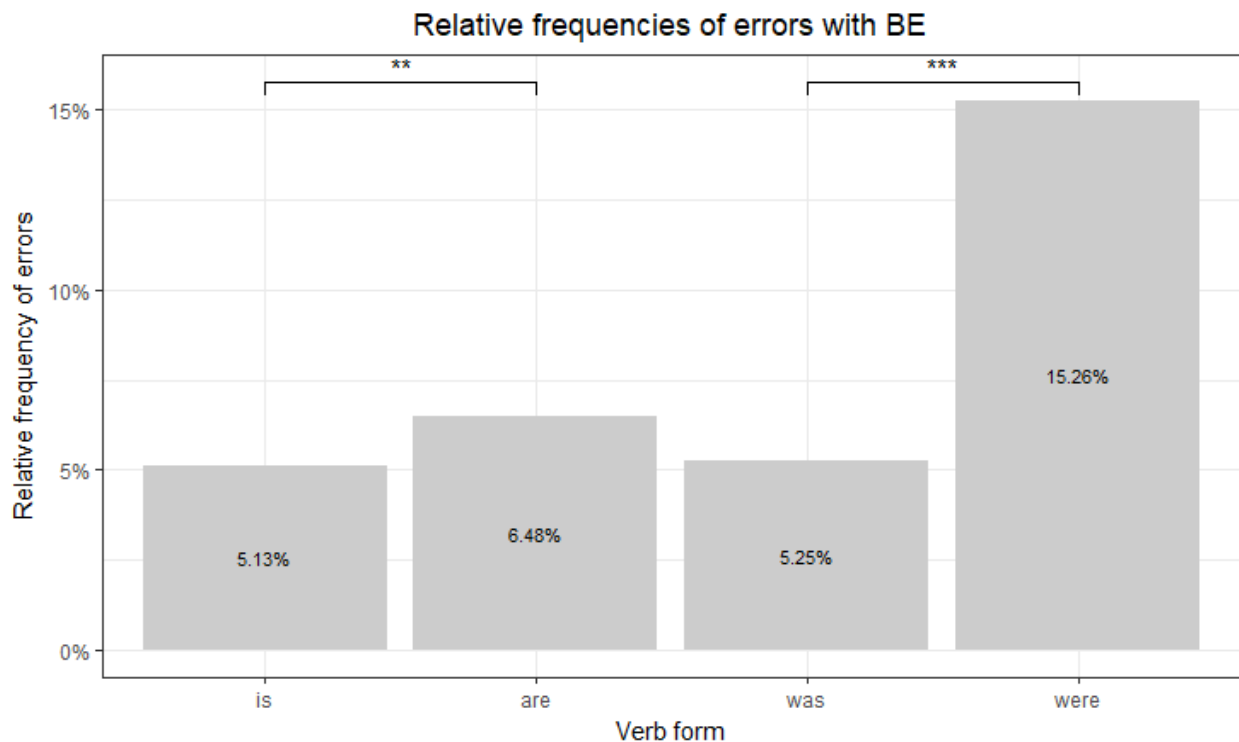
Figure 7. Erroneous use of the different forms of BE as percentages of the total occurrences of the individual finite forms in the corpus.

It seems there are different factors at play in the erroneous clauses with *BE* as the finite verb and in clauses with verbs other than *BE*. The students are more prone to overuse the plural forms of *BE* when they make an error in suppletive agreement as opposed to the overuse of the marked singular verb in affixal agreement errors. Despite the high number of errors involving the verb form *is*, the proportion of the errors with *is* is not high. There are statistically significant differences between the rates of suppletive agreement errors in plural and singular, both in the present tense and in the past tense. Because the possibility of phonological influence on the error rates with *BE* depends on the number of the verb, the errors with *BE* are grouped according to number and not tense in the following:

47

Relative frequencies of errors with BE distributed by number
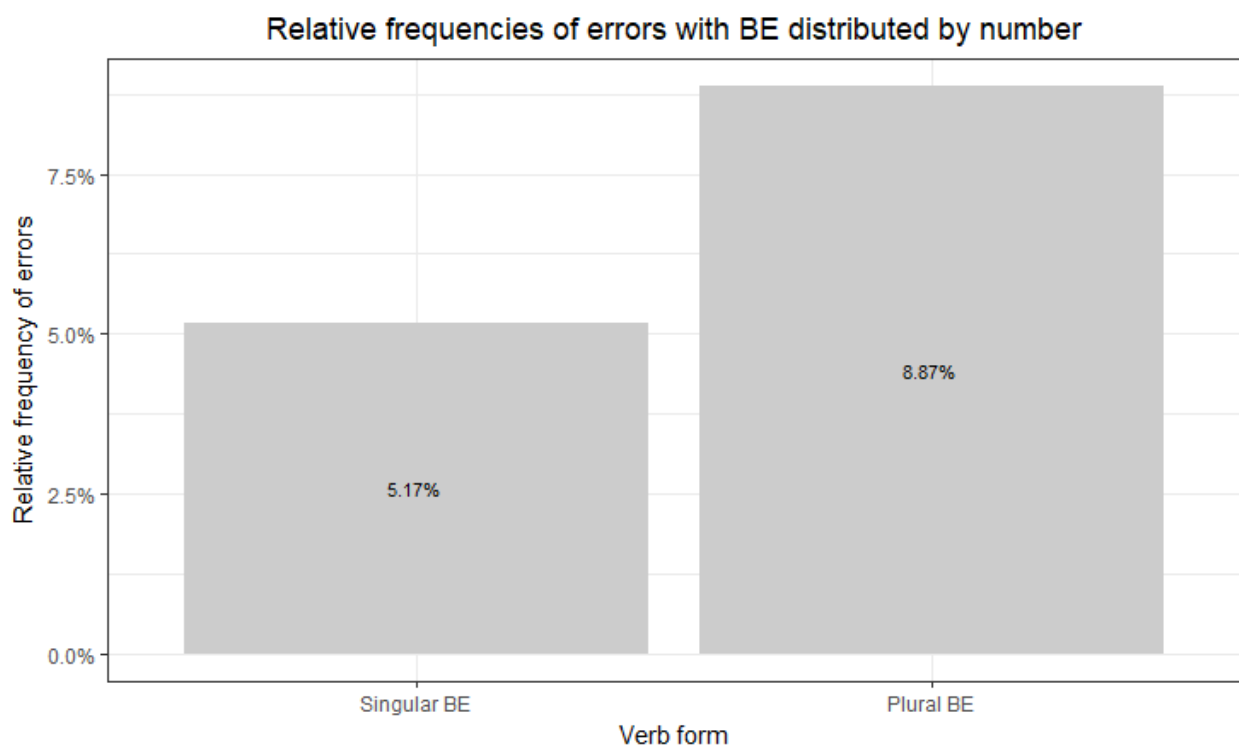
Figure 8. The distribution of errors with BE based on the number of the verb.

5.4   Subject types

Based on previous studies, it is predicted that the complexity of the subject has an influence on the agreement error rate of L2 users (Jensen et al., 2017; Ocampo, 2013; Thagg Fisher, 1985). Certain contexts, such as long NPs with postmodifiers, or subjects which are not immediately adjacent to the finite verbs often cause problems even for native speakers of English (see Biber et al., 1999). In addition, there are potentially different issues involved in erroneous agreement produced in clauses with *BE* as the finite verb, i.e. errors in suppletive agreement, and in clauses with other finite verbs, i.e. errors in affixal agreement (see above). In clauses with *BE* as the finite verb, there are significantly more errors found after singular subjects (overuse of the plural verb forms) while in clauses with other verbs, there are marginally more errors found after plural subjects (overgeneralization errors). In order to be able to discuss the error rates in the different subject-verb combinations, the affixal agreement errors and the suppletive agreement errors are discussed separately for each subject type in the following.

5.4.1 Subjects with other verbs – affixal agreement

There are more overgeneralization errors in the clauses with affixal agreement than omission errors, i.e. clauses where the subject is plural, but the verb is singular, as in (39):

(39)   **Friends** <u>makes</u> us believe in ourselves. (15MKV07_Mar16)

As mentioned in section 5.3.1, the distribution of overgeneralization and omission errors is not the same with all subject types. The data set includes both overgeneralization errors and omission errors in all categories of subjects, except after clauses as subjects where the only recorded error type is an omission error. Examples of errors in each category are given in (40) – (50). Table 7 summarizes the distribution of the two error types across subject types:

(40)   **She** <u>change</u> her behaviour. (15STK09_Sep16)

(41)   **They** <u>loves</u> their flag. (13STK06_Apr14)

(42)   **Everybody** <u>have</u> a weakness. (14STV22_Apr15)

(43)   **Some** <u>starts</u> getting sad. (15MKV15_May16)

(44)   **This** <u>keep</u> on going. (15STK02_Feb16)

(45)   **Those** who <u>needs</u> help. (15STV01_Dec15)

(46)   **There** <u>have</u> been a long and tough fight. (15STV17_Dec15)

(47)   **There** <u>exists</u> many fake friends. (14STV02_Apr15)

(48)   **To take a drink** <u>have</u> been traditional in Britain since… (14STV20_Dec14)

(49)   **The reader** <u>do</u> not get the same feeling. (15STV52_Feb16)

(50)   **Colors** <u>plays</u> a big part in this. (15STV31_Feb16)

Table 7

Distribution of affixal errors across subject types

	PersPro	IndefPro	DemPro	Expletive	Clause	**NP**	Total
Omission	**307**	71	27	6	19	**483**	913
Overgeneralization	**156**	16	13	10	0	**788**	983

The distribution of omission and overgeneralization errors in clauses with indefinite pronouns, demonstratives, expletives, and clauses as subjects is clearly connected to the content of these categories. For example, in the category of indefinite pronouns the most common subjects are compound pronouns, such as *everybody*,

everyone, etc.. These pronouns can be semantically confusing because they denote several people or things but they are syntactically singular. In addition, the Norwegian equivalent of *everybody* (*alle*) is syntactically plural. The most common error with such pronouns is the omission of the obligatory *-s*, probably because of the clash of semantics and syntax or due to the transfer of the properties of the specific lexical items from the L1 of the learners. The distribution of errors in clauses with personal pronouns and noun phrases as subjects (marked bold in Table 7) clearly differs. With personal pronouns as subjects, it is twice as likely for the students, if they make an error, to make an omission error (66.31% of the cases) than an overgeneralization error (33.69% of the cases). In the category of NPs, the reverse situation obtains. The chance for an overgeneralization error (62.00% of the cases) is much higher than for an omission error (38.00% of the cases). These two categories are discussed in more detail below.

5.4.1.1 *Affixal agreement errors with NP subjects*

There are three sub-categories within the NP subject category: coordinated subjects, post-modified NPs, and head-final NPs.[9] In order to make the discussion simpler, the category of post-modified NPs also includes cases where the agreement error involves a verb which is in a relative clause which is part of the NP, and clauses where the subject NP is followed by another clause element before the finite verb. The category of post-modified NPs thus includes all cases of NPs where the head noun is not directly followed by the finite verb. The distribution of the overgeneralization and omission errors in the category of NP subjects is illustrated in *Figure 9*:

---

9 "Head-final NP" is used here as a term for NPs which do not contain any post-head material. English is typologically a head-initial language, but if the NP does not contain any complements, the NP head is the last element in the phrase and it is thus adjacent to the finite verb.
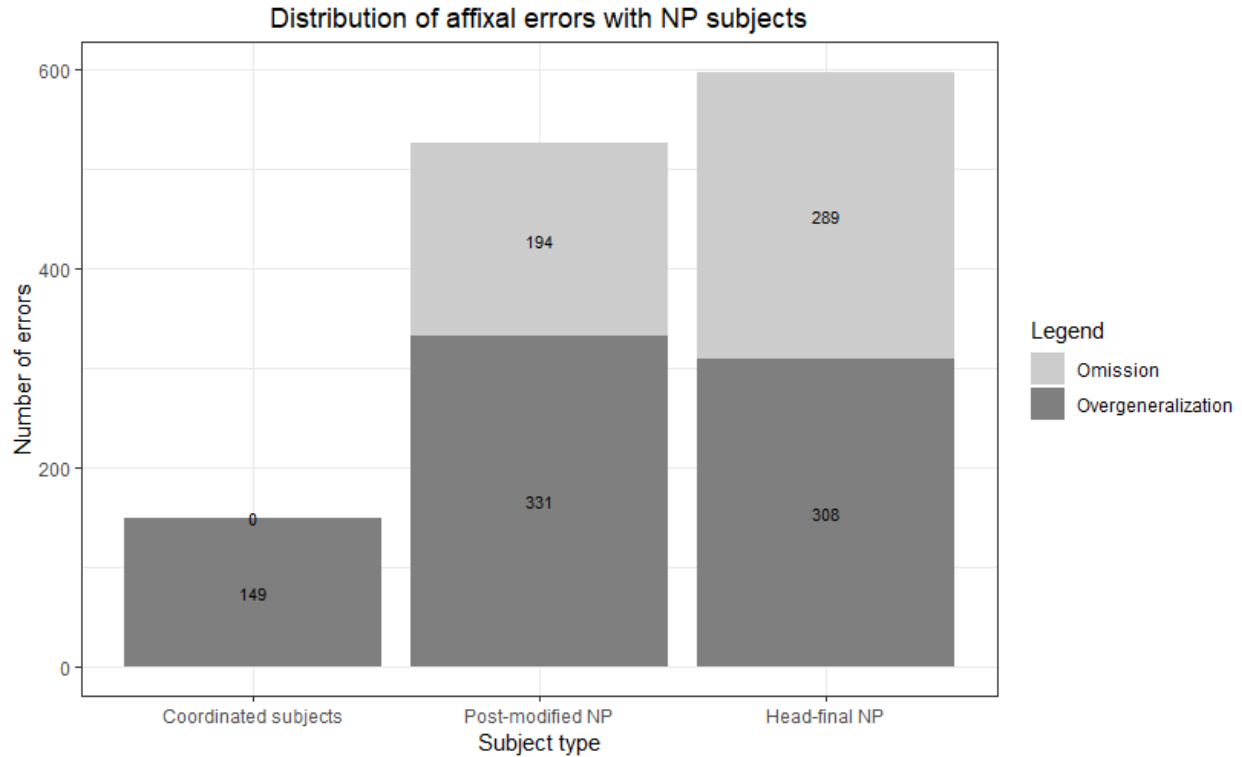
Figure 9. Distribution of errors in clauses with affixal agreement and NP subjects.

When there is any intervening material between the head of the subject NP and the verb, the chance of making an overgeneralization error as compared to an omission error is much higher (63.05% of overgeneralization errors and 36.95% of omission errors). The distribution of errors in the head-final NP category is practically equal between the two categories (51.59% of overgeneralization errors and 48.41% of omission errors). Note that this is a higher occurrence of overgeneralizations than previously reported in the literature (Breiteneder, 2005; Dröschel, 2011; Johansson, 2008; Thagg Fisher, 1985). In addition, agreement errors following a head-final subject are in fact the most numerous kind in the category of NP subjects. This raises questions about the conclusions of among others Ocampo (2013) who claims that the increased distance between the subject and the verb and an intervening noun of a different plurality is the main problem in agreement marking for L2 learners of English.

It is often claimed that if the intervening element (either within or after the subject NP) contains a noun which differs in number from the head of the NP this can cause 'proximity agreement' errors (Biber et al., 1999), i.e. the verb agrees with the last preceding noun instead of the head of the subject noun phrase. Clauses with intervening

elements attested in the corpus either contain postmodified subjects (51), or various adverbials between the subject and the verb, with (52) or without nouns (53) in their internal structure:

(51)   **An example of internal conflict** <u>are</u> Mrs. Forester's fear of being late. (14STA18_Sep14)

(52)   **The Democrats**, on the other hand, <u>wants</u> to draw a much softer line. (14STA13_Apr15)

(53)   **Most families** actually <u>grows</u> on these disagreements. (15STV52_Oct15)

In order to check whether proximity agreement can be a significant factor in the production of non-standard agreement after complex NP subjects, the clauses with NP subjects were further labelled based on the number of the last non-head noun in the subject NP. When the clauses with the post-modified NPs are divided between those which contain a non-head noun with a conflicting number (54) and those which do not (55),[10] it becomes clear that proximity agreement can be involved only in marginal cases in this student population (but see Jensen et al., 2017 for conflicting evidence).

(54)   **Tony's thoughts about the cop** <u>changes</u> throughout the story. (14STA02_Feb15)

(55)   **The pupils who go to state schools** <u>starts</u> at Primary School at the age of 5 as well. (14STA18_Nov14)

The distribution of errors in post-modified NPs with and without conflicting nouns is illustrated in *Figure 10*:

---

[10]   This category also includes clauses with intervening elements which do not include any nouns as example (53).
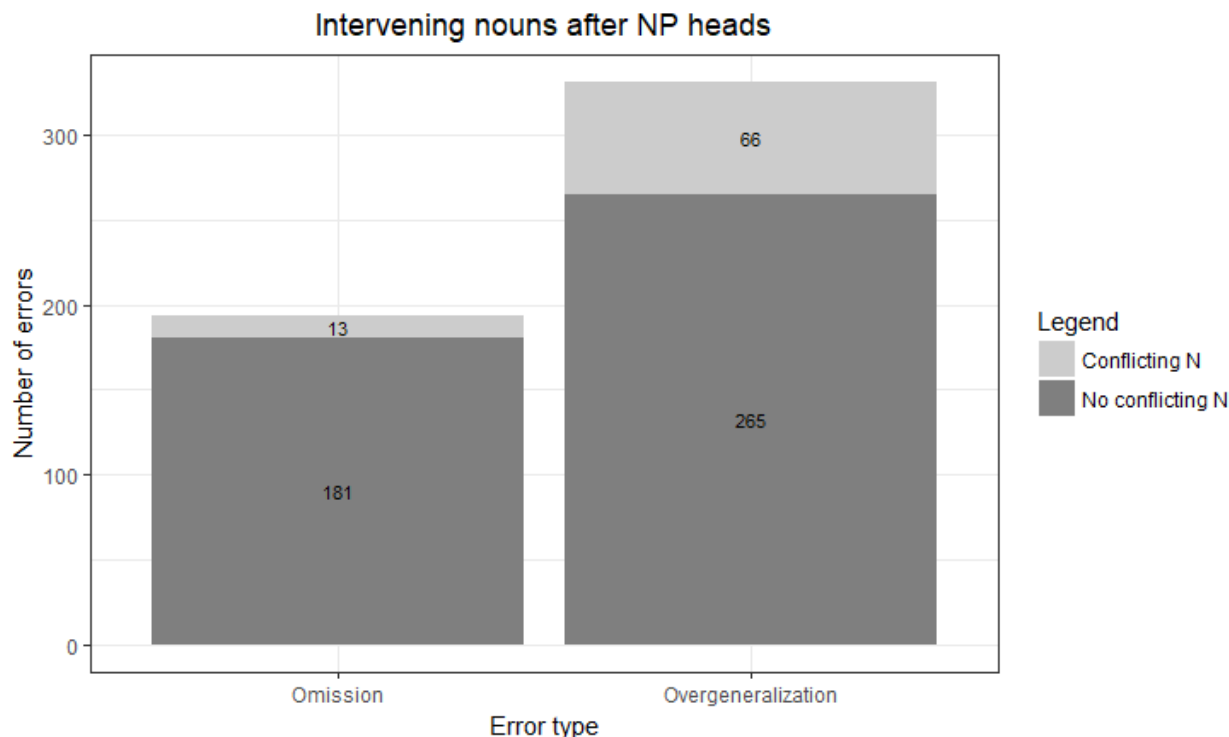
Figure 10. Distribution of errors in post-modified NPs based on the occurrence of a non-head noun with a conflicting number preceding the finite verb.

It seems that the major problem with the post-modified subject NPs is not the existence of a plurality conflict between the head noun and the noun immediately preceding the verb, but the general cognitive load. The students might get distracted by the intervening elements and not assign the number of the verb correctly. It is also conceivable that when such distraction happens, the students resort to their default form, which seems to be the inflected 3rd person singular form (two thirds of the errors involving post-modified NP subjects are overgeneralization errors). The production of the overgeneralization errors is more sensitive to the number on the preceding noun (conflicting number involved in 19.94% of the cases) than the production of the omission errors (conflicting number involved only in 6.70% of the cases), but both these proportions are fairly low as compared to the overall number of errors in this category.

5.4.1.2 *Affixal agreement errors with personal pronouns as subjects*

Agreement errors with personal pronouns as subjects should be infrequent due to the semantic transparency of pronouns as compared to complex or abstract NPs or indefinite and demonstrative pronouns. The number assignment should also be cognitively simpler

with personal pronouns as compared to irregular NPs or other types of pronouns. Despite that, clauses with personal pronouns as subjects constitute about one third of all affixal agreement errors. In contrast to the NP subject category, the most common errors in the category of personal pronouns as subjects are omission errors, which is the type of error commonly reported for speakers of other L1s learning English. *Figure 11* illustrates the distribution of the errors across individual personal pronouns:

Distribution of affixal errors with personal pronouns as subjects
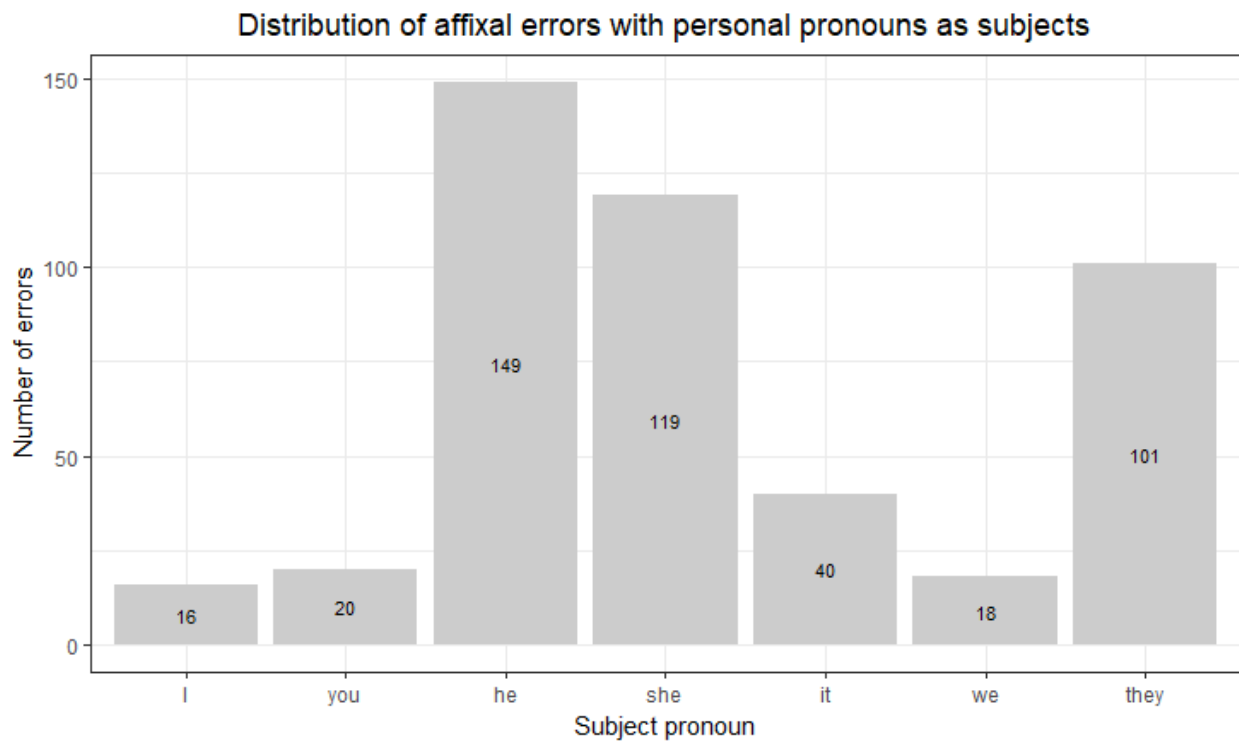
Figure 11. Distribution of affixal agreement errors across personal pronouns as subjects.

The large proportion of omission errors also entails that the most represented subjects will be pronouns *he*, *she*, and *it*. In the cases of overgeneralization errors, the most common subject pronoun is *they*. When overall frequencies of the individual personal pronouns in subject positions are considered, the difference between overgeneralization and omission errors becomes even more pronounced. Table 8 provides an overview of the number of total occurrences of personal pronouns as subjects and the relative frequencies of errors in clauses with pronominal subjects:

Table 8

Relative frequencies of affixal errors after personal pronouns as subjects

Pronoun	I	you	he	she	it	we	they
Total occurrences	2831	1742	2293	1484	2845	1408	2482
Error frequency	0.57%	1.12%	6.50%	8.02%	1.41%	1.28%	4.07%

While the pronouns *I* and *you* occur in clauses with erroneously marked agreement only in about 1% of the cases, the pronoun *they* occurs in clauses with erroneous agreement in 4% of the cases. This could suggest some overgeneralization of the learned rule, i.e. a 3rd person receives the *-s* morph regardless of the number. However, there is a statistically significant difference between the frequencies of overgeneralization and omission errors in the 3rd person (chi-square test (Field et al., 2012, pp. 812-828) for *he* vs. *they*, $\chi^2(1) = 14.17, p < 0.001$; for all omission errors vs. all overgeneralization errors , $\chi^2(1) = 99.28, p < 0.001$). This makes the possibility of an extension of the learned rule into plural less likely because, despite the high frequency of the 3rd person plural pronoun *they*, the agreement errors involving *they* as a subject are significantly less frequent than the errors involving *he* as a subject pronoun.

Another possible explanation for the different distribution of affixal agreement errors following NP subjects and personal pronouns as subjects could lie in the general complexity of the language used. Since the number assignment for personal pronouns is more transparent than for complex NPs it is plausible that the accuracy in number assignment for personal pronouns should be acquired earlier than the number assignment for complex NPs. Thus, students who are less proficient in the L2 might make more agreement errors in clauses with personal pronouns as subjects than students who are more proficient, while both groups continue to struggle with number assignment for complex NPs. At the same time, students who are less proficient in the L2 might resort to simplification strategies, i.e. affix omission, more often than students who are more proficient. In spite of the general error scores being comparable between the GS students and the VS students (see section 5.2), the VS students stand for a higher proportion of errors in clauses with personal pronouns as subjects than the GS students. The texts of the VS students contain a total of 359 affixal agreement errors (see Table 5 in section 5.2)

and 124 of these erroneous clauses have personal pronouns as subjects. The texts of the GS students contain 1537 affixal agreement errors and 339 of them involve personal pronouns as subjects. In other words, the VS students stand for 26.78% of the affixal errors with personal pronouns while they in total produced only 18.93% of all affixal errors in the corpus. This difference in the error proportions is statistically significant (chi-square test (Field et al., 2012, pp. 812-828) for affixal errors with PersPro vs. affixal errors with other subjects for VS and GS students , $\chi2(1) = 24.58$, $p < 0.001$).

As with the NP subjects, the increased cognitive load caused by intervening elements between the personal pronoun and the finite verb could be contributing to the error rates. *Figure 12* shows the distribution of errors across personal pronouns divided by the presence of intervening elements between the pronoun and the finite verb:
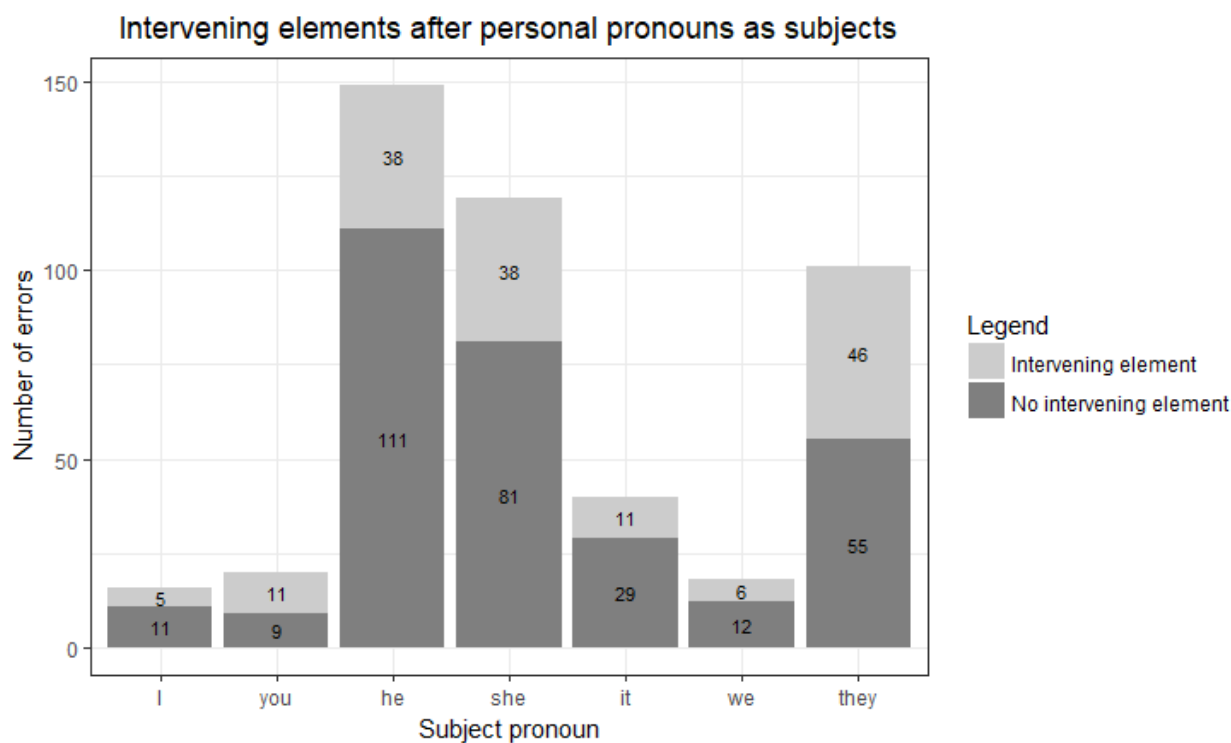
Figure 12. Distribution of errors across individual personal pronouns divided by the presence of intervening elements between the pronoun and the finite verb.

Intervening elements are involved in about one-third of the cases of omission errors while they are present in about half of the cases of overgeneralization errors. As with NP subjects, the increased cognitive load caused by the intervening elements could be part of the explanation of the relatively high rate of overgeneralization errors in clauses

with plural pronominal subjects. And as is suggested in section 5.4.1.1, it seems that the increased cognitive load more often triggers the inflected form of the verb, as if this form was the base form for these learners.

5.4.2 Subjects with BE – suppletive agreement

As with affixal agreement, there are errors in clauses with *BE* recorded in almost all possible categories of subjects. There are some differences in the proportions of the forms of *BE* across the subject categories, but the most common error type overall involves erroneous use of *is* with plural subjects. In contrast to the category of errors with other verbs where there are many different verbs involved, the errors involving *BE* can easily be expressed as relative frequencies of the total occurrence of each verbal form in the corpus. Such approach is more representative of the potential problems with suppletive agreement than just raw error numbers. All recorded error types with *BE* are exemplified in (56) – (78). Table 9 gives an overview of the distribution across the different forms of *BE* and the subject categories:

(56) …something **I** of course <u>is</u> against and do not like. (13STK06_1Mar14)

(57) **She** <u>are</u> upset. (15STK21_Sep15)

(58) **We** <u>was</u> behind one of the goals. (14STV19_Oct14)

(59) **It** <u>were</u> made into a movie. (15STK03_Apr16)

(60) **Some** <u>is</u> aiming at not so good jobs. (15STV40_May16)

(61) **Everybody** <u>are</u> dressed the same. (14STV22_Oct14)

(62) **Both** <u>was</u> against racism. (15STV11_Dec15)

(63) **Everyone** <u>were</u> treated terribly. (14STA08_Apr15)

(64) **That** <u>is</u> questions many will ask. (15STV10_Dec15)

(65) **This** <u>are</u> the reasons why I want to... (15STV36_Dec15)

(66) **This** <u>was</u> the most violence riots… (13STK03_2Mar14)

(67) **That** <u>weren't</u> such a big problem. (14STV18_Oct14)

(68) **There** <u>is</u> many tribal issues. (13STK01_2Oct13)

(69) **There** <u>are</u> still racism. (15STK12_Feb16)

(70) **There** <u>was</u> many journalists. (15HOV12_Oct15)

(71) **There** <u>were</u> criminality. (15STV03_Dec15)

(72) W**hat are people willing to do** <u>are</u> almost limitless. (15STK21_Apr16)

(73) **"Elisabeth, let me explain"** <u>was</u> the last words. (15STK26_Apr16)

(74) …but since **"to eat a haggis"** <u>were</u> on my bucket list… (14STV13_Oct14)

(75) **Differences** <u>is</u> good. (15STV14_Dec15)

(76) **The Indian culture** <u>are</u> very colorful. (15STV20_Oct15)

(77) **Parties** <u>was</u> huge thing. (15STV07_Feb16)

(78) …**his brother** <u>were</u> killed too. (15SSK19_Nov15)

Table 9

Distribution of errors across subject types in clauses with suppletive agreement

Verb form	PersPro	IndefPro	DemPro	Expletive	Clause	NP
is	13	6	2	66	0	376
was	31	2	2	20	1	148
are	10	23	1	35	5	129
were	28	10	2	23	3	112

Unlike with affixal agreement, personal pronouns as subjects are only marginally represented in the erroneous clauses with suppletive agreement. Expletive subjects are the second largest category after NPs. However, the higher error rate with expletives in combination with *BE* is not surprising. Expletives often combine with *BE* and they are challenging both for language learners and for native speakers because the number of the formal subject is determined by the notional subject later in the clause (cf. Biber et al., 1999, pp. 185-186). Sobin (1994, 1997) argues that agreement in expletive constructions is deviant in English and the choice between the formally grammatical version and the commonly used version is mainly a question of prestige. A detailed exploration of the L2 usage of expletives is beyond the scope of this work so this category is not analyzed any further in the following despite the high error rates.

Errors involving plural forms of *BE* are overrepresented in the category of indefinite pronouns. This, again, is not surprising considering that the most common subjects in this category are compound pronouns, such as *everybody*, *everyone*, *everything*, which are singular in English despite referring to several people or things. The error rates with demonstrative pronouns and clauses as subjects are so low that it is difficult to discern any patterns. The errors in clauses with personal pronouns and NPs as subjects with *BE* as the finite verb are discussed in more detail below.

5.4.2.1   Suppletive agreement errors with NP subjects

The discussion of the distribution of suppletive agreement errors across different subject types is complicated by the fact that there are four possible forms of *BE* among the detected errors. Based on the arguments discussed in section 5.3.2 the suppletive forms are group by number, not by tense in the following. Errors involving singular forms of *BE* appear more often both in the present and in the past, but the proportions differ:

Distribution of suppletive errors with NP subjects
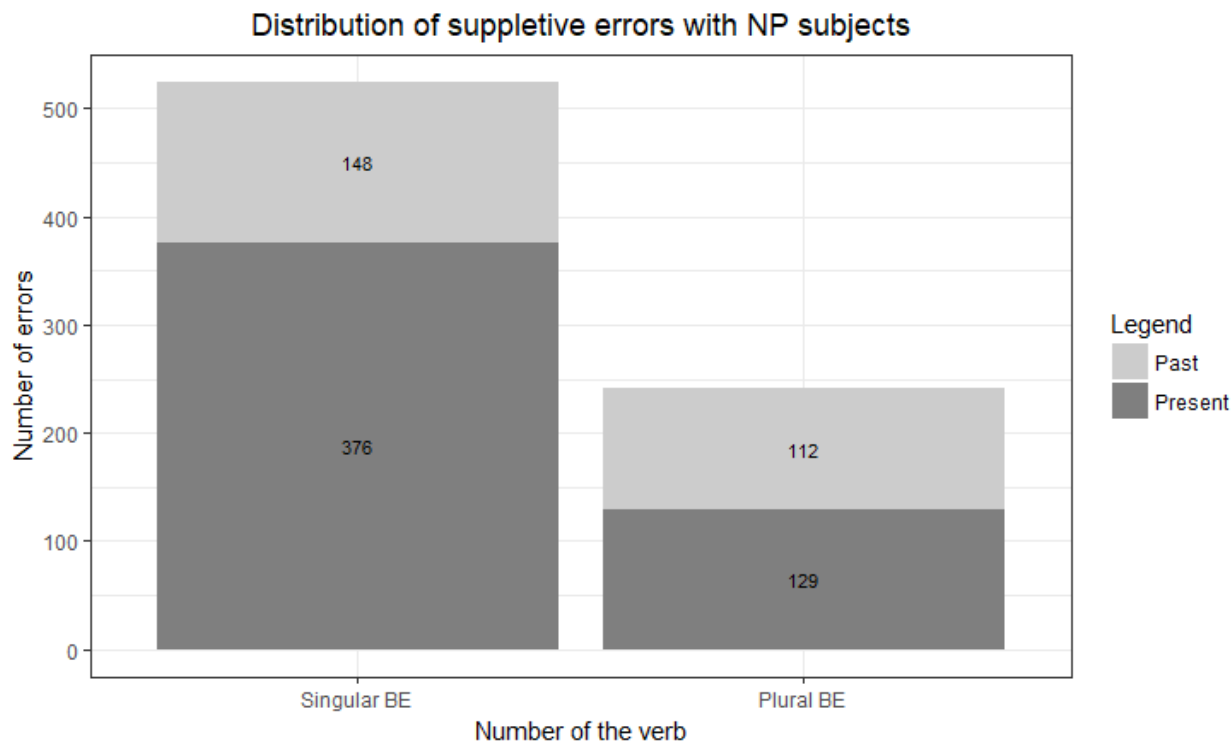
Figure 13. Distribution of suppletive errors in clauses with NP subjects.

The high rate of errors involving present singular *is* can be explained by the high overall frequency of this word in the corpus (9,037 occurrences, 51.4% of all forms of *BE* in the corpus). It is also plausible that this morph is the default option for the learners due to its high frequency in the language in general. However, recall from section 5.3.2 that the proportion of errors to correct usage is much higher for plural forms than for singular forms of *BE*. In the current form of the corpus, it is not possible to extract the clauses with NP subjects and correct suppletive agreement. It is thus not possible to evaluate whether the overall relative frequencies of errors and correct usage described in section 5.3.2 hold also for suppletive agreement after NP subjects in particular.

As with the affixal errors, the suppletive errors are further divided according to the types of NPs. The distribution of errors involving each type of NP and a specific finite form of *BE* is summarized in Table 10:

60

Table 10

Distribution of suppletive agreement errors across different types of NPs

Verb form	Coordinated subjects	Post-modified NP	Head-final NP
is	66	161	149
was	26	46	76
are	0	71	58
were	0	47	65

Head-final NPs as subjects constitute together 673 of the suppletive errors. These are the cases where the head-final subject and the finite verb are adjacent. There are more errors involving *is* in the category of post-modified NPs, but the total number of suppletive errors after post-modified NPs, 325, is less than half of the suppletive errors after head-final NPs. This again suggests that the increased distance between the head of the subject phrase and the finite verb is not the main problem for these learners. *Figure 14* illustrates the distribution of suppletive errors following post-modified NP subjects with non-head nouns with a conflicting number:
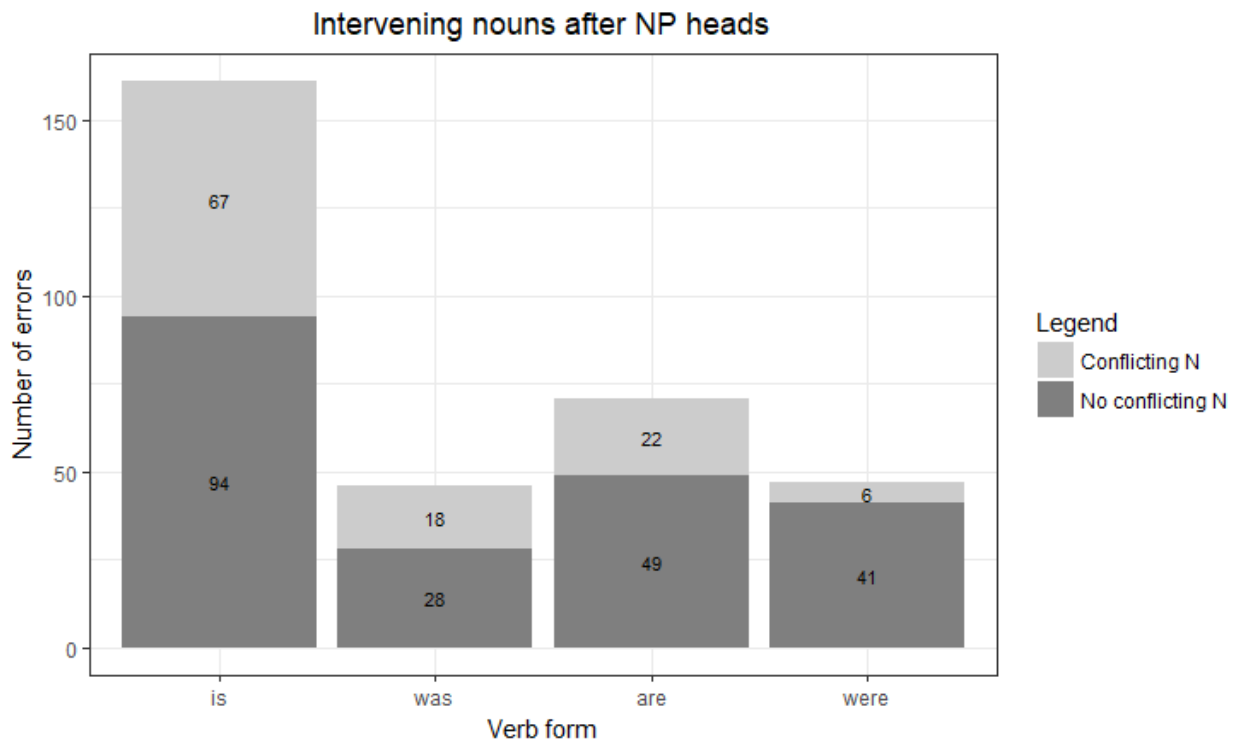
Figure 14. Distribution of suppletive errors following post-modified NP subjects involving nouns with conflicting number.

It seems that the production of suppletive errors is more sensitive to the presence of intervening elements with conflicting number as compared to affixal errors. Recall from section 5.4.1.1 that the proportions of the affixal errors following intervening elements with conflicting number were low in the category of post-modified NP subjects followed by finite verbs other than *BE* (6.70% for omission errors, i.e. singular subject with a final plural non-head noun followed by a plural verb, and 19.94% for overgeneralization errors, i.e. plural subject with a final singular non-head noun followed by a singular verb). In clauses with *BE* involving intervening elements with conflicting number only errors with *were* fall within the same range (12.77% errors involving conflicting number on the last preceding non-head noun). The proportions of clauses with intervening elements among errors with singular *BE* are especially high (41.61% for errors with *is* and 39.13% for errors with *was*). Errors with *are* contain intervening elements with conflicting number in 30.99% of the cases.

In absolute numbers, there are fewer suppletive errors than affixal errors registered in the corpus. Ionin & Wexler (2002) suggest that suppletive agreement is acquired before affixal agreement because finite *BE* raises to I while finite lexical verbs stay in-situ (see chapter 2.2.1). White (2003) also observes that L2 learners of English rarely make errors in suppletive forms. Nevertheless, the Norwegian learners produce suppletive agreement errors more frequently than previously reported for other learners (e.g. Ionin & Wexler, 2002). The different raising patterns could be part of the explanation why suppletive errors are less frequent than affixal errors in the Norwegian learner data. If the suppletive agreement is acquired earlier and it is more automatized than the affixal agreement, there should be more processing power available to the learner to retain the information of the number of the subject across intervening elements, especially if there are no conflicts in number within the subject NP (i.e. in cases where the only factor is the length of retention). The higher proportion of the cases of erroneous use of *BE* after intervening elements which differ in number from the subject as compared to the same type of errors with other verbs could suggest that this is the case. If the agreement with *BE* is more automatized for the learners (i.e. more native-like) they might be more prone to proximity agreement, just as native speakers.

5.4.2.2 Suppletive errors with personal pronouns as subjects

The suppletive agreement errors following personal pronouns as subjects are not a numerous category. There are only 82 errors in this category, i.e. only 2.79% of all recorded agreement errors. *Figure 15* illustrates the distribution of the suppletive errors following personal pronouns as subjects across the forms of *BE*:
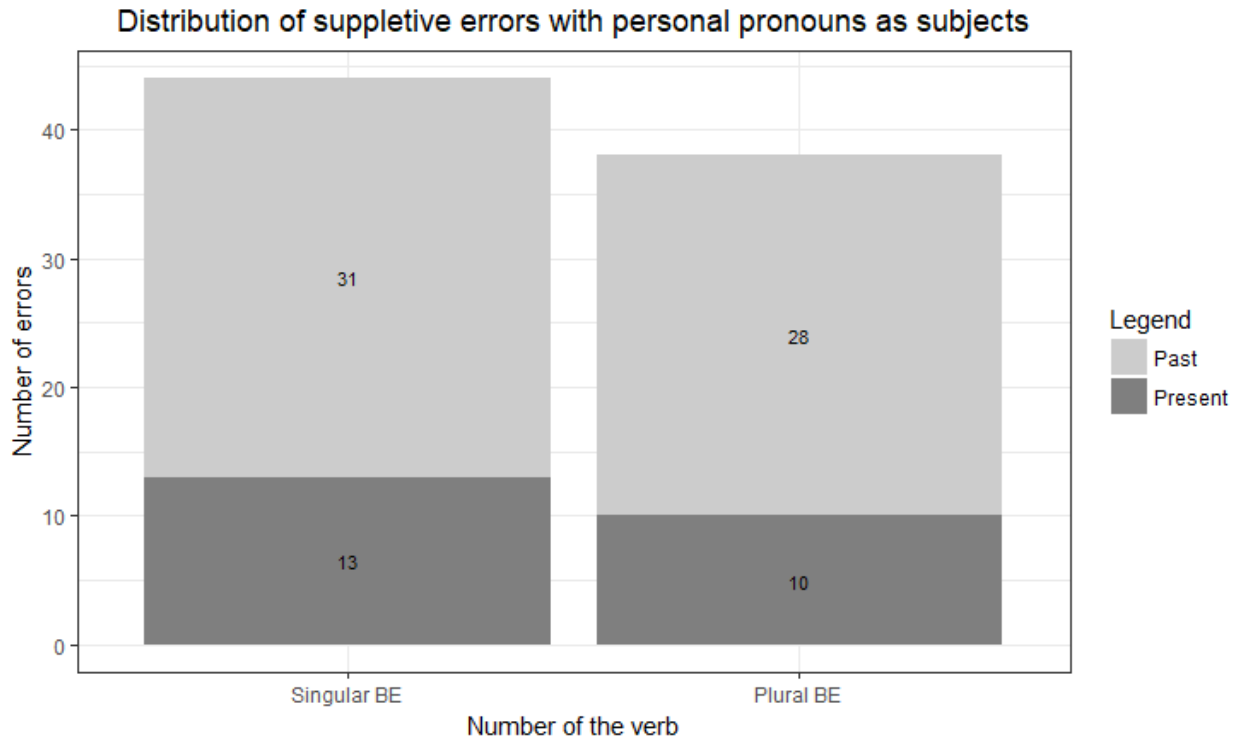
Figure 15. Distribution of suppletive agreement errors following personal pronouns as subjects.

While errors involving singular *BE* are clearly dominant with NP subjects, the clauses with pronominal subjects show a different distribution. There are almost as many errors involving plural *BE* after a personal pronoun as singular *BE*. It also seems that the combination of a pronominal subject and *BE* in the present tense is not problematic for the students whereas *BE* in the past tense leads to problems for the learners. *Figure 16* presents the distribution of the errors among the individual personal pronouns:

63
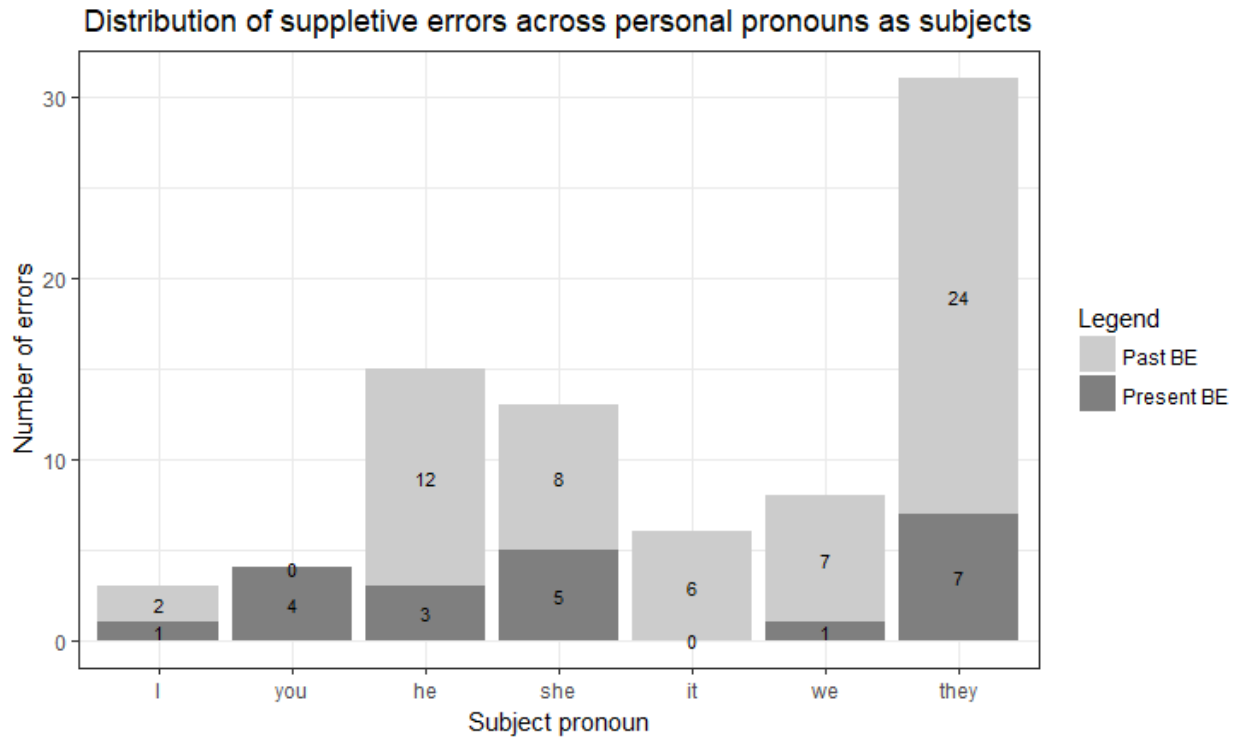
Figure 16. Distribution of suppletive agreement errors following individual personal pronouns as subjects.

With all pronouns, except for the second person, the past tense forms of *BE* create more problems than the present tense forms. However, the individual combinations are not that frequent in the corpus. It is thus difficult to draw any conclusions for such a limited set. Similarly, there are only few erroneous cases with *BE* involving an intervening element following a personal pronoun (20 instances). The distribution of these errors among the individual pronouns is illustrated in *Figure 17*:
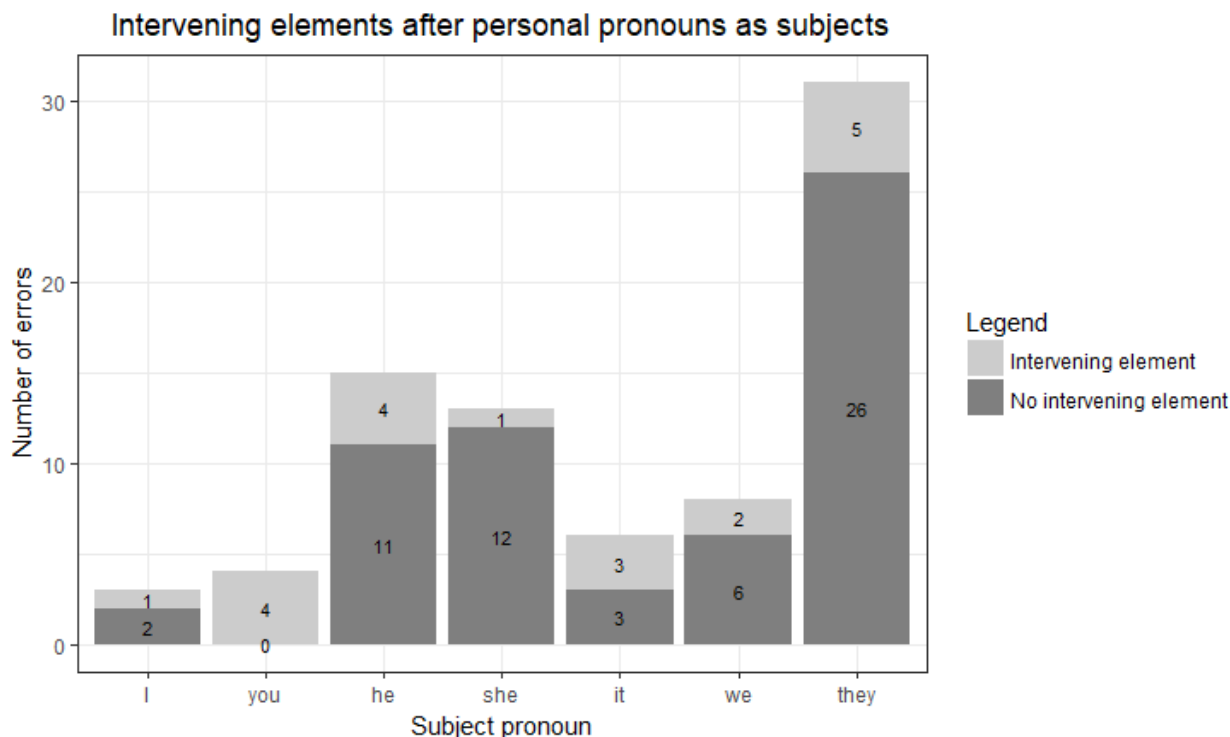
Figure 17. Distribution of suppletive agreement errors in clauses with personal pronouns as subjects and intervening elements between the subject and the finite verb.

The cases with intervening elements after personal pronouns are so few that it is impossible to make any generalizations. Most of the cases involve coordinated structures where two or more conjoined clauses share the same pronominal subject. However, only one of these 'intervening' clauses ends with a noun with a conflicting number (79). Other intervening elements include adverbials, as in (80), or short postmodifiers (81):

(79)   …**they** <u>lost</u> most of their **land**, <u>was</u> put in camps or concentrations and <u>got</u> in fights with the whites. (13STK01_2Mar14)

(80)   **You** still <u>is</u> not sure what you want. (15STV40_May16)

(81)   …**we as individuals** <u>is</u> communicating and connecting more than ever! (14STV13_Apr15)

5.5   Possible error triggers across categories

So far, the discussion has only concerned the distribution of errors based on the morphosyntactic properties of the subjects and the finite verbs. However, there are other factors to consider in order to get the full picture of agreement marking difficulties. If

there is a semantic uncertainty involved, such as with notional agreement and semantically ambiguous nouns, the resulting agreement error could be seen as semantically triggered and not a result of lacking competence. Similarly, the processing load increases for the speaker when there are intervening elements involved between the subject and the verb. If the result is an agreement error, this could be triggered by the distance between the subject and the verb and thus not be due to lack of competence. Such errors are not uncommon even among native speakers, especially in spontaneous oral production. In written production, the learners should, in theory, have more control of the output as they can proof-read and revise their production. However, it is not certain whether these students had time (or skills) to proof-read their texts.

The following sections first discuss the errors which could be triggered by the semantics of the subject or by the distance between the subject and the verb. If an erroneous clause contains a subject with a semantically difficult head noun which is also post-modified, such errors are included in the distance-triggered category. The remaining errors are labelled idiosyncratic in the following (marked bold in Table 11):

Table 11

Overview of agreement errors in different subject categories divided into possible triggering categories

	NP	**PersPro**	IndefPro	DemPro	Expletive	ClaS
Semantically triggered	445 (21.87%)	6 (1.11%)	73 (56.15%)	28 (59.57%)	158 (98.75%)	14 (50%)
Distance-triggered	852 (41.87%)	191 (35.30%)	57 (43.85%)	19 (40.41%)	2 (1.25%)	14 (50%)
Idiosyncratic errors	**738 (36.27%)**	**344 (63.59%)**	-	-	-	-
Total	2035 (100%)	541 (100%)	130 (100%)	47 (100%)	160 (100%)	28 (100%)

5.5.1 Semantically triggered errors

Semantically difficult subjects can be divided into two categories: those which are specific to English and are thus difficult for learners regardless of their L1s, and those which are difficult for contrastive reasons. Subjects which are semantically difficult in

English include for example nouns which are ambiguous with regard to number (e.g. nouns such as *team, government*) and can be treated either as singulars or plurals depending on the context and the variety of English (Biber et al., 1999). Other subjects which can be semantically challenging for learners regardless of their L1s include nouns which are irregular and do not take the regular plural morph *-s* (e.g. *children, the British, mice*). Nouns and indefinite pronouns which are inherently singular or plural (e.g. *police, people, everybody*) as well as non-count nouns (e.g. *money, gold*) pose difficulties mainly if their properties differ in the target language and the L1 of the speakers.

Agreement errors involving irregular, non-count and inherently singular or plural nouns are marked as errors in the data. All such errors fall into the category of semantically triggered errors. Clauses with errors involving nouns from the first category (subjects with ambiguous number, such as example (82)) are excluded from the material because it is often problematic to determine which variety of English the students intend to use, and consequently whether such use would be erroneous in that variety:

(82)   **The government** <u>do</u> not stop them. (14STA18_Apr15)

All agreement errors in clauses with clauses, demonstratives or indefinite pronouns as subjects can be considered semantically triggered as these categories of subjects have distinct inherent morphosyntactic properties which make correct agreement marking difficult for L2 learners of all L1 backgrounds. Erroneous clauses with coordinated subjects and expletive subjects are also considered semantically triggered.

5.5.2   Distance-triggered errors

Subject-verb agreement errors may also be triggered by the distance between the head of the subject noun phrase and the finite verb. The first difficulty here, which is also common among native speakers, arises when the last noun in the subject noun phrase differs in number from the head noun – so called proximity agreement (see Biber et al., 1999, pp. 189-190):

(83)   **Some people**, like <u>Lieutenant Dan</u>, **has** difficulties. (15STK28_Dec15)
(84)   **Jules mum**, just as <u>many others</u> **are** not supporting gays and lesbians. (15STV40_Dec15)

Previous studies (Jensen, 2016; Ocampo, 2013) show that also L2 learners may mark proximity agreement instead of grammatically correct agreement. However, Jensen

and Ocampo come to opposite conclusions as to which combination of the number of the head of the subject NP and the last preceding noun is the most difficult (see chapter 3.3). The data from the corpus of Norwegian young learners do not show strong influence of proximity agreement. Such errors are found, but they constitute a minority as compared to agreement errors after head-final subjects or after intervening elements with no number conflict (see *Figure 10* above for affixal agreement after NP subjects, *Figure 12* for affixal agreement after personal pronouns as subjects, *Figure 14* for suppletive agreement after NP subjects, and *Figure 17* for suppletive agreement after personal pronouns as subjects).

Another issue with the distance between the subject and the finite verb is connected to the short-term memory and processing capabilities – the learner may 'forget' what number the subject has by the time the finite verb is produced. This problem is more pronounced in spontaneous oral production than in writing, but it is a plausible argument that the attention and processing capacities of the L2 writers are overloaded by the task of producing a text in a foreign language, and they do not have enough resources to keep track of the number of the head of the subject noun phrase until the finite verb is produced. The fact that the students sometimes manage to produce the correct verb form immediately following the subject while they fail to produce the correct form later in the sentence (viz. (85) and (86)) also lends support to this argument:

(85)   **Nelson Mandela** <u>was</u> the main person to fight this bad idea, but **were** sendt to jail for life. (13STK04_Oct13)

(86)   In this passage, **Dahl** <u>slows</u> down the pace to the very second and therefore **let** the readers have a moment thinking what actually happens. (14STA12_Sep14)

Errors which can be ascribed to proximity agreement are in a sense legitimate as these are not uncommon also among native speakers of English. Errors which seem to be caused by distance only, i.e. with no intervening nouns, or with intervening nouns with the same number as the head of the subject NP, cannot be explained by feature checking applied on a 'wrong' element. *Figure 18* illustrates the distribution of potentially distance-triggered affixal agreement errors across NPs and personal pronouns as subjects excluding proximity agreement:

Distribution of distance triggered affixal agreement errors across subject types
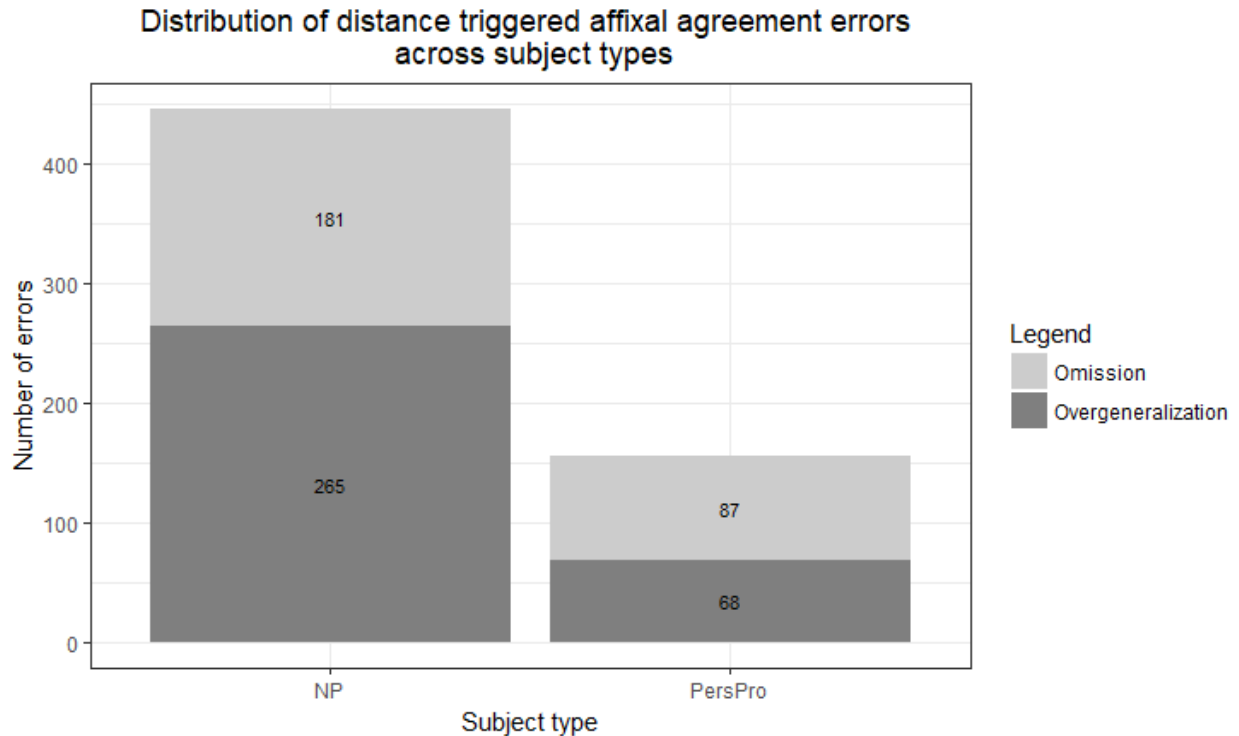
Figure 18. Distribution of potentially distance-triggered affixal agreement errors across NPs and personal pronouns as subjects.

Recall from section 5.4.1 that clauses with NP subjects are more prone to erroneous overgeneralization of the affixal agreement, while clauses with personal pronouns as subjects are more likely to contain an erroneous omission of the agreement marker. This ratio holds also in the group of potentially distance-triggered affixal agreement errors, although it is more pronounced in the case of NP subjects. If it is assumed that the distance between the head of the subject NP and the finite verb increases the processing load for the learner, an unmarked form of the verb would be expected as the less costly default option.[11] However, the Norwegian learners use the marked form of the verb in almost 60% of the clauses where the agreement marking is complicated by intervening elements after the head of the subject NP and only few of these errors can be attributed to proximity agreement (see section 5.4.1.1).

Figure 19 illustrates the distribution of suppletive agreement errors across subject types in clauses with intervening elements excluding potential proximity agreement:

---

[11] I follow Haspelmath's definition of markedness as overt coding, i.e. the marked form is morphologically more complex and often also phonologically longer than the unmarked form (Haspelmath, 2006, pp. 29-30).
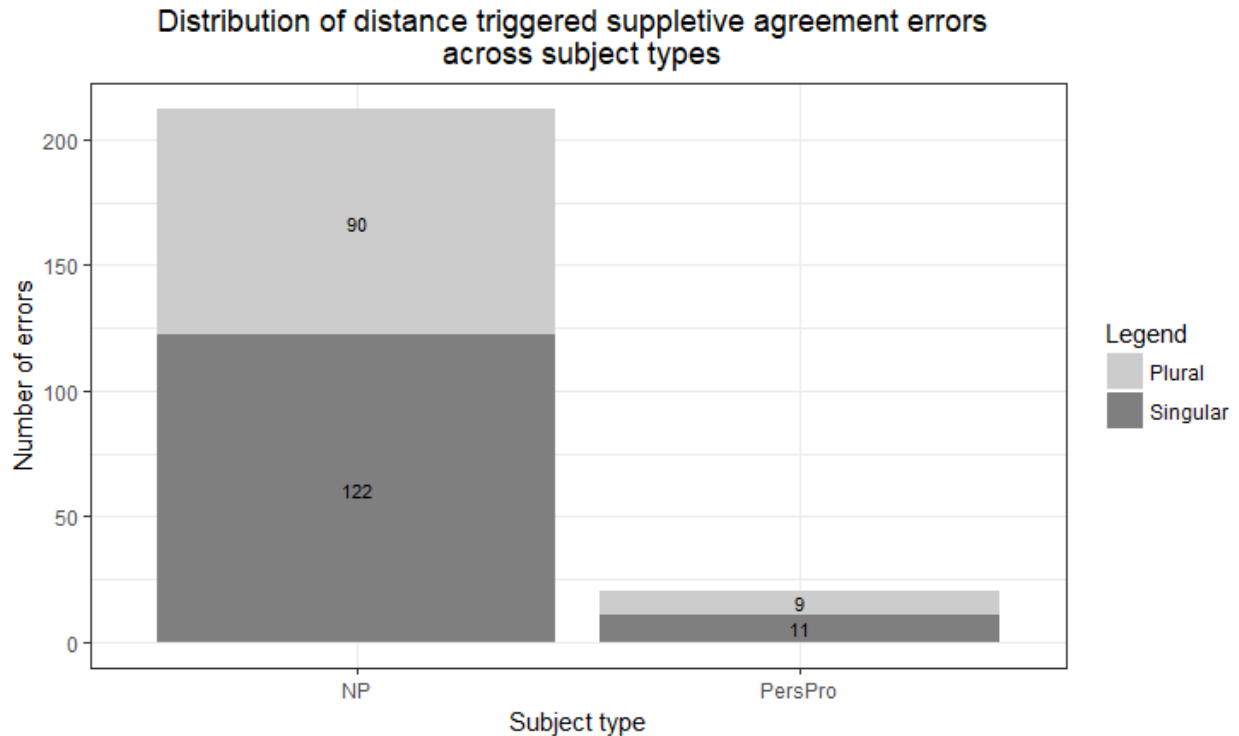
Figure 19. Distribution of potentially distance-triggered suppletive agreement errors across NPs and personal pronouns as subjects.

As discussed in section 5.3.2, there is a possibility that the suppletive agreement errors in the texts of Norwegian L2 learners of English could be phonologically triggered. Even though in raw numbers the errors involving singular forms of *BE* constitute a majority of the suppletive agreement errors, recall that singular forms of *BE* occur in the corpus more than three times as often as plural forms (12,922 occurrences of *is* and *was* as compared to 4,262 occurrences of *are* and *were*). Taking in the consideration the distribution of the individual suppletive forms, it seems that the erroneous use of plural forms of *BE* is overrepresented in the category of potentially distance-triggered errors despite the lower total count. As with affixal errors this distribution is more pronounced with NP subjects than with personal pronouns as subjects.

5.5.3 Idiosyncratic errors

When all the errors that seem to be justifiable in terms of semantic properties or processing constraints (distance and intervening elements) are removed from the data, there are still some errors left that cannot be accounted for in any other way than as 'idiosyncratic' errors (see Table 11 above). It is expected that L2 learners will struggle

with agreement in situations where the subject, to some degree, has idiosyncratic grammatical features (e.g. irregular plural, uncountability, etc.) which need to be included in the mental lexicon instead of rule learned (semantically triggered errors). It is also expected that short-term memory and processing constraints may play a role (distance-triggered errors). What is of concern here is the errors which appear to be completely idiosyncratic. In the category of noun phrase subjects, the portion of such errors constitutes 36% of the cases, while in the category of personal pronouns as subjects almost 64% of the errors seem to be idiosyncratic. It could be argued that such errors may only be performance errors which do not reflect the actual competence of the speakers. However, results of a recent acceptability judgment study conducted by Jensen (2016; see also Jensen et al., 2017) show that Norwegian learners often accept faulty agreement in grammaticality judgement tasks, which would suggest that the incorrect agreement marking cannot be attributed only to performance (see chapter 3.3 and chapter 6 for discussion). As in the previous sections, affixal agreement errors and suppletive agreement errors are treated separately. *Figure 20* shows the distribution of affixal agreement errors which appear to be idiosyncratic:
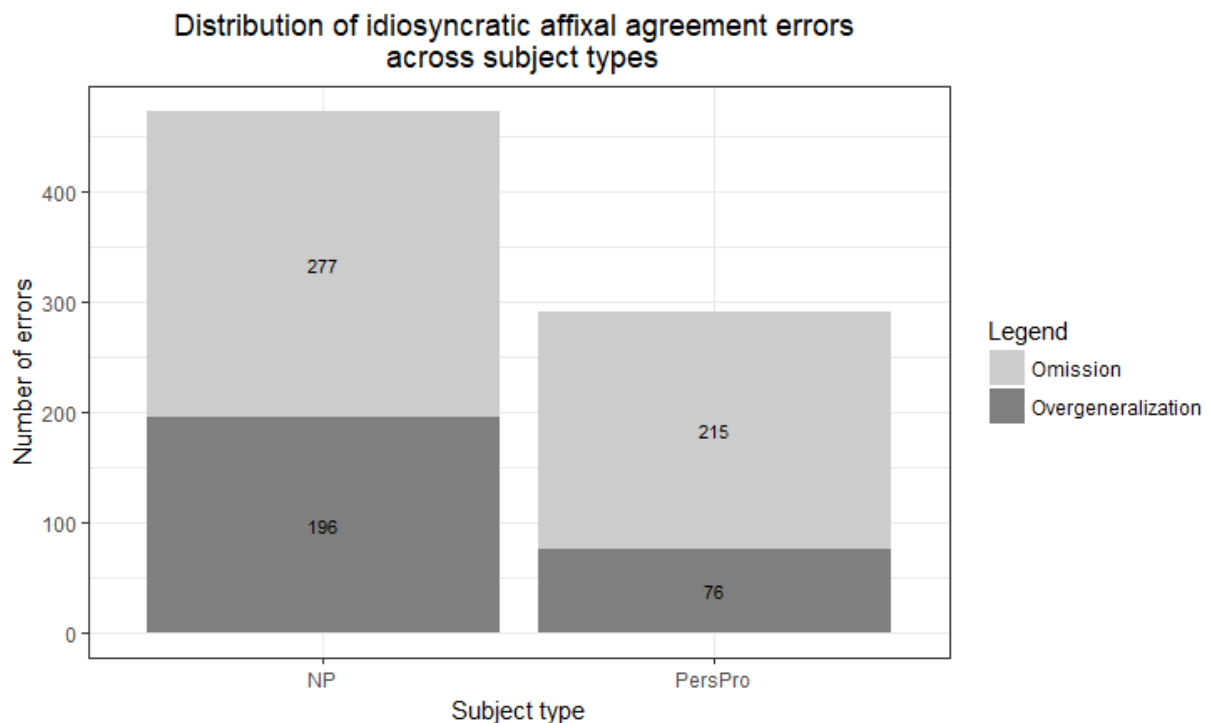
Figure 20. Distribution of idiosyncratic affixal agreement errors across NPs and personal pronouns as subjects.

Figure 20 shows that when only the idiosyncratic errors are considered, the students are more prone to make omission errors than overgeneralization errors in both subject categories. This pattern is common among many other first language groups. It seems that in situations where agreement marking is complicated by factors such as (idiosyncratic) grammatical features on the noun or processing limitations the Norwegian learners tend to use the marked verb form more often when they make an affixal agreement error. While when the situation is as clear as it can be (i.e. regular noun or personal pronoun as subject immediately followed by the finite verb), the error rates are leaning more towards the omission errors (i.e. the 'normal' learner behavior). However, an overgeneralization rate of over 40% (196 out of 473 errors with NP subjects) is still unusually high considering that most studies of L2 agreement marking find only sporadic cases of incorrect agreement marking. In addition, the studies which report overgeneralization rates of over 30% (Dröschel, 2011; Housen, 2002) do not distinguish the errors potentially triggered by the semantics or the distance between the elements which would probably decrease their numbers.

The last aspect to consider in the category of idiosyncratic errors is the distribution of suppletive agreement errors across the different subject types as illustrated in *Figure 21*:
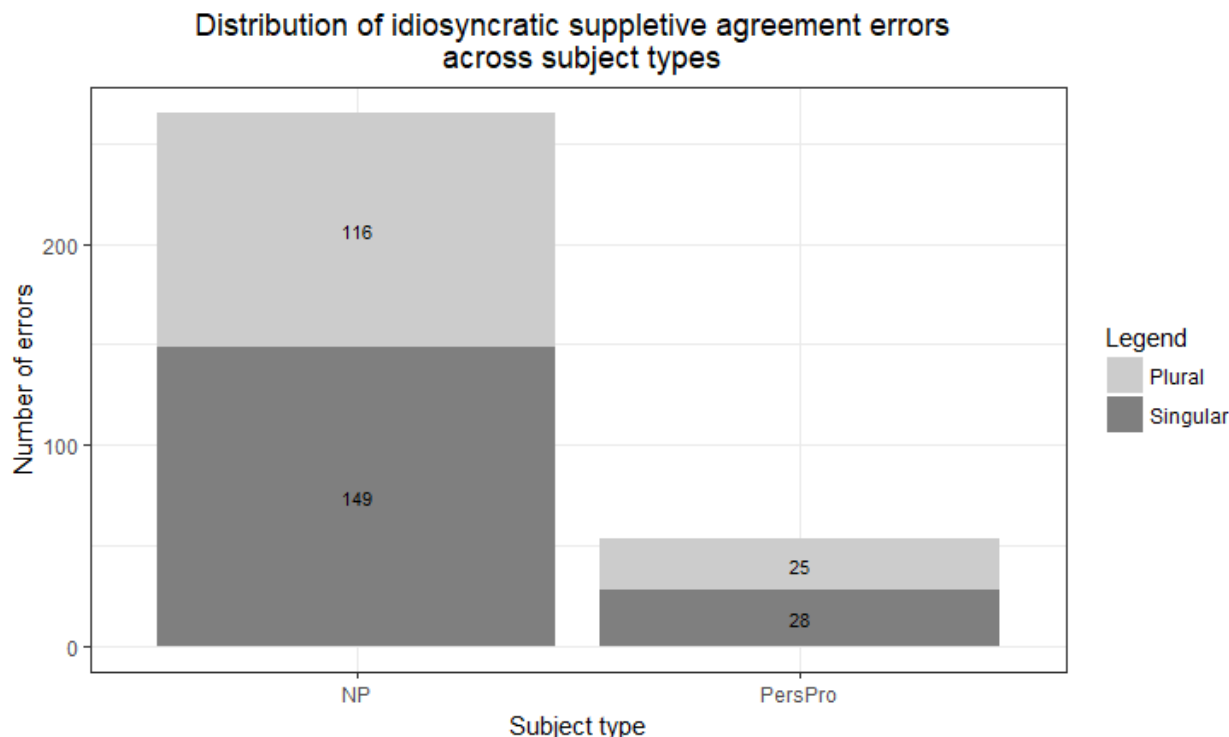
Figure 21. Distribution of idiosyncratic suppletive agreement errors across NPs and personal pronouns as subjects.

As with the overall suppletive error distribution and with the distribution of suppletive errors which are potentially distance-triggered, the singular forms of *BE* are overused when raw error counts are considered. Similarly, as with all the other categories considered, there are fewer errors after personal pronouns as subjects and the distribution of erroneous forms is more even in this sub-category of clauses than after NP subjects. However, the singular forms of *BE* are three times as frequent in the corpus as the plural forms and the distribution of singular and plural errors does not correspond to this frequency of occurrence. The plural forms are used incorrectly in 43.77% of the idiosyncratic suppletive agreement errors after NP subjects while only 24.80% of the finite *BE* forms in the corpus are plural.

5.6   *Summary of the results from Norwegian students*

Norwegian L2 learners produce subject-verb agreement at relatively stable rates even at fairly advanced stages of L2 English. There are nearly 3,000 agreement errors recorded in the corpus material of Norwegian high school students. Generally, the students who

attempted to use more complex structures (e.g. embedded clauses, longer sentences with several subordinate clauses, etc.) made more errors related to the distance between the subject and the verb, while the students who used structurally simpler language made more errors which can be related to erroneous acquisition or lack of monitoring skills. The errors are analyzed according to the type of verb (affixal or suppletive agreement) and type of subject (NP, personal pronouns, indefinite pronouns, demonstrative pronouns, expletives, and clauses as subjects). The first two categories of subjects are analyzed in more detail as these are more numerous than the others and there are fewer idiosyncratic grammatical features connected to NPs and personal pronouns than to the other categories. This section offers a summary of the results from the exploration of the Norwegian L2 corpus.

Suppletive agreement errors are less frequent in the Norwegian L2 data than affixal agreement errors. This confirms the predictions based on the production of other learner groups that suppletive agreement is acquired before affixal agreement. However, despite higher accuracy with suppletive agreement, over 35% of the recorded errors involve suppletive agreement and this category thus cannot be dismissed as fully acquired. Even though the raw numbers of errors suggest that the most common error in suppletive agreement is the overuse of *is* as a default form for both singular and plural subjects, the relative frequencies of suppletive errors, i.e. the proportion of erroneous use as compared to all use, show that it is more common for the students to overuse the plural forms *are* and *were*. It is hypothesized that there could a phonological reason for this overgeneralization. The Norwegian forms of *BE* in the present and the past tense, *er* and *var*, phonologically resemble the English *are* and *were* respectively, and the Norwegian students might thus overuse these forms based on phonological similarity.

Affixal agreement errors are common in L2 English across speakers of various L1 backgrounds. However, while the commonly reported problem in L2 English is an omission of the 3rd person marker, the Norwegian learners commonly overproduce the marker in contexts where it is not allowed in Standard English. The proportions of the omission and overgeneralization errors differ based on the type of subject involved. The majority of errors in clauses with NP subjects are overgeneralization errors while the majority of errors in clauses with personal pronouns as subjects are omission errors. It is suggested that the marked affixal form of the finite verb might be considered a default finite form for the Norwegian learners and they resort to this default form when the

74

context is complex, e.g. after complex NPs as subjects. Personal pronouns as subjects are simpler, both semantically and syntactically, than complex NPs and the learners make in total fewer errors in clauses with personal pronouns as subjects. It is also shown that the VS students make more errors in clauses with personal pronouns as subjects than the GS students. It is thus suggested that agreement marking in clauses with personal pronouns as subjects is more problematic for the less advanced learners who are then prone to the more economical omission errors in these clauses.

In addition to the combinations of subject and verb types, also the presence of elements intervening between the head of the subject phrase and the finite verb is explored. Intervening elements increase the distance between the agreeing elements and may also play a role in proximity agreement, as is attested among native speakers. However, the data from this learner population do not show evidence of high occurrence of proximity agreement. It seems that the increased distance between the agreeing elements could be part of the explanation, but only few of the errors involve intervening elements with conflicting number. The errors possibly caused by proximity agreement are more common with suppletive forms than with affixal forms. Since the learners are more accurate with the suppletive agreement in general, perhaps they behave more like native speakers within this category also when it comes to proximity effects.

For personal pronouns as subjects, the omission errors are the most frequent ones across the board. For NPs as subjects, the overgeneralization errors are dominant overall and also in distance-triggered contexts. However, in clauses with head-final subject NPs, the omission errors are more numerous. In other words, the Norwegian learners behave similarly to learners with other L1 backgrounds in syntactically and semantically simple contexts while they overproduce the marked verb forms in more complex contexts. It seems that they might use the marked 3rd person form the verb as their default form for affixal agreement in syntactically difficult contexts.

5.7 Data from Swedish learners (from Thagg Fisher, 1985)

Swedish and Norwegian are closely related languages and share, among other features, the verb inflection patterns (including *BE*). If the atypical error production of Norwegian L2 learners stems from L1 influence, Swedish L2 learners should have similar problems. Thagg Fisher's (1985) survey of agreement errors in Swedish L2 English production (see also chapter 3.2) provides enough details to allow reanalysis of the data. In this section I

thus reanalyze Thagg Fisher's data in a way that allows comparison with the Norwegian data from my corpus. However, it is important to keep in mind that the students contributing to Thagg Fisher's corpus were older (university students) and they were studying English as their major field of study. It is thus expected that they are more advanced L2 learners than the Norwegian high school students who contributed to my corpus. On the other hand, Thagg Fisher (1985, pp. 12-13) reports that language teaching in Sweden during the period of collection (1960s) was marked by the transition from the traditional grammar-translation approach towards behaviorist-based audiolingual method. Even though English was present in the society in Sweden via magazines, books, films, and pop music, it was presumably not as dominantly present as it is today. Thus, the Swedish students who contributed to Thagg Fisher's corpus might have had less naturalistic input during their L2 acquisition than the Norwegian students whose data are analyzed in this thesis.

Thagg Fisher does not report error frequencies for the individual learners. Instead, she uses error frequency per word. The Swedish students produce on average one agreement error per 481 written words in argumentative texts (Thagg Fisher, 1985, p. 69). The Norwegian students produce on average one agreement error per 147 written words. As mentioned above, this difference is most likely due to their lower age and the presumably lower proficiency. In addition, the Swedish students wrote 120-580 words each, while the Norwegian students wrote over 2000 words each, which gave them more opportunities to produce errors. Only written data from Thagg Fisher's corpus are analyzed below.

5.7.1 Verb types

Thagg Fisher distinguishes between full verbs, which always function as main verbs, and primary verbs, which can function both as main verbs and auxiliaries, i.e. DO, HAVE, and BE (1985, p. 41). However, since she provides the number of errors for each category separately, it is possible to transform her counts to match the approach chosen in my analysis, i.e. distinguish suppletive errors involving a form of BE and affixal errors with other verbs. The total error counts from the Norwegian data (from section 5.3) and the Swedish data from Thagg Fisher (1985, pp. 82-88) are summarized in Table 12 and 13:

Table 12

Summary of suppletive error counts from Swedish and Norwegian data

Error type	Nr. of errors - Swedish learners	Nr. of errors - Norwegian learners
Suppletive singular	81 (57.45%)	668 (63.86%)
Suppletive plural	60 (42.55%)	378 (36.14%)
Suppletive total	141 (100%)	1,046 (100%)

Table 13

Summary of affixal error counts from Swedish and Norwegian data

Error type	Nr. of errors - Swedish learners	Nr. of errors - Norwegian learners
Affixal overgeneralization	121 (38.66%)	983 (51.85%)
Affixal omission	192 (61.34%)	913 (48.15%)
Affixal total	313 (100%)	1,896 (100%)

The proportions of suppletive errors with singular and plural *BE* in the texts by Norwegian and Swedish learners are similar: both groups more often use singular forms incorrectly (*is* and *was*) than plural forms (*are* and *were*). However, recall from section 5.3.2 that when the relative frequency of errors to the total use of each form is calculated, the plural forms are overrepresented among the errors for the Norwegian learners. Unfortunately, Thagg Fisher's corpus is not digital, and she does not provide any frequencies of use of the individual verb forms. It is thus not possible to compare the relative frequency of errors to correct use of *BE* in the Swedish data.

The proportion of the two types of affixal errors differs between the Norwegian and Swedish learners. Unlike the Norwegian learners, the Swedish learners produce omission errors more often than overgeneralization errors. There are close to 39% of overgeneralization errors (i.e. the use of a singular verb with a plural subject) in the Swedish data, which is comparable with the proportions reported by Housen (2002, see chapter 3.3) for young Dutch and French learners in the low to higher intermediate groups (but not in the high proficiency group). Both Breiteneder (2005) and Dröschel (2011, see chapter 3.1) report lower overgeneralization rates (around 20% and 33% respectively, both for adult learners). However, it is expected that the level of proficiency in English of

the Swedish university students majoring in English should be higher than the level of the higher intermediate Dutch and French high school students. In addition, all three mentioned studies analyzed spoken data, while the Swedish and Norwegian corpora comprise written data. Spoken data tend to contain more errors than written data due to the higher requirement for automatization and spontaneity in oral production. All in all, it seems that the Swedish learners produce more overgeneralization errors in their writing than what would be expected based on their proficiency level and the mode of production.

5.7.2   Subject types

Thagg Fisher's analysis of the subjects of the erroneous clauses differs from mine (see section 5.4). Thagg Fisher first analyzes only the contiguous S-V constructions, which she defines as "subject heads which are in immediate contact with the verb" (1985, p. 38). She provides an overview of the distribution between nominal and pronominal heads combined with both plural and singular verbs, but she does not separate the errors involving *BE* and other verbs, nor does she separate different types of pronouns (e.g. personal, indefinite, etc.). She further analyses the non-contiguous S-V constructions but only in regard to the type of verb (lexical, *do, have, be*) and not in regard to the type of subject. However, with the data provided in the report, it is possible to reanalyze Thagg Fisher's data for lexical verbs and divide the error counts into categories based on the subject type (nominal or pronominal). I have thus collapsed my data into the categories available in Thagg Fisher's data for the comparison below. The data for auxiliary verbs are not available in the same detail. Auxiliaries are therefore excluded in the following. Thagg Fisher's data (1985, pp. 83-84, Table 3.4), adjusted to the terminology used in my analysis in this chapter, are reproduced as Table 14 together with the Norwegian data from my corpus collapsed into Thagg Fisher's categories:

Table 14

Summary of error counts from Swedish and Norwegian data divided by the subject type and number of the lexical verb

	Nr. of errors - Swedish learners		Nr. of errors - Norwegian learners	
	Nominal head	Pronominal head	Nominal head	Pronominal head
Singular verb (overgeneralization)	82 (52.90%)	9 (11.84%)	788 (62.00%)	195 (32.18%)
Plural verb (omission)	73 (47.10%)	67 (88.16%)	483 (38.00%)	411 (67.82%)
Total	155 (100%)	76 (100%)	1,271 (100%)	606 (100%)

As shown in section 5.6, the young Norwegian learners are more likely to produce overgeneralization errors with nominal subjects and omission errors with pronominal subjects. The Swedish learners show the same distribution of overgeneralization and omission errors across these two subject categories although the difference in frequencies with NP subjects is smaller and the difference with pronominal subjects is larger than in the Norwegian data. It is thus clear that even though the Swedish learners produce more omission errors than overgeneralization errors in total (see chapter 3.2), when the erroneous clauses are divided based on the type of subject, the same overgeneralization pattern as in the Norwegian data emerges. Unfortunately, it is not possible to conduct more detailed analysis based on Thagg Fisher's data, but it seems that at least in the category of lexical verbs, disregarding possible intervening elements, these two groups of learners behave similarly. As their L1s are closely related, it is possible that this error pattern is caused by L1 influence.

5.7.3 Possible error sources

As briefly described in chapter 3.2, Thagg Fisher uses three broad categories to discuss the possible sources of the agreement errors in her data (repeated here for convenience):

1. Difficulties inherent in the English construction, e.g. collective nouns, notional agreement, proximity agreement, coordination within the subject, etc.

2. Contrastive problems, e.g. irregular plurals, mismatch of number or countability of specific nouns between the L1 and the L2, etc.
3. Performance problems, e.g. using the unmarked form as a simplification strategy, '-s preservation', and overgeneralization of the 'one -s principle.'

Errors from the first category are occasionally produced by native speakers of English as well, especially in oral language. Errors from the second category do not normally occur among native speakers but are common among L2 learners depending on the L1/L2 combination due to transfer of semantic and syntactic properties of specific lexical items. Errors from the third category are rare among native speakers and since they are L1 independent, they should occur among all L2 learners regardless of their L1s.

Recall from chapter 2.1 that omission errors (i.e. using the unmarked form as a simplification strategy) are, indeed, commonly reported among L2 English learners from various L1 backgrounds. On the other hand, '-s preservation' and 'one -s principle' are not commonly reported as L2 English problems. Thagg Fisher explains that in some cases the plural marker on the subject noun can be reflected or preserved on the verb creating thus an overgeneralization error. Such errors are rare but reported among native speakers as 'slips of the tongue' (H. H. Clark & Clark, 1977; Fromkin, 1973). Thagg Fisher (1985, p. 45) argues that L2 learners "are likely to commit these errors more often [than native speakers] owing to insufficient automatization of the relevant L2 rule." However, -s preservation cannot explain errors which involve a pronoun or an irregular plural noun which is not marked with the plural morph -s. For such cases, Thagg Fisher suggests an overgeneralization of the 'one -s principle,' which she defines as a regular pattern in English according to which "the elements in a concord relation should be marked with at least one, but not more than one, -s" (Thagg Fisher, 1985, p. 40). However, it seems strange that such principle should be noticed by the learners and established as a hypothesis in L2 acquisition. There are several grammatical S-V combinations occurring in English which contradict such pattern, e.g. singular nouns which end with -s as part of the stem occur with singular verbs which also end with -s (the 3rd person singular marker), and plural nouns which end with the plural marker -s occur with plural verbs which also end with -s as part of their stem. There are also potential S-V combinations which follow the 'one -s principle,' but are ungrammatical in English and should not normally occur in the input, e.g. singular nouns which end with -s do not occur with plural verbs, and plural nouns which do not end with -s do not occur with singular verbs. Furthermore, modal

verbs do not follow this pattern at all. In other words, this principle seems too inconsistent in Standard English to be subconsciously noticed as a rule by the learners.

In non-contiguous S-V constructions, i.e. cases where the head of the subject noun phrase and the finite verb are not adjacent, Thagg Fisher recognizes four categories of intervening elements: adverb or postmodifying clause, postmodifying prepositional phrase, coordinated verbs with the same subject, and finite verb in a postmodifying relative clause. Unfortunately, she does not provide any overview of the types of errors (i.e. omission vs. overgeneralization) in these categories. She states that about half of the errors after postmodifying prepositional phrases can be explained by 'attraction' from a preverbal or 'anticipation' of a postverbal noun, i.e. a proximity agreement (Thagg Fisher, 1985, p. 94). She also observes that the distance between the subject and the finite verb may play a role in coordinated constructions as the most frequent error combination in such constructions involves an agreeing first verb and non-agreeing second verb (Thagg Fisher, 1985, pp. 97-98). Finally, the category of finite verb in a postmodifying relative clause is the only one where overgeneralization errors clearly dominate (65% of such errors involve a plural antecedent and a singular verb in the relative clause). However, Thagg Fisher (1985, p. 98) states that many of the antecedents can be characterized as difficult for the L2 learners as they are irregular or susceptible to notional agreement. These findings, albeit not supported by frequencies, seem to confirm the results from the Norwegian corpus regarding distance-triggered errors. Some of the errors can be ascribed to proximity agreement, but others seem to depend only on the distance between the head of the subject NP and the verb. In addition, with the increased complexity of the subject NP, the rates of overgeneralization errors rise, as if the marked verb form was the default form for the learners.

5.8 Data from Austrian learners

As mentioned in chapter 2.4.3, a sub-corpus from ICCI (International Corpus of Crosslinguistic Interlanguage) is used to compare the agreement error patterns of the Norwegian learners to the German speaking Austrian learners. German L1 speakers are chosen because, unlike Norwegian, German marks subject-verb agreement, and the verbal inflection paradigm in German is morphologically richer than the English one. German-speaking learners of English thus should not easily be led to assume that the paradigm from their L1 can be transferred into their L2. The advantage of the ICCI corpus is that

81

the learners can be matched with age (11th grade). The disadvantage is that the selected sub-corpus is rather small (68 learners, 11,342 words), which results in few agreement errors available for analysis.

There are altogether only 15 errors detected in this sub-corpus (produced by 12 students). Four of the errors involve the number marking of the subject in (87). Seven errors are omission errors as in (88). Three errors are overgeneralization errors as in (89) and one error is a case of marking on an infinitive as in (90). In addition, three of the erroneous clauses contain intervening elements, but only one of these can be labelled as proximity agreement (91):

(87)   **Many teenager** <u>have</u> a very big problem with money. (subject 41146)

(88)   **Dancing school** <u>start</u> late in the evening, so I have enough time to learn in the afternoon. (subject 41121)

(89)   …because **the bosses** <u>thinks</u> that teenagers don't know a lot about their rights. (subject 41118)

(90)   I think there are a lot of things that are better than work and that **you** <u>can</u> only <u>makes</u> as long as you are at school. (subject 41144)

(91)   **Every human** especially <u>teenagers</u>, <u>have</u> to spend money on something. (subject 41104)

The sample of (Austrian-) German L2 English is, of course, very limited. Nevertheless, it seems that it is more likely for the L1 German learners to produce omission errors than overgeneralization errors at this stage, unlike the L1 Swedish and L1 Norwegian learners. However, these results must be confirmed on a larger sample before any conclusions are drawn.

6 Discussion of the results

Chapter 5 reports on the analysis of agreement errors extracted from a corpus of young Norwegian L2 learners of English are analyzed with the aim to identify the subject and verb combinations which are most problematic for the learners. It was shown that overgeneralization errors are more common among Norwegian learners of English than what has previously been reported among learners of English with other L1s. Clauses with NP subjects which are separated from the finite verbs by some intervening elements seems to be especially vulnerable. Similar error pattern is also found in the Swedish data described by Thagg Fisher (1985) but not in the Austrian-German data extracted from the ICCI corpus. It seems that Austrian learners are not as likely to overgeneralize the agreement pattern to other persons in L2 English as Norwegian and Swedish learners. Since Norwegian and Swedish have similar patterns of verb inflections, the higher rates of overgeneralization errors could be caused by L1 influence. However, the ICCI corpus sample is too small and the general error count is low, so it is difficult to draw any firm conclusions.

There are several possible hypotheses which would explain incorrect agreement assignment in L2 English. Since a corpus analysis as a method does not facilitate an investigation of individual differences, such consideration will be put aside here, and only group-wise trends will be considered. As shown in chapter 5, Norwegian learners of English overproduce the English 3rd person singular marker. Their general accuracy in agreement marking is high (on average around 90%), but when they produce agreement errors, these are as often faulty agreement as a lack of agreement. Norwegian learners also have problems with suppletive forms (verb *BE*) which, especially in the past tense, are often incorrect. The explanations for this behavior could include, but are not limited to the following:

1. The agreement properties have not yet been acquired
2. Dissociation of syntax and morphology (i.e. agreement is acquired, but the appropriate morphological representations are not)
3. Performance errors
4. Overextension of a learned linguistic behavior due to extensive explicit teaching
5. L2 learners do not have access to UG parameters and the agreement paradigm has to be acquired in a piece-meal fashion (weak UG approach)

6. L2 learners have access to UG parameters, but the acquisition resulted in an incorrect feature reassembly (i.e. the feature bundles in the L2 English of the Norwegian learners differ from the bundles in Standard English)

The following sections investigate these possible explanations in the light of the theories described in chapter 2.

6.1 Agreement properties have not been acquired

The most straightforward explanation of a persistent error pattern in L2 production is that the relevant aspect of the language has not been acquired yet. However, this clearly is not the case for the Norwegian L2 learners who contributed to the corpus. The median accuracy rate for all contributing students is over 90% (median error rate of 8.02%) which is normally considered a ceiling performance. Even the weakest students produce accurate agreement in almost 70% of the cases. In addition, other aspects of the L2 which are associated with agreement marking and functional categories above VP are unproblematic for these learners. The Norwegian learners have no persistent problems with auxiliary raising, question formation, nominative case assignment, etc. All these properties are associated with the IP or CP and the accurate L2 structures involving elements in these categories show that the L2 learners have acquired the functional categories required for agreement marking. There is no evidence of the agreement properties missing all together from the L2 repertoire of the Norwegian learners.

6.2 Dissociation of syntax and morphology

A second possible explanation of inaccurate agreement marking could be that the learners have acquired the abstract morphosyntactic properties associated with agreement in English but fail to produce the correct forms. White (2003, p. 188) refers to dissociation of syntax and morphology reported in several longitudinal L2 studies. In other words, the learners have the correct abstract representations in their internal grammar but have mapping or retrieval problems in production. However, White (2003) specifies that such potential dissociation rarely leads to suppliance of incorrect agreement morphs but rather to omission of the marking all together (see also Lardiere, 1998b). The Norwegian data clearly show that incorrect agreement marking is common in this population, both with affixal and suppletive forms. The only option remaining within the dissociation hypothesis would then be that the Norwegian learners for some reason mapped the φ-

features onto incorrect person-number combinations. In that case, the incorrect agreement pattern should be fairly stable, at least individually, but there are no cases of students consistently marking specific person-number combinations in the same way throughout the year or even throughout one text. The trends reported in chapter 5 seem to be stronger in some contexts on the constituent level than in others regardless of specific person-number combinations.

6.3 Performance errors

According to Johansson (2008, p. 139) the agreement errors in the Norwegian section of ICLE are just "slips which could easily by corrected by the learner." The high overall accuracy rates (around 90%) also suggest that many Norwegian L2 English users have good command of agreement marking in English. However, labeling the agreement errors as a pure performance issue is problematic for at least two reasons. First, the learners in fact do not correct their errors despite the high focus on accuracy in their graded written work. And second, ungrammatical sentences with agreement errors similar to those found in the L2 production are widely accepted in a grammaticality judgement study performed by Jensen et al. (2017). In addition, if the problem was only in performance, it would be more likely that the more economical, unmarked verb form would be overused by the learners and not the marked 3rd person singular form.

Jensen et al.'s (2017) study shows that the Norwegian L2 learners of English have preference for the overgeneralized marked verb form. Jensen et al. (2017) test the judgements of the learners in four subject-verb conditions: local singular agreement, local plural agreement, long-distance singular agreement, and long-distance plural agreement. All tested verbs are lexical, subjects are NPs, and the subjects are either head-final NPs or they contain a post-head prepositional phrase. When all proficiencies are considered together, Jensen et al.'s learners incorrectly accept the ungrammatical sentences in 65-71% of the cases (and also incorrectly reject grammatical ones in 7-27% of the cases), depending on the number of the subject and the distance between the head of the subject NP and the finite verb. When only the advanced students are considered, i.e. the group which best matches the learners who contributed to my corpus study, the results of the grammaticality judgment tests copy the most frequent error patters from the corpus study.

To sum up, the advanced students reject local uninflected correct plural forms significantly more often than inflected correct singular forms. In other words, they reject

what I argue is a 'non-default' form for them even though it is unmarked in the traditional sense and it is grammatically correct. They also reject grammatical uninflected plural verbs in long-distance subject-verb combinations significantly more often than in other conditions. Furthermore, the learners accept incorrect inflected singular verb forms after plural subjects in 48% of the cases, which is a significantly higher acceptance of the overgeneralization errors than of the incorrectly uninflected plural verb forms after singular subjects. All in all, the grammaticality judgment tests show a clear preference for the inflected marked verb form whether it is grammatical or not. Since grammaticality judgement tests do not require any production, the results should reflect the competence of the learners. These results thus suggest that the agreement marking problems of the Norwegian L2 English learners are not only a performance issue.

6.4   Overextension of a learned linguistic behavior

It could be argued that the overgeneralization of the 3rd person -s could result from an overextension of a learned paradigm. If the learners were exposed to extensive explicit teaching, they could be prone to hypercorrection in their production. Schwartz (1993) argues that explicit grammar teaching cannot change the internal competence of the learners, but it can result in 'learned linguistic behavior.' However, as discussed in chapter 9.2.2 below, the English teaching in Norway is highly communicative and the explicit grammar instruction, if present, can hardly be considered extensive. It is thus unlikely that the Norwegian learners should suffer from hypercorrection in their production.

6.5   Weak UG – no access to parameters

Clahsen & Hong (1995) argue that L2 learners do not have the same access to UG properties as L1 learners. They argue for a 'weak UG,' which entails that the process of second language learning is still constrained by the principles of UG, but the L2 learners do not have access to the UG parameters. In other words, the L2 learners must acquire the parametric features of their L2 on a case-by-case basis and cannot benefit from the instant acquisition and clustering effect of parameter (re-)setting. If the parameter resetting is in fact unavailable for the L2 learners, they might be prone to pattern matching based on their L1, as Bley-Vroman (1990) suggests.

As described in chapter 2.2.2 Norwegian does not mark subject-verb agreement but marks present tense on the verb in all persons with the suffix *-(e)r*. The same suffix is also used to mark plural on masculine and feminine nouns. This superficially resembles the pattern in English where the 3rd person singular agreement marker has the same form as the regular plural marker. The Norwegian L2 learners of English could thus incorrectly assume that their L1 and L2 have the same patterns of noun and verb inflections. This assumption could be strengthened by the phonological resemblance of the forms of the verb *BE* in these two languages, as discussed in chapter 5.3.2. If L2 learners do not have access to the UG parameters, they might be prone to use pattern matching as a strategy in their learning process. When some of the overt inflection patterns seem to match between the L1 and L2, e.g. the plural marking on nouns, they might assume that other superficially similar patterns are based on the same principles in the two languages as well. Learners with L1s which do not superficially resemble the English pattern would not have grounds for the same assumptions and would thus presumably produce other types of errors.

Swedish displays the same parallels with English as Norwegian and Swedish learners make similar agreement errors as Norwegian learners. It is thus plausible that the prevalence of overgeneralization errors in the texts of Norwegian (and Swedish) learners, in contrast to learners of other L1 backgrounds, could be due to L1 influence, possibly caused by pattern matching. The Swedish data have not been specifically analyzed for suppletive agreement errors, nor have these types of errors been tested in grammaticality judgment tests in either of the two languages. If the phonologically based overgeneralization was confirmed by experimental data, it would strengthen the pattern matching hypothesis.

6.6   Incorrect feature reassembly

Another explanation why Norwegian learners seem to be treating English agreement marking differently than speakers of other L1s could lie in the feature reassembly hypothesis (Lardiere, 2008, 2009, 2012). Recall from chapter 2.3.1.1 that Lardiere proposes that the problem with abstract features in SLA is not only feature selection from the UG options but also the correct bundling of the features and their association with specific morphological representations. In her 2012 work, Lardiere describes this process as "difficulty in L2 grammatical acquisition related to the extent to which formal features

that have already been 'packaged' or assembled into certain morphemes in the L1 must be isolated and redistributed among different morpholexical items in the L2" (Lardiere, 2012, p. 113). In Norwegian, the verbal suffix *-(e)r* marks present tense, not agreement. In English there is no present tense or finiteness marker, but since the suffix *-s* marks agreement only in the present tense, it could be erroneously associated with present tense marking in the developing L2 grammar. It is plausible that the Norwegian learners could, based on their L1, assume that the English agreement marker also marks present tense or finiteness and use it as they use the present tense marker in their L1.

The feature reassembly hypothesis would explain why the Norwegian L2 English learners do not have persistent problems with other abstract features associated with the IP category. Under this hypothesis, the learners have acquired the agreement feature and associated it with the correct morphological form, but they bundled it incorrectly with present tense marking based on their L1. The same would hold for the Swedish data, but learners with L1s which mark both agreement and present tense or finiteness would not necessarily have the same problem. However, the feature reassembly hypothesis would not explain the problems with suppletive agreement in the past tense. Recall from chapter 5.3.2 that the Norwegian L2 English learners significantly overuse the plural past form *were*. English and Norwegian both mark past tense on lexical as well as non-modal auxiliary verbs. English in addition marks also agreement but only for the verb *BE* in the past tense.

English lexical verbs in present tense, expect in the 3rd person, are identical in form to the non-finite infinitival form of the same verb. This is not true for the verb *BE* which has suppletive forms in all finite contexts. If the Norwegian learners would associate the agreement marking in the present tense with finiteness and consequently overgeneralize the affixal agreement marking as if it were a tense or finiteness marker in English, they should not have similar problems in the past tense because all finite verbs are marked for tense in the past in English. It would be plausible that the learners would, at some lower level in their L2 development, use only one of the suppletive *BE* forms for all persons, marking thus only tense and not agreement. However, it would seem more likely if they would overuse the more frequent form as their default for *BE* [+past]. *Was* is almost three times as frequent as *were* in both the Corpus of Contemporary American English (Davies, 2008-) and the British National Corpus (Davies, 2004-). In the Norwegian L2 corpus *was* is also over three times as frequent as *were*. Nevertheless, the Norwegian learners overuse

the plural form *were* if they make an error in suppletive agreement in the past. It thus seems that the feature reassembly hypothesis alone cannot explain the error pattern in the suppletive agreement among the Norwegian L2 learners of English.

7 Summary of the discussion and suggestions for further research

Most generative studies focusing on acquisition of agreement in L2 assume that it is the surface morphology which is problematic, and not the abstract features of the language. The learners, for various reasons such as economy, processing limitations, stress, etc., fail to produce the required morph despite having acquired the underlying features. It is claimed that when the surface morphology is present, it is mostly correct (White, 2003). Such results are compatible with the Full Transfer Full Access Hypothesis. The learners acquire the features of their L2 in a similar fashion as their L1 using principles of UG as their starting point and setting parameters based on cues in the input. As soon as a specific parameter is (re-)set to the L2 value, the internal L2 competence should be approximating the target language and any discrepancies are explained by the gap between competence and performance. It is of course difficult to measure internal language competence, but grammaticality judgment tests or sentence matching tests (see Freedman & Forster, 1985) are often used to compare the competence of L2 learners to the native speaker norm.

The corpus data presented in chapter 5 paint a different picture of subject-verb agreement errors in L2 English than that previously reported in experimental studies. The Norwegian learners both omit the surface morphology and produce incorrect morphs in their L2 English. In addition, grammaticality judgment tests (Jensen, 2016; Jensen et al., 2017) reveal the same pattern. The overproduction of the affixal agreement marker is also attested among Swedish learners of English, and to a lower degree also among German learners. In addition, there seems to be a pattern of overuse of the form *were* in the category of suppletive agreement, which is not reported in previous studies. Several hypotheses which could explain this discrepancy have been discussed in chapter 6. Clahsen & Hong's (1995) Weak UG hypothesis and Lardiere's (2012) Feature Reassembly hypothesis are both considered plausible. The former assumes limited access to UG in L2 acquisition (only principles, not parameters), while the latter assumes full access but opens the possibility that the features which are bundled into parameters in L1 do not necessarily match the features bundled into the same parameter in L2 which can result in lengthy (or permanent) mismatch between the production of the L2 learners and native speakers of the same language.

There are several possibilities for further investigations which could shed more light on the topic. First of all, since the current study is a corpus investigation, it might be relevant to look at the results on an individual level, i.e. track the error rate development and proportion of the different error types in the production of individual students instead of in the corpus in general. Another option would be to track the development and the proportion of the error types among Norwegian L2 learners of English in a wider cross-sectional study. A combination of data from the CORYL corpus (9-16-year-olds, see chapter 2.4), from the corpus investigated in this thesis (15-16-year-olds), and from NICLE (university students, the Norwegian section of ICLE, see chapter 2.4) would give an overview of the development of agreement marking from early stages of L2 development until highly advanced learners. One of the specific questions which could be answered by such an investigation is whether the young Norwegian learners go through a stage when they only omit the agreement morphology, as is reported for learners of other L1 backgrounds, before they start overproducing the affixal marking. If that would be the case, the proposal that extensive explicit teaching may be causing hypercorrection or overextension of a learned rule would have to be reconsidered.

A second potentially interesting option would be to investigate the suppletive agreement errors further both in different corpora and experimentally. I propose in chapter 5.3.2 that there might be a phonological influence underlying the overuse of the suppletive form *were* among the Norwegian learners of English. If these errors were in fact phonologically triggered, *were* should also be overused by learners of other L1s with similar phonological properties of the verb form, such as Swedish or perhaps also German (some forms of the past tense of the verb *sein* could be considered similar to *were*), but not by learners whose L1s do not bear such similarities, e.g. Russian. Ionin & Wexler (2002) report that errors involving suppletive agreement are rare in their data from Russian learners of English. Due to the nature of the data reported in Thugg Fisher's (1985) study of Swedish learners it is not possible to calculate the relative error frequencies compared to all instances of use in the Swedish data. In the Austrian-German data from ICCI, there were no errors involving *were* registered, but this sample of L1 German learners of English is very limited. Larger corpora of data from these populations would be needed for a comparison with the Norwegian data. If the reason for overuse of *were* was phonological resemblance with the L1, this would strengthen the hypothesis of Weak UG and pattern matching as a substitute for parameter resetting in L2 acquisition.

91

In addition to looking at suppletive agreement in corpus data from speakers of other L1s than Norwegian, it would also be necessary to experimentally test the proposed phonologically triggered pattern matching. Jensen (2016) did not include any sentences with *BE* in her grammaticality judgment tests so there are currently no experimental data available which could give an indication of the internal competence of Norwegian learners of English regarding suppletive agreement. A grammaticality judgment test or sentence matching test focusing on suppletive agreement could give an indication whether the patterns described in the corpus data are representative of the competence of the L2 learners or whether they are more likely only performance errors.

A third question is addressed by White (1992) and Ionin & Wexler (2002). They argue that suppletive agreement is acquired earlier and is more accurate in L2 English than affixal agreement because of the different raising possibilities for *BE* and other verbs in English. Since *BE* raises to IP in English, it can check agreement there in contrast to the less economical affix-lowering which is necessary for lexical verbs which remain in-situ (VP). If the economy of derivation is the reason for earlier and more accurate acquisition of suppletive agreement in English, then the accuracy rates should also differ for other auxiliaries which mark agreement, i.e. *do* and *have* and their lexical counterparts. Auxiliaries are assumed to be generated above VP in English and a finite auxiliary always moves to I (Adger, 2003, pp. 177-178) while their lexical counterparts remain in-situ in VP. If corpus or experimental data would reveal that there are differences in agreement marking accuracy between *do* and *have* as auxiliaries and their lexical counterparts, this would confirm the hypothesis that "L2 learners initially associate morphological agreement with overt movement to Tense" (Ionin & Wexler, 2002, p. 116) which then explains the earlier and more accurate acquisition of suppletive agreement. If there would be no significant difference in accuracy between the auxiliaries and lexical verbs, then there would have to be an alternative explanation for the different error patterns in affixal and suppletive agreement.

Part II

8 Introduction

Part I of this thesis focused on an exploratory analysis of agreement errors in a corpus of young Norwegian learners of English. Part II describes a development of an Inverted Classroom course in English grammar based on the learner corpus data from Part I. The resulting 12-module course is tested in a didactic intervention in Norwegian schools. The aim of the intervention is to test whether the application of the Inverted Classroom Methodology in teaching explicit English grammar can decrease the incidence of subject-verb agreement errors in the students' writing.

The students who contributed to the learner corpus in Part I are used as a comparison group in the intervention. The comparison and intervention groups of students have similar backgrounds and they can be matched for schools, study programs, and English teachers. This should allow the evaluation of the Inverted Classroom Methodology as the only independent variable. The frequencies of agreement errors are assessed at three measurement points during the school year for both groups of students and the developmental trajectories in the two groups are then compared. User log data from a subgroup of the students are also reviewed to help in the evaluation of the intervention process.

Chapter nine first provides a discussion of the relevant theoretical approaches and previous research within instructed second language acquisition. In the second part of the chapter, the English language curriculum is discussed, both from the historical perspective and in its current form. The chapter also includes some comments on the role of prescriptivism in a language classroom and concludes with a discussion of the theories underlying the Inverted Classroom Methodology and previous research of the use of this method in various educational contexts.

Chapters 10 and 11 describe the intervention. Chapter 10 provides some demographic information on the intervention and comparison groups. The design and development of the Inverted Classroom course is described in detail and some issues with the implementation of the methodology in schools are also addressed. Chapter 11 analyzes the results of the intervention. The first part of the chapter describes the scores of the two groups of students separately before they are statistically analyzed in the second part of the chapter. As the results of the intervention are not significant, the potential

reasons and suggestions for improvement of the intervention methodology are discussed as well.

Chapter 12 is devoted to an evaluation of the methodology, both from the point of view of the teachers and the students. Some suggestions for improvement of the course and the intervention design in further research are also discussed. Chapter 13 provides a conclusion of the intervention part of the project.

9 Theoretical framework

This chapters addresses some of the issues of second language acquisition from the instructional perspective, i.e. how classroom activities contribute to the development of L2. Some researchers distinguish between the terms 'language learning' and 'language acquisition', and claim, following Krashen (1985), that acquisition is an unconscious process which is the result of an exposure to language while learning is a conscious effort, predominantly happening in a classroom setting, in which explicit rules are explained and practiced. These two processes are often claimed to be independent of each other and only the result of the former is seen as automated subconscious knowledge of language. However, in the current globalized world with wide-spread internet access it becomes increasingly difficult to divide the exposure to a language as English into classroom and naturalistic. In addition, with the promotion of communicative approaches to language learning and the use of authentic materials in the classroom it seems that the division between classroom exposure and naturalistic exposure becomes more and more blurred. I have therefore chosen not to make a distinction between the terms 'language acquisition' and 'language learning'.

Another formal divide in the domain of language learning is whether an L2 should be considered a second or a foreign language for the learners. Traditionally, the distinction between foreign and second language instruction has been tied to the distinction between language acquisition and language learning. A second language is acquired more or less naturally (in the society and/or in the classroom) in the country where it is spoken as L1, while a foreign language is learned only in the classroom in a country where it is not spoken (Loewen, 2015, p. 145; Nassaji & Fotos, 2011, p. 122). This distinction often has implications for the type of teaching offered in the classroom. If there are ample opportunities for genuine communication outside of the classroom, as there are in a second language setting, there is more space to focus on language accuracy in the classroom. In a foreign language setting, on the other hand, the only communicative opportunities in L2 happen in the classroom. Nassaji and Fotos (2011, p. 123) also suggest that second language classrooms often focus on cultural and pragmatic competence necessary in the L2 culture, while foreign language classrooms might avoid these topics as they would "constitute a threat to the learners' own ethnic identities and also might not be favorably received by native speakers of that [L1] culture." However, with English becoming the global lingua franca, this distinction is not as clear anymore.

English is present in many societies where it is not spoken as L1 and the culture of the English-speaking world is also heavily present in the everyday life of many ESL/EFL students (see among others Kachru, 1992). In Norway, for example, the knowledge of English is considered a basic skill (one of the three core subjects taught in school: Norwegian, math, and English) and even though it is not an official language of the country, it is spoken and accepted at almost all levels of society. Movies and TV-shows from the English-speaking world are not dubbed (only subtitled) and the Norwegian learners of English have ample opportunities to encounter English also outside of the classroom. In the following I will, therefore, refer to the English instruction in Norway as second language instruction, not foreign language.

Section 9.1 of this chapter provides an overview of different approaches to instructed second language acquisition (ISLA) and the related research findings. Section 9.2 summarizes the development of language teaching approaches and policies in Norway to provide the necessary background information for the current educational situation and the improvements suggested in the intervention. Since the current teaching approach in Norway is described as heavily communicative with little explicit grammar teaching, section 9.3 of this chapter argues for targeted use of prescriptivism in the classroom. Section 9.4, finally, introduces the teaching approach used in the intervention, the Inverted Classroom Method, together with an overview of research findings related to the use of this method in various educational settings.

9.1   Instructed second language acquisition

The Grammar-Translation Method dominated instructed second language pedagogy for an extended period of time. In the 1970s, a new pedagogical approach to instructed second language acquisition started to emerge – the communicative language teaching (CLT). It emphasized the importance of communication and focus on meaning in the classroom activities (Spada & Lightbown, 2008). This trend is also reflected in the changes of the Norwegian English language curriculum (see section 9.2). An increase in genuine communication in the language classroom has some undeniable benefits for the L2 learners, such as an increased communicative competence among L2 students or the possibility to use the L2 in practice when travelling abroad (Loewen, 2015, p. 12), which in itself can increase the motivation in learning the L2. However, this fundamental shift in focus from grammar-centred to communication-centred instruction in some cases led to

the opposite extreme: that grammar was 'considered undesirable' and teachers were "encouraged to believe that grammar instruction was old-fashioned, uninteresting, and best avoided" (Nassaji & Fotos, 2011, p. vi). Later studies have shown that the sole focus on communication and complete avoidance of grammar instruction has noticeable negative effects on accuracy in the language production. Nassaji and Fotos (2011, p. 9) in their review of research of full immersion programs conclude that "some type of focus on grammatical forms is necessary if learners are to develop high levels of accuracy in the L2." The question which remains is how much and what kind of grammar instruction should be included in the second language classroom.

9.1.1 Approaches in ISLA

In recent years, much research has been devoted to the evaluation of the merits of various approaches to classroom instruction (Norris & Ortega, 2000 provide a comprehensive summary). Even though virtually no one proposes a return to some form of heavily grammar-focussed drill-based instruction (as Long, 1988 warned against), there is disagreement in the field regarding the type and amount of explicitness which would be optimal for instructed second language learners. The following sections review some of the arguments of the main trends along the continuum of approaches in ISLA from focus on meaning to focus on forms.

9.1.1.1 Focus on meaning

One of the basic tenets of the shift from grammar-focused to communication-focused instruction is the claim that explicit grammar instruction does not change the L2 competence and therefore does not have any long-term effect on the linguistic behaviour of the learners. Only meaningful input (as in L1 acquisition) can trigger change in the inner L2 grammar system. This approach was among others heavily influenced by Krashen's (1985) Input Hypothesis which argues that language can only be acquired via comprehensible input while language teaching can only provide information about language. Krashen further claims that, even though explicit grammar knowledge cannot modify the L2 competence, it can serve as a monitor for the L2 production: helping learners notice their own errors and trigger self-correction (see also Loewen, 2015, p. 12). Since the Norwegian students are expected to be able to monitor their own production and adapt it to the form and purpose of communication as well as develop strategies in their

own language learning (see section 9.2.2), it is necessary for them to learn about the language in addition to learning the language. In other words, even under the most meaning-focused version of communicative language instruction, the learners need to be exposed to some explicit grammar teaching in order to be able to develop their learning strategies and monitor their production in situations when it is required.

9.1.1.2 Focus on form

Since the rise of communicative teaching in the 1970s, it has often been assumed that exposure to natural language is enough for the language learners to develop sufficient command of the L2. However, numerous studies show that the improved communicative abilities of the students go hand in hand with decreased accuracy of their language production (Ellis, 2011; Granger & Tribble, 1998; Loewen, 2015). Although there are undeniable advantages of communicative language learning, there is growing support in the field of SLA research for the claim that passive encounter with grammatical features in the input does not lead to acquisition for older learners. Schmidt (1990, p. 145) argues that adults[12] are able to control their attention better than children and that they "do not deliberately attend to form, especially for redundant and communicatively less important grammatical features" which can result in an incomplete acquisition of some forms. It thus seems necessary to provide the learners with some form-focused instruction if such features are to be acquired, especially if the learners are older and at a more advanced stage of language learning (Granger & Tribble, 1998; Nesselhauf, 2004, p. 140). Nassaji and Fotos (2011, p. 10) describe several classroom studies which show that "form-focused instruction and corrective feedback provided within the context of communicative program are more effective in promoting second language learning than programs which are limited to an exclusive emphasis on accuracy on the one hand or an exclusive emphasis on fluency on the other." In other words, neither extreme is the best approach.

Various approaches have been suggested to integrate focus on form into communicative classrooms. Some suggest very implicit forms of instruction, such as 'input flood' and 'input enhancement' (Sharwood Smith, 1993). Others recommend consciousness rising task or corrective feedback (Li, 2010) but preferably only when

---

[12] The learners who contributed with their texts to this study are all 15-16 years old, which, from the perspective of language development, puts them into the adult category.

communication breaks down. The strongest version of the focus on form approach includes versions of explicit grammar teaching such as mini-lessons on specific problematic topics (Nassaji & Fotos, 2011; but see Sheen, 2005 below for criticism of such grouping). The arguments to use these activities are usually connected to the assumption that, even though explicit instruction does not seem to influence the acquisitional sequences for specific features of the L2, it can help the learners progress quicker through the individual stages (Nassaji & Fotos, 2011, p. 79). Loewen (2015, p. 86) also points out that, if the explicit instruction is provided in the L2, it can also serve as input.

The problem with the degree of focus on form is that, if the attention to grammatical forms becomes too explicit, it might hinder communication and disrupt the attention to meaning. On the other hand, if the attention to forms is too implicit, the learners might not notice the linguistic features at all (Loewen, 2015, p. 58). Loewen (2015, pp. 68-69) further states that, while corrective feedback seems to have an effect on pronunciation, lexis, and semantics, it does not seem to be equally efficient in morphosyntax, not even in written activities. In other words, there are areas of L2 learning which might require even more explicit attention than usually provided in a communicative classroom, if the teaching is to have an effect.

9.1.1.3 Focus on forms

Focus on form (FoF) is often contrasted with focus on forms (FoFS) – where the former usually denotes an implicit focus on linguistic features in an otherwise communicative environment, while the latter denotes explicit explanations of grammatical forms and their function. This distinction has been used at least since Long (1988) and the most common association with the FoFS approach is the denounced grammar-drill. However, Sheen (2002, 2005) points out that, should one follow Long's definition of what constitutes a focus on form instruction (i.e. "drawing students' attention to linguistic elements as they arise incidentally in lessons"), the majority of the approaches mentioned in the previous two sections would be disqualified.

Sheen (2005) sharply criticises the dismissal of FoFS by ISLA researchers claiming that the boundaries between FoF and FoFS are not as sharp as it may seem. He classifies all instructional approaches as FoFS if they involve any explicit grammar explanations (including contrasting L1 and L2 structures), planned communicative and

non-communicative exercises which entail using specific syntactic structures, or any feedback on grammar which is not a result of a communication breakdown (Sheen, 2002). In their meta-analysis, Norris and Ortega (2000) come to the conclusion that FoF and FoFS approaches are equally effective in the classroom. Sheen (2002, 2005) criticises both their selection of studies and their categorization of the approaches. He explains that Norris and Ortega compared both the FoF approaches and the FoFS approaches to communicative approaches, but not the FoF and FoFS to each other. When FoF and FoFS approaches are compared to communicative approaches they both show strong effect on accuracy. In order to evaluate which approach is more efficient, FoF and FoFS must be compared to each other.

In his study, Sheen (2005) compares the effects of FoF and FoFS instruction implemented into otherwise communicative immersion classes (French speakers learning English). He finds that FoFS has significant effect on the increase of accuracy, including in a delayed post-test (the effect is lower, but still significant). This finding is in accord with White's (1991) study of the effects of instruction on the placement of adverbs in L2 English by French speakers. She also found that the most efficient way to change the learners' linguistic behaviour was explicit instruction including negative feedback. However, the effect was not retained in White's delayed post-test, which she interprets as a change in learned knowledge only, not in the underlying language competence.

Explicit FoFS instruction is well integrated into the Skill Acquisition Theory proposed by DeKeyser (2015). He claims that learned knowledge about language can be proceduralized and automatized via continuous communicative practice. Loewen (2015, p. 26) adds that, even while learned knowledge is not proceduralized and automatized, it can serve as a monitor, in careful production, such as prepared speech or writing, when the learners have time to revise. Without explicit knowledge, the learners lack the tools to revise their own language. Finally, Nassaji and Fotos (2011, p. 9) suggest that monitored output can serve as 'auto-input' for the implicit acquisition of the L2 and Sheen (2005, p. 298) points out that, if the erroneous production is always left uncorrected, it can also serve as input reinforcing the incorrect forms in the learner's L2 system.

As is mentioned above, many of the FoF techniques actually involve explicit grammar teaching, either during a communicative activity (e.g. consciousness rising or corrective feedback), or before/after a communicative activity (e.g. mini-lessons explaining the difficult issues (Nassaji & Fotos, 2011, p. 14)). As such, these techniques

can be seen as a modification of the typical FoFS paradigm of Present-Practice-Produce (PPP, see among others Harmer, 2007). The new (or problematic) linguistic feature is presented in an explicit mini-lesson (potentially contrasted with L1), then this feature is practiced in a controlled setting (e.g. with a fill in the gaps exercise or an elicitation task) before the students are asked to produce the forms in context (either orally or in writing). If the practice and production sessions are sufficiently meaning-focused, the learners would get both short explicit information and space to practice the new structures in a meaningful way. This approach also lends itself easily to 'inverting' as discussed below in section 9.4.

9.1.2   Focus on form(s) in a communicative classroom

In light of the discussion above, it seems that communicative instruction without any form of focus on form(s) is not the best way to help learners develop near-native L2 competence. However, it is not clear how much and what kind of form-focused instruction is the most efficient to reach this goal. Williams (1995) discusses research studying the efficiency of form-focused instruction in communicative classrooms and suggests that learners who study in communicative classrooms "are generally more successful communicators than their predecessors who received instruction focused solely on structure" (Williams, 1995, p. 12). On the other hand, she states that as a result of the lack of any explicit or implicit grammar instruction the learners "continue to make numerous errors, primarily in second language morphology and syntax, some of them rather basic" because "the emphasis on fluency and communicative success may not push the learners towards accuracy" (Williams, 1995, pp. 13-14). Some may argue that communicative competence is enough for L2 learners and that native-near attainment is not necessary. Some level of communicative competence and not necessarily near-native production seems to be the intention behind the current Norwegian English language curriculum as well (see section 9.2.2). However, if there are any expectations that the students should be able to use their L2 in a professional setting later in life, it seems ill-conceived not to equip them with ability to improve their L2 competence further (see the students' attitudes towards this question reviewed in chapters 10.1.2 and 10.2.2).

The practice regarding focus on form or corrections of form in communicative classrooms is said to be highly unclear and inconsistent (e.g. 'Good!"), leaving the learners uncertain of what the target of the teacher's comment is – form, content, or

something else (Williams, 1995, p. 13). The practice in final evaluations in Norwegian schools is governed by the centrally issued instructions for the evaluators. These instructions contain three equally weighted sections: content, text structure, and language (Norwegian Department of Education, 2012), but the teachers in the classroom can choose to focus only on some parts of the evaluation guidelines at specific times for specific purposes during the school year. This, together with the lack of focus on accuracy in the curriculum, can lead to a mis-alignment of the curriculum to the evaluation criteria, which can lead to the students being evaluated on their language accuracy in the final evaluation whether they have received feedback on their accuracy during the school year or not.

Williams (1995) repeatedly states that her suggestion is not to abandon the communicative approach to language instruction and return to some form of grammar-translation method. However, she suggests that form-based knowledge might be useful for the learners in the following situations (Williams, 1995, p. 13):

1. planning and monitoring output
2. noticing features in the input
3. noticing the gap between their own production and the target
4. speeding passage through developmental sequences
5. destabilizing fossilized forms

Points one, three, and five are especially relevant for the Norwegian students. Their non-standard agreement production seems to be fossilized and they are not able to monitor their production even in a written, accuracy demanding context (i.e. graded written assignments and tests).

In order to mediate the lack of accuracy among students who attend communicative language classes, Williams (1995, p. 15) proposes that "for forms that are infrequent in the input, and therefore perhaps, unlikely to be noticed, it may be enough simply to point out their existence and increase their presence in the input and practice," i.e. teach them implicitly, while other forms which are problematic "because of irregularities or subtle differences in usage" or those which are "largely superfluous for successful communication" might require more explicit approaches including corrective feedback. Harley (1993, p. 251) suggests that the most likely candidates for form-focused instruction

1. differ in non-obvious ways from the L1
2. are not salient because they are irregular or infrequent in the input
3. are not important for successful communication
4. are likely to me misinterpreted or mis-analyzed by learners

Agreement morphology in English clearly fulfills points two and three: the 3rd person -*s* is not salient, and it is in most cases not important for successful communication. Furthermore, I argue in chapter 6 that -*s* is likely to be misinterpreted by some learners based on their L1. Thus, agreement marking in English is a good candidate for form-focused instruction, especially if the non-standard use is already fossilized.

Williams (1995, p. 15) concludes her paper stating that FoF approaches in communicative classrooms seem to have an effect on accuracy and the most efficient form of FoF instruction actually is instruction which includes explicit rules and feedback, "particularly in eliminating errored forms" (note that Sheen would classify such approaches as FoFS). However, Williams claims that such instruction should only be used for more advanced learners and the FoF must be maintained over time if the effect is to be long-lasting. In other words, it is not sufficient to concentrate the explicit rules into one grammar session and then expect the learners to implement this explicit knowledge into the acquired language competence immediately. Focus on formal accuracy should be integrated into the communicative classroom on regular basis, until the explicit knowledge becomes integrated into the language competence of the learners, otherwise the effects of the instruction will fade over time (see White, 1991).

9.1.2.1 Error corrections and metalinguistic knowledge

Another open question is whether errors should be corrected if they are not impeding the communication and how much place such error corrections should take in a communicative classroom. The general consensus among the generative strands of SLA theories is that explicit knowledge about language cannot serve as input for the implicit acquisition processes and that explicit corrections usually do not have lasting effects on the learners' language production. On the other hand, some researchers (among others Harley, 1993; Schwartz, 1993; Sheen, 2005; Williams, 1995) claim that especially for older learners who appear to have fossilized forms in their L2, only explicit corrections and negative evidence have any effect on production. Schwartz (1993, p. 150) distinguishes between linguistic performance which stems from the unconscious language

competence and *learned linguistic behavior* (LLB) which stems from *learned linguistic knowledge* (LLK). She explains the observable changes in learners' production after negative evidence are a result of changes in the LLK, not in the underlying competence which is UG based.

From a didactic perspective though, if a certain approach can help learners to develop the metalinguistic knowledge (or LLK) necessary to monitor their own production which would then result into changed L2 production, it seems less relevant whether this change is conscious or unconscious. This approach seems to be the core of several cognitive and emergentist approaches to instruction (Ellis, 2011), including the Skill Acquisition Theory (DeKeyser, 2015). According to these theories, the repeated practice of L2 forms will eventually lead to proceduralization and automatization of the learner explicit information. In addition, Sheen (2005, p. 298) warns against approaches which pay too little attention to the incorrect forms which otherwise do not hinder communication because of the self-assuring effect such forms can have. Just as examples of correct use during explicit instruction can become input for the learners, so can repeated examples of incorrect use become auto-input for the learners and their classmates. Nassaji and Fotos (2011, p. 80) suggest that teachers should specifically prioritize frequent errors and errors which can have "stigmatizing effects on the interlocutor." It is, nevertheless, important to keep in mind that all these researchers agree that the focus on accuracy should be embedded into a rich communicative language environment.

9.1.2.2   Errors in verbal morphology

As discussed in chapter 2.3.1.3, inflectional morphology is particularly a vulnerable part of L2 acquisition. Verbal morphology in L2, which is the focal point of this thesis, shows a large degree of variability among learners in general, even at high levels of proficiency (Schwartz, 1993; White, 2003). Schwartz (1993) claims that the large difference in accuracy of inflectional affixes in spoken and written production of L2 learners confirms that learned paradigms cannot affect the internal grammar and only come to use in monitored production, because they are only part of the learned knowledge and not the L2 competence. That being said, the data discussed in this thesis are written and should be carefully monitored by the learners because they are part of their final evaluation in the subject. The fact that the learners still produce many errors in verbal inflections even in

monitored production then suggests that they have neither the implicit competence, nor the explicit knowledge necessary to produce native-near agreement. Alternatively, it is possible that they do not know how to use their explicit knowledge for monitoring their language production.

9.2 English language curriculum in Norway

Intervention studies which lead to a development of didactic theories, such as those described above, are necessarily influenced by the education policies and specific curricula in the countries or regions where the studies are performed. Despite some global trends in language teaching, e.g. the current popularity of different varieties of the communicative approach, there are potentially large differences in the specific interpretation of a given approach in different national curricula.

This section provides a description of both the current Norwegian English language curriculum and its historical development in order to offer the necessary comparison of the current methods used in Norwegian schools to the method used in the intervention. It is important to consider both the current state of the curriculum and its historical development because the attitudes towards explicit grammar instructions are tightly connected to the curriculum development and the public discussions leading to it. In addition, most active teachers have been educated under one or more of the previous versions of the curriculum when the communicative approach was not yet in the center and their attitude towards explicit grammar instruction may be colored by their own experience as learners.

9.2.1 Historical development

English instruction was established in the Norwegian school system already from the start of the general nine-year elementary school in 1959.[13] Even though the curriculum from 1960 specifies that aim with teaching English in school is to give students practical language competence, which would enable them to have oral and written contact with other parts of the world (Forsøksrådet for skoleverket, 1960, p. 204), Drew and Sørheim (2009, pp. 21, 28) describe the practice of English teaching at this time as heavily

---

[13] The English curriculum was also present in the respective curricula for the town schools and village schools prior to 1959, but here I will only focus on the development from the point of establishment of the common general elementary school for all students in Norway.

influenced by the Grammar-Translation Method. They further state that oral English was neglected and usually limited to reading out loud. Regarding grammar teaching, students were supposed to repeat and practice main features of morphology and syntax to the degree that is necessary to understand and use the language (Forsøksrådet for skoleverket, 1960, p. 206).

The next reform of the curriculum in 1974 (*Mønsterplan 1974*, or M74) was influenced by the popularity of the behaviorist-based audio-lingual method. Understanding and communicating were considered important but were often limited to heavily controlled oral repetitions. Written English was still dominant and the instruction was heavily corrective (Drew & Sørheim, 2009, pp. 28-29). Grammar teaching was supposed to focus on systematic practice anchored in speech or text and new structures were supposed to be presented in such a way that the students are from the beginning made aware of the principles behind the structure that is to be practiced (Kirke- og undervisningsdepartementet, 1974, p. 149).

There was a clear shift from the M74 curriculum to the M87 which was introduced in 1987. The M87 curriculum clearly stated that English was a 'communication subject' and stressed the need for as much authentic communication as possible in the instruction (Kirke- og undervisningsdepartementet, 1987, p. 204). The curriculum further recommended that the instruction can include both structured practice and varied and meaningful language modelling, but at the same time pointed out that students can to a large degree learn to use English language correctly also without knowing grammatical rules and terminology (Kirke- og undervisningsdepartementet, 1987, pp. 205, 208). Despite this shift in focus in the instruction, the curriculum also included a reminder that a certain degree of metalinguistic knowledge and awareness of the principles behind the grammatical structures used in communication was helpful and necessary. It was also pointed out that students should always understand the meaning of what they were learning and mechanical drilling exercises without context could easily undermine the aim of the subject (Kirke- og undervisningsdepartementet, 1987, p. 209).

The development towards communicative methods of instruction continued in the following reform of the curriculum in 1997 (*Læreplan 97*, or L97). The L97 curriculum moved the start of English language instruction from the third to the first grade, which, theoretically, should increase the chance for natural acquisition as children start being exposed to English from the age of five or six. The students were supposed to start their

contact with English through games, rhymes, songs, and children books before they moved to more 'fact-oriented' instruction (Kirke- undervisnings- og forskningsdepartementet, 1996, pp. 224, 226). Drew and Sørheim characterize this curriculum as natural language-oriented with a focus on authentic language in various genres, both oral and written (2009, pp. 29-31). There was an increased attention to the learning process itself and students were encouraged to become independent learners and users of grammar books and dictionaries (Kirke- undervisnings- og forskningsdepartementet, 1996, p. 231). In addition to the purely communicative method, the students were also supposed to master some metalinguistic terminology and learn about: different types of sentence structures, clause constituents, word classes and their morphology and function in the language (Kirke- undervisnings- og forskningsdepartementet, 1996, p. 232).

9.2.2 The current curriculum

The current curriculum for English language instruction in Norwegian schools was published in 2006 with updates in 2010 and 2013. The curriculum lists the aims and competences students should achieve during their compulsory education but does not specify the methods the teachers should use. There is also greater freedom in the choice of topics and materials as compared to the previous curricula. There is a heavy focus on communication and the development of own learning strategies. The content of the instruction is anchored in the culture of the English-speaking countries and students are supposed to learn how to work with sources, adapt their language to different genres, and use various aids in their own learning (Drew & Sørheim, 2009, pp. 41-42; Norwegian Department of Education, 2013).

The last update of the curriculum in 2013 did not introduce any major changes, but it is interesting to note that all reference to explicit metalinguistic knowledge, accuracy in language production, and even the word *grammar* disappeared from the description of the competences the students should have achieved at the end of their compulsory English education. In 2006, the competence aims in the area of language learning strategies called for the students' "use of basic terminology to describe *grammar* and text structure" and the competence aims in the area of communication required the students to "use basic *grammatical* and text structures of English orally and in writing; express himself/herself […] with some precision, fluency and coherence" (Norwegian Department of Education,

2006, p. 9, my italics). In 2013, the reference to using basic terminology to describe grammar disappeared completely and the description of grammatical knowledge within the communicative competence was changed to "use the central patterns for pronunciation, intonation, word inflection and different types of sentences in communication; express oneself fluently and coherently, suited to the purpose and situation" (Norwegian Department of Education, 2013, p. 9), without any reference to *precision*. It can be argued that the basic linguistic terminology itself is not relevant for the students, however, if the students are supposed to be independent language learners and are supposed to use dictionaries and grammar books as aids in their own work with language, they need to have some basic command of the linguistic terminology. An experienced teacher will probably incorporate such information into their instruction, but the curriculum does not require it any longer. Similarly, it might not be important whether the authors of the curriculum call the morphosyntactic features of the language *grammar* or whether they provide a list of features that should be covered (e.g. patterns of word inflection), but it is unclear whether the list is more accessible for the teachers and teacher students than the well-known and understood term *grammar*. It is, nevertheless, noteworthy that students in Norway no longer need to aim for accuracy or precision in their language production.

As is mentioned above, the Norwegian English language curriculum does not specify any methods of language teaching. It is often assumed that exposure to natural language is enough for the students to develop a sufficient command of English. Grammar is presumably taught, but the form and frequency of grammar teaching is dependent on each individual teacher. Because there are currently no systematic surveys of grammar teaching in Norwegian English classes available, it is difficult to say to what degree grammar is taught and which methods are employed. This results in high communicative competence of the Norwegian learners, but, as has been shown in the Part I of this thesis, some features of English grammar are fossilized in a non-standard form for many Norwegian learners despite more than 10 years of formal English instruction.

This is presumably not the case only in Norway, as Granger and Tribble (1998, p. 199) point out: "While the communicative approach adopted in current foreign language teaching has undeniably helped improve learners' fluency […] it has been accompanied by a loss of accuracy, especially grammatical accuracy." On the one hand, it can be argued that the native norm is not necessarily what all L2 learners need to aim for. On the

other hand, use of non-standard English, especially when the L2 users are not aware of the standard, can have undesired consequences in professional settings.

9.3 The role of prescriptivism in the classroom

The current English language curriculum in Norway has a general focus on communicative competence and practical use of English in various contexts. In the introduction, the authors of the curriculum also mention the need for English in professional settings in a globalized world and call for the ability to use English in different contexts and communicative situations. This includes also adjusting the language to different modes (written, spoken), registers (formal, informal), and accommodating different cultural norms and conventions (Norwegian Department of Education, 2013, p. 2). Prescriptivism in language instruction is often viewed negatively, as an unnecessary restriction to L2 usage. Using a non-standard variety of English is certainly legitimate in some contexts, but it can influence the reception of the message by others. Widdowson (1994, p. 381) labels standard grammar and spelling as a shibboleth which grants the entry to the good company of educated and influential speakers. The aim of the English curriculum must thus be seen in light of cultural and social belonging as well. L2 users need to be made aware of registers and the real-world consequences of their linguistic choices. In other words, prescriptive rules should not be seen as a goal, but rather as a part of the language repertoire the learners should possess.

Curzan (2014) distinguishes four institutionalized strains of prescriptivism: standardizing, stylistic, restorative, and politically responsive. With the exception of the restorative prescriptivism, i.e. the idea that the language is on the verge of decline and needs to be preserved in its current form, prescriptivism actually has a strong place in the curriculum despite it not being explicitly mentioned. Like Widdowson, Curzan also argues that the standard variety of English (and by extension, the standardizing prescriptivism) carries prestige and it is useful for social mobility (Curzan, 2014, pp. 29-30). She further argues that stylistic prescriptivism is "the equivalent of etiquette rules for how to use the language properly" and thus distinguishes "those who know better from those who don't" (Curzan, 2014, p. 33). Both the knowledge of what is prestigious and socially appropriate in which situation, and the ability to adjust the style and register to various situations are included in the introductory section of the English curriculum. Finally, the politically responsive prescriptivism, which Curzan defines as "the rules or

judgments that aim to promote inclusive, nondiscriminatory, and/or politically correct or expedient usage" (Curzan, 2014, p. 38), is clearly reflected in the section of the Norwegian English curriculum devoted to multi-cultural competence and the requirement to "take cultural norms and conventions into consideration" (Norwegian Department of Education, 2013, p. 2).

Curzan (2014, p. 75) criticizes the typical use of prescriptivism in the classroom when it is reduced to teaching grammar "to help students identify and eliminate 'errors' in their sentences." However, the awareness of prescriptive rules, and the consequences of breaking them, undeniably helps students to write more idiomatic English. Even the most proficient writers and speakers break the standards (consciously or not). The aim of prescriptivism in the classroom should therefore not be purism or persecution of all violations of any given standard but raising the awareness of the L2 users of the existing standards and what their violation entails. Awareness of styles and registers is unavoidable in professional settings, and learners need to be trained in proofreading and editing their own texts. Thus, despite the pervasively negative attitude towards prescriptive rules, the English curriculum in Norway requires prescriptive teaching in many respects. Nevertheless, the teachers are given the freedom to choose their methods and there are currently no survey data available as to what of the above mentioned is actually taught in the average classroom.

9.4   The Inverted Classroom Methodology

The term Inverted Classroom (or Flipped Classroom) is relatively new, but the concepts underlying it have been applied in higher education for several decades. The basic tenet of the method is to move some or all theoretical instruction outside the classroom and focus on problem-solving tasks or concept discussions in the classroom. A typical example would be to ask the students to study a chapter from their syllabus or a primary text sample before they come to a lecture or a seminar where the instructor builds on the reading done outside of the class and leads the students in discussions or practical work. One of the early proponents of this approach, Eric Mazur (Crouch & Mazur, 2001; Mazur, 1997), simply stopped lecturing on what the students were supposed to have read according to the syllabus and focused on problem solving and practical tasks in his teaching hours. He labeled his approach 'Peer Instruction' since the students were

working together to comprehend and apply the information they gained from their reading.

In the past couple of decades, distance learning and internet-based courses have also become more and more popular as the technology advancement makes it easier for people to access resources across the globe. The common criticism of distance learning is that the students cannot benefit from the face-to-face interaction with their instructors and peers. As a result, an attempt to get the best of both worlds resulted in the concept of Blended Learning, i.e. a combination of distance instruction and in-class participation (Information Resources Management Association, 2016). Blended Learning can involve the study of written material, as in Peer Instruction mentioned above, but more often it involves watching a video which explains some theoretical concepts before these concepts are discussed or applied during a face-to-face session. Regardless of whether this approach is called Blended Learning, Peer Instruction, Inverted or Flipped Classroom, the basic purpose is "to change classroom dynamics by using technology to present direct instruction outside of class" and devote the class time to "more interactive tasks and additional scaffolding" (Egbert, Herman, & Chang, 2014). See also Baker (2000), Lage, Platt & Treglia (2000), Bergmann & Sams (2012), and Handke, Kiesler & Wiemeyer (2013) for more discussion.

In lower education, at least in Norway, it is not very common to ask the students to study the theoretical concepts prior to class. The common methodology for introducing new concepts is a present-practice-produce method, or PPP (see among others Harmer, 2007, pp. 66-68). This approach is heavily teacher-centered as the teacher first presents the new material, be it algebraic equations or verb inflection paradigms; the students are then involved in controlled practice of the new knowledge for a limited period of time, and the individual application of the new concepts, again, be it math problems or composition writing, is assigned as homework. The problem is that when the students are in the classroom, they are mostly passive recipients and when they are supposed to apply their newly gained knowledge and could benefit from interaction and collaboration, they are left alone to complete their individual assignments at home. The Inverted Classroom approach turns this traditional methodology around and moves the theoretical explanations outside of the classroom, which consequently opens the classroom time for discussions, collaborations, and individual help when needed.

9.4.1 Theoretical background of the Inverted Classroom

The theoretical background of the Inverted Classroom Methodology is built on assumptions which stem from the work of Vygotsky (1978) and Piaget (1968). Vygotsky's 'zone of proximal development,' which limits what the learners can and cannot do by themselves and Piaget's 'constructivism,' i.e. the assumption that each learner needs to reconstruct knowledge based on the new information and their previous experience, are combined in various theories of cooperative learning (see Bishop & Verleger, 2013 for an overview). In cooperative learning, the traditional 'transmittal model,' where the teacher transmits knowledge from their head into the learners' heads, is replaced by a constructivist model, where the learners interact with the new information to construct knowledge and the teacher moves from being the 'sage on the stage' to the 'guide on the side' (King, 1993). In other words, the lower levels of Bloom's Taxonomy for the cognitive domain (see Anderson & Krathwohl, 2001, also illustrated in Figure 22 below) are activated outside of the class, while in the classroom the instructor can focus on the higher level activities, e.g. analysis and application (Gilboy, Heinerichs, & Pazzaglia, 2015, p. 110; Kiesler, 2013, pp. 90-91):
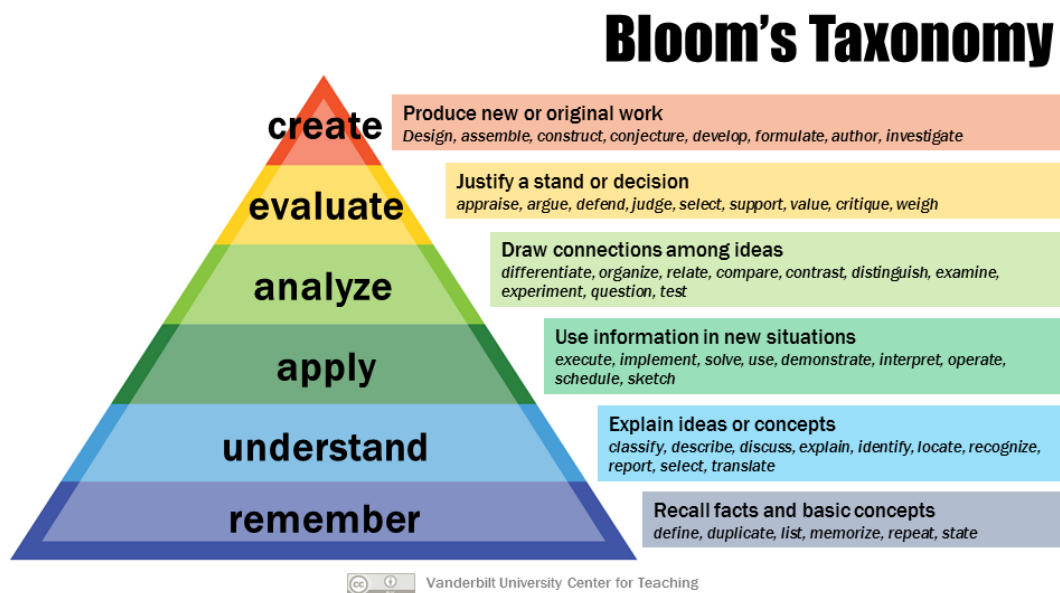
Figure 22. The revised version of Bloom's Taxonomy. Picture credit: Vanderbilt University Center for Teaching (2016).

It is important to note that the Inverted Classroom Method does not simply replace traditional in-class instruction with videos. What is essential is what happens in the

classroom. As the teacher-centered instruction is removed, there should be more time and focus on learner-centered activities (Spannagel & Spannagel, 2013, p. 113). Abeysekera & Dawson (2015, p. 4) further stress that the success of this approach relies on the students doing their homework and coming to class prepared, which can be a problem in some student groups. However, they also claim that the changed classroom dynamics should increase both the extrinsic and intrinsic motivation of the students by satisfying their need for competence, autonomy and relatedness: "Students feel more competent when they are active participants in the creation and dissemination of knowledge than when they are passive recipients of knowledge dictated by an instructor, as done through traditional lectures" (Abeysekera & Dawson, 2015, p. 7). In addition, the Inverted Classroom methodology can be used to tailor the instruction to the individual students, and students should be encouraged to pause and rewind when needed (Abeysekera & Dawson, 2015, p. 9). Self-pacing through the theoretical instruction should also help them manage the cognitive load when necessary (R. C. Clark, Nguyen, & Sweller, 2006). Abeysekera & Dawson's description of the Inverted Classroom methodology is illustrated in *Figure 23:*
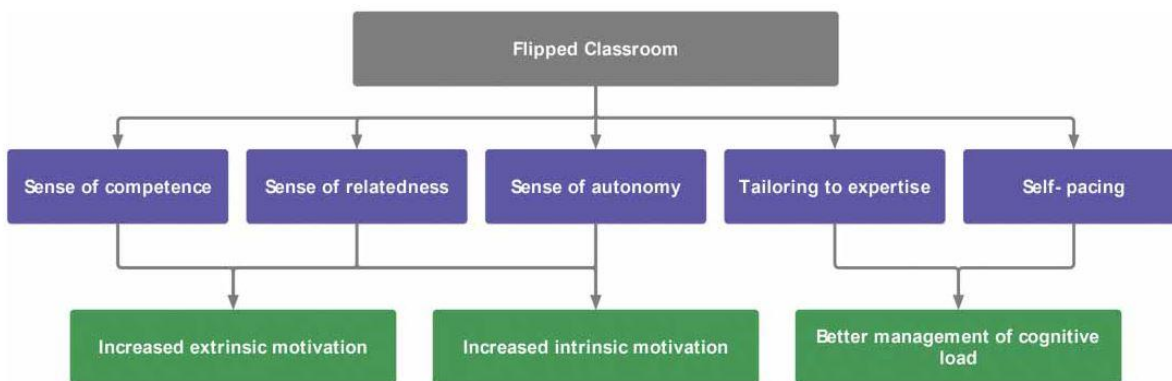
Figure 23. Rationale behind the Inverted Classroom methodology. Reproduced from Abeysekera & Dawson (2015, p. 10).

9.4.2 Previous research on the Inverted Classroom Method

Traditionally, the Inverted Classroom Method has been applied mainly in the so-called STEM disciplines (science, technology, engineering, and mathematics) because of the

theory-heavy nature of these fields of study (among others Bergmann & Sams, 2012 - high school chemistry; Lai & Hwang, 2016 - elementary school math; Roach, 2014 - undergraduate economics; Sørensen, 2016 - undergraduate IT; Steen, 2013 - high school math; Strayer, 2012 - undergraduate statistics). However, there are some examples of studies which tested the Inverted Classroom methodology in other subjects, such as physical education (Østerlie, 2018), Chinese as a foreign language (Egbert et al., 2014), English as a foreign language (Lee & Wallace, 2018; Leis, Tohei, & Cooke, 2015), English linguistics (Handke et al., 2013), and English argumentative writing (Prodoehl, 2016). Some of these studies are reviewed below.

Many of the studies in STEM disciplines report improved results in the form of final grades (Bergmann & Sams, 2012; Sørensen, 2016). Others focus on student satisfaction and report that students generally consider the inverted method helpful and the classes more interactive than other classes (Roach, 2014). The instructors also report higher student engagement (Bergmann & Sams, 2012; Sørensen, 2016), but some students struggle with the adjustment to the new classroom dynamics and with connecting the online and in-class parts of the course (Strayer, 2012). Studies involving the Inverted Classroom in language classes report improvement in text length and complexity (Leis et al., 2015), final grades (Lee & Wallace, 2018), and increased time for language practice (Egbert et al., 2014). However, students also report frustration with managing the unfamiliar technology (Prodoehl, 2016) and with the increased out-of-class workload (Egbert et al., 2014).

Much of the systematic research involving some form of Inverted Classroom approach focuses on higher education. One of the reasons for this trend might be the lower availability of technology, e.g. personal computers with internet connection, in the homes of students enrolled in lower levels of education. Prodoehl (2016, p. 76) points out that high school students can benefit even more than university students from the advantages of individualized and self-paced instruction the Inverted Classroom Method potentially offers. Luckily, this model is easy to implement in Norwegian high schools since all high school students receive a laptop from their institution (Fylkenes informasjonstjeneste for søkere til videregående opplæring, 2014; Opplæringslova, 1998) and 97-100% of the households in Norway have internet connection (Statistics Norway, 2017a). All students should thus be able to access the out-of-class study materials regardless of their home situation.

Two of the studies mentioned above tested the Inverted Classroom method in Norwegian lower education context and one in Taiwanese lower education. Østerlie (2018), who designed an intervention in high school physical education classes, reports increased motivation to actively participate in the classes among high school students after the Inverted Classroom was implemented to introduce game rules and some basic principles of appropriate exercise prior to class. The effect on motivation was higher among girls, who statistically show lower motivation and achieve lower grades than boys in PE in Norway (Østerlie, 2018, p. 5). Steen (2013), who applied the Inverted Classroom methodology in high school math classes, reports satisfaction with the increased predictability of homework assignments among the students (theory videos instead of math problems). Steen's students also report that the ability to rewind the videos helped them comprehend the new material. Lai & Hwang (2016) focused on self-regulation in their study of Inverted Classroom in elementary school math in Taiwan. They did not use videos but assigned readings outside of class followed by quizzes to check comprehension and they compared a group which received assistance in self-regulation of the out-of-class assignments to a group which only received Inverted Classroom instruction without assistance in self-regulation. Their findings confirm the importance of self-pacing (see Abeysekera & Dawson, 2015) for the success of the Inverted Classroom method. It seems that younger learners need extra help organizing their own studies and it is not enough to give them the tools without clear instructions how to use them. Such findings do not surface in the research focusing on higher education, perhaps due to higher maturity of the subjects.

10 Intervention design

As described in chapter 9.2.2, the curriculum for English language in Norwegian schools is heavily communicative and the teachers have freedom to use any methods and materials they see fit to help the students reach the competence aims of the curriculum. One of the results of such an approach is that students often have varying degrees of explicit knowledge of formal English grammar when they fulfill their obligatory schooling. The requirements of the curriculum in this regard are vague and their interpretation and following depth and frequency of explicit grammar instruction depends on the individual teachers. In order to equip teachers with a tool which would allow them to provide targeted explicit grammar instruction an Inverted Classroom course in English grammar was developed.

The Inverted Classroom Method is a relatively recent didactic approach and it has been tested mainly in higher education. The aim in the second part of this project is to test whether the use of the Inverted Classroom course in English grammar could decrease the number of subject-verb agreement errors in writing of Norwegian high school students. The frequency of subject-verb agreement errors from the corpus of the comparison groups (described in Part I) is compared to the frequency of the same type of errors in a corpus collected from the students in an intervention group. The students in the intervention group have been exposed to the Inverted Classroom course in English grammar developed by the author (see section 10.6). Based on the background information collected, the data from the two corpora can be broken down into smaller subgroups, which can then be matched for specific schools, types of study programs and the English teachers involved.

In this chapter, first, the background information of both the comparison group and the intervention group will be described in sections 10.1 – 10.3. Then, I will move onto the description of the data from the subjects who could be matched between the two groups based on their backgrounds and describe the measurement and statistical procedures in section 10.4. Section 10.5 is devoted to the discussion of some alternative methodology options, and finally, section 10.6 describes the teaching intervention. The results from the intervention are discussed in chapter 11.

10.1 Comparison group

There are 199 students included in the comparison group. The students come from three high schools in the Agder region, ten different classes (five different tracks: General

Studies; Healthcare, Childhood and Youth Development; Media and Communication; Technical and Industrial Production; and Service and Transport), and they were taught by ten different teachers. All these students contributed to the corpus analyzed in Part I of this thesis. However, for the evaluation of their progress through the school year and comparison with the intervention group, only the students who contributed with at least three texts are included in the subjects described below. The following sections describe the demographics, the experience with additional languages, and the attitudes to language learning of the 134 students who contributed with at least three texts to the comparison group corpus.

10.1.1 Demographic data[14]

There are 80 females, 52 males, and two students with an unassigned gender in the comparison group. 117 claim to be monolingual Norwegian (i.e. no other languages are used at home), seven use Norwegian and an additional language at home (Vietnamese; Spanish; Turkish and Arabic; English and Danish; and three students use English in addition to Norwegian), six students only use languages other than Norwegian at home (Somali; Kurdish; Chechen; Arabic and Tigrinya; and two students use only Polish). Four students did not answer this question. Four students report having a learning-related diagnosis (such as ADHD or dyslexia; specific information regarding the diagnoses was not collected).

Five students did not take their entire education in Norway (and four did not answer this question). These students started their education in Norway in the third, sixth, eighth and two in the tenth grade. Naturally, there is an overlap between the multilingual group and the students who studied partially abroad. Altogether, there are 14 students who were either bilingual prior to learning English in school or did not take their entire education in Norway. Out of these 14 students, six have total error scores above the median for the whole group (above 7.97%) and eight score below the median. In other words, their total error scores are comparable to their monolingual peers.

All four students who report having learning-related diagnoses make more agreement errors than the median for the whole group. Two of these students also belong to the previous group, i.e. they are both speaking additional languages at home, and one

---

[14] See Appendix 1 for the full questionnaire in Norwegian.

of them started attending Norwegian school in the sixth grade. It seems that learning-related diagnoses can be contributing to the high error scores of these students. However, the sample size is too small to draw any conclusions. In addition, the confounding of bilingualism, education from abroad, and learning-related disorders for half of the students in question makes it impossible to judge which of these factors are relevant. On the other hand, the purpose of the intervention is to test whether the given methodological approach could help the student population improve their agreement accuracy, so it is relevant to include also the students who perhaps have different starting conditions than a typical monolingual Norwegian speaker without any learning-related diagnoses.

10.1.2 Use of additional languages and attitudes towards language learning

In addition to the basic background data described above, the students also answered some questions about their language use and attitudes towards language learning (47 of the 134 students did not respond to this part of the questionnaire). 48 students report using a language other than Norwegian in their daily life outside of school. 37 of these use English, and the other students use the languages they reported using at home (two of those who speak Norwegian at home reported using Hindu and Punjabi, and Kurdish in their daily lives outside of home and school). When they use English outside of school, they use it equally often in the online environment, e.g. gaming, social media, etc. (17 students), and in offline communication (17 students). Nine students report using English for travel, culture, and hobbies.

62 of the students think that the foreign language(s) they know will be important or very important in their future work, 20 are not sure, and five think that languages will have little importance in their future work. For 72 of the students, it is important or very important to become independent users of the foreign language they are learning, nine are uncertain, and six do not consider it important. For 63 of the students, it is important or very important to speak grammatically correctly, 14 are uncertain, and for 10 of the students, grammatical correctness of the oral language is little or not important at all. For 66 of the students, it is important or very important that they write grammatically correctly, 15 are uncertain, and for six students grammatical correctness of the written language is little or not important at all. Out of the students who are not particularly interested in grammatical correctness, three do not consider either oral or written correctness important, seven care more about the correctness of their written production,

while two care more about the grammatical correctness of their spoken language (some of these students are 'uncertain' about the other mode). All in all, most of the students consider foreign languages important, and they also desire to be able to use their L2 independently and grammatically correctly.

10.2 Intervention group

Due to unforeseen circumstances, such as leaves of absence and changes in schedules, not all teacher-track combinations used in the comparison group were available for the intervention. 119 students from the same three schools, but only representing four tracks (General Studies; Healthcare, Childhood and Youth Development; Media and Communication; and Sports and Physical Education), and six different classes participated in the intervention. 100 of the students in the intervention group submitted at least three texts during the school year so their progress could be tracked. In the following sections, I present the demographic data, experience with additional languages, and attitudes towards language learning of the 100 students from the intervention group who contributed with at least three texts.

10.2.1 Demographic data[15]

There are 58 females and 42 males in the intervention group. 85 claim to be monolingual Norwegian (no other languages used at home), eight use Norwegian and an additional language at home (French, Romanian, Vietnamese, Persian, German, and three say they use English), and seven students only use a language other than Norwegian at home (Persian, Amharic, Romanian, and two each for Dutch and Polish). Only two students report having a learning-related diagnosis. Nine of the students did not take their entire education in Norway (three started in the second grade, three in the third, and one each in the fourth, fifth, and eighth grade). There are altogether 16 students who were either bilingual prior to starting to learn English or did not take their entire education in Norway. The error scores of these students are actually marginally better than their monolingual peers (only five of them have error scores higher than the median for the whole group). One of the students with learning-related diagnosis scores higher than the median, the other one scores lower. It seems that neither bilingualism, nor learning-related diagnoses

---

[15] The questionnaire used in the intervention group is identical to the one used in the comparison group.

have an effect on the number of agreement errors in this group, but, as with the comparison group, the sample size is too small (16 bilinguals and two with learning-related diagnoses) to offer any conclusions.

10.2.2 Use of additional languages and attitudes towards language learning

As the comparison group students, the intervention group students also answered questions about their language use and their attitudes towards language learning. 48 of the 100 students in the intervention group report using a language other than Norwegian in their daily life outside of school, 32 of these use English, and the others use the languages they reported using at home (two of those who speak Norwegian at home reported using Danish and Dutch with distant relatives). When they use English outside of school, they use it both offline (16 students), and online for gaming, social media, etc. (14 students). 13 students report using English for travel, culture, and hobbies.

67 of the students consider their additional language(s) important or very important for their future work, 32 are not sure, and only one thinks that languages will have little importance in their future work. For 89 of the students, it is important or very important to become independent users of the L2, nine are uncertain, and two do not consider it important. 62 of the students state that it is important or very important for them to speak grammatically correctly, 15 are uncertain, and 23 of the students consider grammatical correctness of the oral language little important or not important at all. For 69 of the students, it is important or very important that they write grammatically correctly, 13 are uncertain, and 18 students consider grammatical correctness of the written language little important or not important at all. Out of the students who consider grammatical correctness unimportant, 13 do not consider either oral or written correctness important, ten care more about the correctness of their written production, while five care more about their spoken language. Thus, the attitudes towards language learning and the importance of grammatical correctness in L2 are very similar in both student groups.

10.3 Representativeness of the student backgrounds in the samples

Since the data were collected from a limited geographical area in Norway, questions could be raised regarding the representativeness of this sample in relation to the whole population of high school students in the country. Statistics Norway database on Upper secondary education (2017b) allows data extraction on several of the variables described

above. In order to check whether the students involved in the intervention could be considered a representative sample of all high school students in Norway regarding their background, I have extracted data regarding study tracks, gender, and immigrant background. All data are aggregated over the four years the data collection took place (2013-2017) and converted to an average value for this period. Health related data (e.g. the incidence of learning-related diagnoses among students) are not publicly available. However, since there are only six students in both samples who reported having a learning-related diagnosis and these students' scores are distributed fairly evenly around the average, the existence of such diagnoses is not considered an influential factor in the following.

The distribution of students between the two study tracks (general studies and vocational studies) in Norway is about 60.5% in GS and 39.5% in VS (Statistics Norway, 2017b, table 09378). The distribution of the subjects in the comparison and intervention groups is more skewed towards the GS track as compared to the national average. There are only 14.2% and 20% VS students in the comparison and intervention groups respectively. This discrepancy is caused by the general aim at the GS students in the development of the first version of the Inverted Classroom course. Only some teachers were willing to try the course in VS classes because the VS students have fewer teaching hours and generally less theoretical instruction. A second version of the course is planned to target the VS groups specifically.

Male and female students are represented equally in secondary education in Norway (50.6% females, 49.4% males). However, the proportions of the genders differ in different study tracks. I have extracted the relevant study tracks from table 06382 (Statistics Norway, 2017b) and calculated the proportional representations of males and females with respect to the compositions of the comparison and intervention groups. The proportion of female students in the tracks represented in my samples on the national level is 56%. There are 60.6% and 58% female students in the comparison and intervention groups respectively. This gender distribution is thus close to representative of the general distribution of gender among high school students in the specific study tracks in Norway.

The proportion of students with additional home languages is usually closely connected to the proportion of students with immigrant backgrounds. The questionnaires do not include questions about ethnicity or nationality, but it seems reasonable to assume

that additional home languages are for most students connected to immigrant background either in the parent generation or in the generation of the students themselves. The proportion of high school students with immigrant background in Norway is 15.3% (Statistics Norway, 2017b, table 09236, both first and second generation immigrants are included in the count). There are 10% and 15% students with additional home languages in the comparison and intervention groups respectively. This sample is thus representative of the general distribution of students with additional home languages in Norway.

10.4 Intervention data

The data material in Part II consists of two sub-corpora: the texts of the control group of first-year high school students collected without any intervention and the texts of other first-year high school students who, in addition to their ordinary classes, received inverted classroom instruction. Each sub-corpus is screened for subject-verb agreement errors which are recorded in two separate samples (see chapter 4.3.1 for details of the procedure). Each student is given a score based on the number of agreement errors the student produced in each of the submitted texts throughout the school year. The error rate is calculated as the amount of detected subject-verb agreement errors divided by the number of potential occasions to mark overt agreement in each text.[16] This is necessary in order to accurately measure the ability of the students to use the correct inflection (zero, -*s*, or suppletive) in each context. Otherwise, texts which are written in the past tense (and thus contain fewer occasions to overtly mark agreement) could score much lower on errors than texts which are written in the present tense, regardless of the student's competence. The error scores thus measure how accurately the students are able to use the correct verb forms to mark agreement. Potential occasion analysis is used to capture both missing and illicitly supplied morphs as opposed to obligatory occasion analysis which captures only the production rates of the overt morphs in obligatory contexts (see Thewissen, 2015, pp. 45-48, 143-144).

10.4.1 Measurement points for the individual students

The texts of each student are divided into three measurement points which are used to track his or her progress throughout the school year: *Fall, Midterm,* and *Spring*. Fall

---

[16] English marks SV agreement overtly in 7 contexts: lexical verbs in the present, auxiliary *do* in the present, auxiliary *have* in the present, and auxiliary and copula *be* both in the present and in the past.

includes all texts written from August to November; Midterm includes all texts written from December to February; and Spring includes all texts written from March to May. Each teacher autonomously decides when his or her class submits written assignments and tests, but each class typically writes at least two whole-day school exams in English during one school year. The texts were collected without any influence on the contents or the frequency of writing. This results in different dates and different amounts of written material from each of the classes. The division into three arbitrary points of measurement is thus necessary to compare the data between different classes and schools. The majority of the subjects contributed with only one text for each measurement point. For those who submitted more texts within the time frames of the individual measurement points, all texts were considered as one within each point. Some subjects, especially from the vocational groups, do not have recorded measurements for all three points. These were excluded from the statistical tests.

If overall scores were used to measure the total number of errors made by each student during the whole school year, the difference between a weak student with steep improvement and a strong student making little progress could be blurred. For this reason, the error scores are tracked for each individual student at three measurement points for the school year. These three measurement points for each student are used to investigate the relative developmental trajectory for each student and for groups of students (see sections 11.2.1.3 and 11.2.2.3). The expected result is a steeper learning curve, i.e. fewer errors towards the end of the school year, for the intervention group than for the comparison group.

The students from both the comparison group and the intervention group can be divided into subgroups and matched for their school, the type of program, and their English teacher. Each combination of variables can then be tested separately to investigate whether the inverted classroom instruction had any significant effect on their agreement marking accuracy. As many of the external variables, such as the school, the teacher, and the type of program are controlled for, and the demographics of the two groups are comparable, the results of the group-wise statistical tests should reflect the effects of the inverted classroom instruction on the agreement marking accuracy.

10.4.2 Statistical preliminaries for the experimental data

Some common concerns with experimental data involve the potential invalidity of the experiment due to problems in the experimental design, lack of a control group, or problems with the type of variables in the statistical calculations. To avoid these problems the following experimental design is proposed:

- Null hypothesis: all the students involved in the project belong to the same population (of first-year high school students attending the compulsory English course) and will thus not show any statistically significant difference in the number of agreement errors they produce at each point of measurement within one school year beyond the general progress curve.
- Dependent variable: error scores calculated as the total number of clauses with incorrect agreement divided by the total number of clauses in which agreement can be overtly marked on the verb in English in each submitted text (see section 10.4).
- Independent variable: introduction of the Inverted Classroom course.
- Controlled variables: school, school year, study track, teacher, number of hours of instruction per week, study program.

The students in both groups are selected from the same schools, the same programs, and taught by the same teachers, but during two different years. The high school students in year 1 (no intervention) serve as the control group for the high school students in year 2 (using the inverted classroom course). There should be no contamination of the results between the two groups because the teaching materials were not created before the collection of the data from the comparison group was finished. The materials were thus not available to the teachers in the first year and were only used in the second year, in addition to the normal instruction. The only independent variable should, therefore, be the teaching method.

10.4.3 Linguistic and non-linguistic factors

Despite the control of several external factors, there are other linguistic and non-linguistic variables to be considered. Some background data from the subjects are collected, such as age, gender, (non-)existence of diagnosed learning-related disabilities, and the amount of time spent living outside of Norway. However, the influence of factors such as home languages, learning disabilities, or socio-economic status on the effect of the instruction is not investigated in this project in detail. The intention with the testing of the Inverted

Classroom Method is to introduce an alternative methodology for grammar teaching in Norwegian high schools. Should such a model be implemented, it needs to be beneficial to the majority of the students enrolled in Norwegian schools, regardless of their background or predispositions. Should it not be beneficial to the majority of the students, targeted use for specific groups can be investigated in a follow-up study.

The grammatical accuracy of the students can also be influenced by their general motivation, their attitudes towards the importance of foreign languages, and the importance of formal correctness as opposed to 'getting the message across.' In order to understand the relevance of these factors for the student population, the background questionnaires also include questions investigating the students' attitudes towards the importance of foreign languages in professional and personal contexts, and the importance of formal correctness in speech and writing. Even though these data are not used in the analysis in this thesis, they are available for follow-up studies.

10.5 Alternative methodology options

Research within linguistics has a long qualitative tradition. Many of the core theories build on qualitative descriptions of certain language phenomena and their possible theoretical explanations (e.g. much of Chomsky's work). However, with the current development in corpus research, quantitative approaches have become increasingly common in the field. The choice to apply a quantitative methodology to the main variable(s) aims at capturing patterns which are frequent in the data, and not only the interesting exceptions to the rules. Since the assessment of the student texts only focuses on subject-verb agreement errors and uses error scores as the only measurement of proficiency, it can be characterized as a discreet-point objective assessment (Leclercq & Edmonds, 2014).

The assessment approach taken in this intervention has its weaknesses. The lack of agreement errors does not necessarily reflect fluency in the language, nor does it in any way reflect the communicative competence of the writer. On the other hand, students are constantly evaluated on the accuracy in their written production (the three areas of evaluation included in the English language curriculum in Norway are content, structure, and grammatical accuracy). Therefore, one of the goals of this project is to give the teachers new tools to improve grammar instruction and thus equip the students with some explicit grammar knowledge which they could use to monitor and improve their own

written output. In the following, three alternative methodology options for a study of grammatical accuracy development are described and evaluated.

10.5.1 Grades as a measurement of progress

A recent Ph.D. dissertation at the University of Agder has also focused on the development of selected grammatical features in English in the writing of Norwegian high school students (Horverak, 2016a). Horverak chose to evaluate the progress of her subjects by using a traditional grading approach with a series of specific criteria applied by three graders. However, as grading is at best a partially subjective matter, it is unsuitable for a quantitative analysis of the students' linguistic development. This view is supported by many, among others Chalmers, who states that "[i]t should be no news to anyone that the perceptual judgments of individuals can be unreliable for a range of reasons. The challenge, in science, is to arrange the observable situations in such a way that the reliance on such judgments is minimized if not eliminated" (1999, p. 21). Even the use of several graders with precise grading instructions does not guarantee the same result from different graders. According to Horverak's study, the inter-grader correlation was as low as 0.66 for some evaluated aspects (average of 0.76 for the final grades) (Horverak, 2016b, p. 76). Thus, in almost one third of the cases, the graders did not agree. The use of teachers from different schools poses even more problems as graders working at the same institution can be biased by the local standards used in that particular colleague group, which might differ from the standards at other institutions.

In addition to the problem with the subjective nature of grading, the teachers, and even the external graders (when used), do not always have more than one year of university level education in the subject they teach and grade. They also work under enormous time pressure when grading. Some of the material collected for this study includes teachers' comments and corrections. There are several instances of ignored agreement errors and even correct student language incorrectly marked as errors in the corrected texts. Another concern with using grades as the measurement of the improvement of grammatical accuracy is that grades always combine several measurements, which makes it difficult to disentangle the ingredients of the final grade (content, structure, accuracy of the language, creativity in the text, accuracy of the answers as related to the assignment, etc.). Grades can reflect achievement of the educational goals, but they do not directly measure accuracy in language.

10.5.2 Evaluation of spoken language

Another option would be a study of spoken production. Written language differs from spoken language in several respects and the choice to focus on the written language in this project is not incidental. Spoken language often contains significantly more errors and slips, as the speakers do not usually have time to rehearse and correct their utterances. In written production, and especially in the case of assignments and exams that are submitted for a formal evaluation, a proof-reading process and moderation should be involved before submitting the finished text. Evaluating the subject-verb errors in spoken production would perhaps generate larger numbers of errors, but it would bring no insights into the problems at hand, such as the lack of explicit grammar knowledge, the inability to moderate one's production, and to adapt it to the medium and purpose of communication (which are some of the competence aims in the English subject curriculum (Norwegian Department of Education, 2013)).

10.5.3 Longitudinal case study

A language learning study could also be performed longitudinally, e.g. by studying the development of one or several students from elementary school to university level. Such an approach could uncover the developmental trajectories behind the recurring errors. However, longitudinal case studies of this sort have several disadvantages. Apart from the time concern and the possible subject drop out, a longitudinal case study might not be generalizable and might be prone to misinterpretations. According to Franklin, "[t]here is generally much less error in higher-level data [i.e. aggregate data] because individual-level error is averaged out during the process of aggregation" (Franklin, 2008, p. 248). By conducting a cross-sectional study with three measuring points for each subject across almost 200 subjects, I hope to achieve an acceptable level of generalizability and, at the same time, the possibility to track the development of the individual students over a period of one school year.

10.6 Development of the Inverted Classroom course for English grammar

The Inverted Classroom course in English grammar was developed on the Campus Inkrement platform (Thue, 2016) which contains mainly courses in middle and high school math, but opens also for self-made courses in other subjects. The basic version of

the platform is publicly accessible and free of charge. During the intervention, the English grammar course was available only to the invited teachers and their students. After the evaluation of the results, the course will be made public and will thus be available to all interested teachers in Norway.[17] Several aspects from the previous studies in Inverted Classroom implementation among various learner populations were considered in the preparation of the English course (see chapter 9.4.2). In addition, the experience from other Inverted Classroom developers was taken into consideration (see Handke et al., 2013).

The first concern was to limit the out-of-class work-load so the students would not feel overwhelmed by the amount of information (see Egbert et al., 2014). Second, it was important to create predictability for the students (see Steen, 2013), limit the time they have to invest at home, but use this time more efficiently. The intention was that the students should not use more than 20 minutes preparing for their English class. The third aspect was interactivity. The students should be able to receive immediate feedback while also be able to ask questions and address issues they do not understand (see Gilboy et al., 2015). In the following, the general design of the course is described, followed by a description of each of the components with examples and justifications for the choices of method.

10.6.1 Design of the course

The inverted classroom course used in the intervention has 12 modules. The topics for the modules were selected based on an exploration of the corpus of the comparison group collected a year prior to the intervention. Recurring grammatical and stylistic issues were collected and cross-checked with the course books used in the obligatory English course in the first year of high school. Most of the collected topics are covered in the course books only very briefly and superficially, while some are not covered at all (e.g. work with dictionaries and other resources, essay writing strategies). Each module of the course has the following four components:

1. Video lecture which introduces the topic
2. Interactive exercise with immediate automatized feedback
3. Feedback video

---

[17] The language of the platform itself is Norwegian, which makes it difficult to make the course available to teachers outside Norway.

4. Student feedback questions

The students were asked to work with the modules at home, and the teacher could choose to go through the homework with the students in the classroom if there were any questions or issues. An overview of the topics and their content is presented in Table 15. The teachers were free to rearrange the order of the topics to suit their year plan:

Table 15

Overview of the 12 modules included in the intervention material

Module	Topic	Content
1	Grammar Terms	• understanding the basic terminology • word classes
2	Dictionaries & Other Resources	• words in context • work with dictionaries & online resources
3	Concord / Agreement	• agreement – general rules • agreement with specific nouns (*people, everyone*, etc.)
4	Word Order	• word order rules in English and Norwegian • V2 & adverbials / questions
5	It or There?	• *it* & *there* vs. *det* in Norwegian
6	Future Reference	• expressing future • *shall* vs. *skal* in Norwegian
7	Simple or Continuous?	• simple & continuous aspect
8	Adjectives & Adverbs	• adjectives & adverbs
9	Relative Clauses	• relative clauses • which, who, that & commas
10	Prepositions & Punctuation	• prepositions • punctuation (apostrophes & capital letters)
11	Spelling & Easily Confused Words	• spelling (homophones: *to/too/two; were/where*; etc.) • easily confused words (false friends)
12	Writing Essays	• strategies for essay writing • essay structure

The aim of implementing detailed and explicit grammar instruction (including some general style and self-help guidelines) into the obligatory English course is to help teachers bring more metalinguistic awareness into their teaching. A pilot project conducted by the author (unpublished) shows that there is very little improvement in the formal aspects of the students' writing during the school year. Informal conversations among teachers often reveal that they feel the students never seem to learn the grammar they are trying to teach them. The use of the Inverted Classroom course should help both the teachers who perhaps lack the technical knowledge of linguistics to explain certain grammar phenomena and the teachers who feel that grammar teaching is boring or useless because the students 'never learn.' The students should benefit from learning at their own pace with the possibility to replay and repeat the lessons as many times as necessary (e.g. before tests and assignments). This course can also be used to adjust teaching to students at different competence levels (both advanced younger students and older students who have issues with elementary grammar).

The use of video-lectures instead of being personally present in the classroom is not very common in intervention studies. However, there are several reasons why video-lectures are more suitable for this type of didactic intervention. First of all, an online based course with videos and exercises is flexible and reusable by many teachers regardless of their physical location. Such a course can be a lasting contribution to the proverbial toolbox of English teachers across the country and not only a tool for one intervention. Second, by removing the researcher from the classroom, the potential Hawthorne effect, i.e. the act of observing in itself leading to improvement, should be minimized (McCarney et al., 2007). The students were asked to participate in the study towards the end of the semester and all the data were accessed only retrospectively for those students who contested, which also eliminates the potential for students adjusting their writing because they are part of a research project. And finally, the use of video mini-lectures followed by discussions of the reviewed topics in the classroom is ranked very high in Hattie's meta-study of teaching and learning methods: it is the fourth most effective method out of 138 tested approaches (Handke, 2013b, p. 64; Hattie, 2009, p. 112). The disadvantage of the researcher not being present in the classroom is the lack of control of the follow-up discussions.

10.6.2 Video lectures

Handke (2013b, p. 63) suggests that videos should be no longer than 20 minutes for college students. I have decided to limit the presentation of each topic to a maximum of 10 minutes since the students involved are younger and English is just one of many subjects they need to prepare for in an average school week. Prodoehl (2016, p. 77) points out that grammar teaching tends to be considered boring and confusing by many L2 learners and it is thus the part of the English curriculum which could benefit the most from being moved into a calmer and more controlled homework domain where the students would have the option of self-pacing, pausing, and rewinding when necessary. Most of the topics of the video lectures are grammar topics but also general learning strategies, such as the work with aids and sources, and essay writing strategies are included. The grammar topics are explained in a simple but linguistically sound manner and, where applicable, they include contrastive points between English and Norwegian. The language of the videos is English.

The instructional videos are recorded in a studio with professional lighting and sound support to ensure image and sound quality. According to Handke (2013b), the presenter should be visible during the presentation of the topic in order to better preserve the viewers' attention. The chosen format thus contains a large screen with a Power Point presentation and the instructor standing next to it.[18] During the course development, a short survey was conducted among first year university students attending an introductory course in English grammar to check whether the presence of the instructor was important for the students and about 60% of the students prefer to see the presenter while they watch a video-lecture (Gaustad & Horn, Unpublished). The videos are unscripted but based on a prepared Power Point presentation with mostly self-made visual aids such as pictures, comic strips, and animations. The intention is not a rehearsed perfection but liveliness and relevance for the students (see Kiesler, 2013).

10.6.3 Exercises and feedback videos

In addition to the video-lecture, each module includes one or two sets of exercises with automatized feedback. The requirement for automatized feedback limits the form of the exercises to some degree. Most of the exercises require gap filling, rearrangement of

---

[18]   See Appendix 2 for examples from the course platform.

elements, or choice from a set of options. Such exercise forms are often criticized for offering only right-wrong as feedback and no explanation as to why the chosen form is correct or incorrect. Automatized feedback is in itself not formative and has only limited learning potential. In order to compensate for this problem, short feedback videos follow each exercise set. In the videos, I comment on each sentence or example from the exercise explaining which option is correct, why it is correct and, if applicable, also why other options are not possible. The short feedback videos are recorded as screencasts (using Kaltura CaptureSpace) with the sentences from the exercise in a Power Point presentation and a recording of the instructor's voice.

The exercises not only add practice but, combined with the feedback videos, they also give 'instant formative evaluation' (Loviscach, 2013, p. 8). In other words, the students get feedback on the reasoning they used in choosing their answers and they can redo the exercises if they wish to after watching the feedback video or at a later point as part of their revision. Hattie's meta-study (2009, p. 181) evaluates formative feedback as having a considerable learning effect (see also Handke, 2013a). Adding exercises to the video lectures can also contribute to sustaining the attention of the students and avoiding mind-wondering (Szpunar, Khan, & Schacter, 2013).

10.6.4 Student feedback

At the end of each module there are three survey questions. The first two focus on students' understanding and the perceived difficulty of the topic. Students answer with a five-point Likert scale which is then displayed as corresponding 'smiley faces' on a course overview page available to the teacher. The last question is open, and the students can write any comments or questions they may have about the topic. The teachers can review these questions before the class, which gives them an opportunity to adjust the content of the subsequent in-class lesson to the difficulties or questions which arise from the homework assignment.

Handke (2013a, pp. 16-17) explains that, in order to create a connection between the theoretical videos and the learning activities in the classroom, there has to be an assessment system which would be available to the instructor prior to the class. The Campus Inkrement platform offers a simple interface where the teacher can quickly check the progress of individual students and the class as a whole. Color-coding of exercise answers (red for wrong answers, green for correct ones) and feedback forms with

132

emoticons offer a quick overview for the whole class, while it is also possible to review each student's progress in more detail by clicking on the student's name.[19]

10.6.5 Didactic setup

The participating teachers were asked to continue in their practice of teaching as before (including teaching grammar with the methods and frequencies they used to), but, in addition, to assign the inverted classroom course as homework. They were instructed to encourage the students to use the course for revision before written tests. However, the actual frequency of use was not controlled for. The teachers were generally satisfied with the course and in three of the four classroom most of the modules were used.[20] Errors in the software setup reported by the students or the teachers during the intervention period were corrected immediately, however the corrections cannot be implemented before the launch of a new school year.

10.6.6 Students' reported use of the material

Each student participating in the intervention was given a short questionnaire about their experience with the inverted classroom course, in addition to the background questionnaire. The data were collected in early May, that is after nine months of use. Most participating students responded to all the questions in the questionnaire. There were seven questions in the survey – four of them regarding the students' use of the English grammar Inverted Classroom course measured on a four-point Likert scale (see Appendix 4 for the full questionnaire). Two of the questions focused on the frequency of use and two focused on the perceived experience with the course. The most relevant questions for the evaluation of the effect of the intervention are:

- How often have you used the inverted classroom course in English during the current school year?
- Have you used some of the modules several times?

Out of the 119 students involved in the intervention 11 state that they used the course *often*, 87 state that they used the course *sometimes*, 15 state that they used the course *rarely*, 2 state they *never* used the course, and two did not answer this question.

---

[19] See Appendix 3 for examples of the teacher overview.

[20] There was one vocational class included in the intervention and this class only used 4-6 modules. However, vocational classes only have 3 hours of English instruction per week while general studies classes have 5 hours per week.

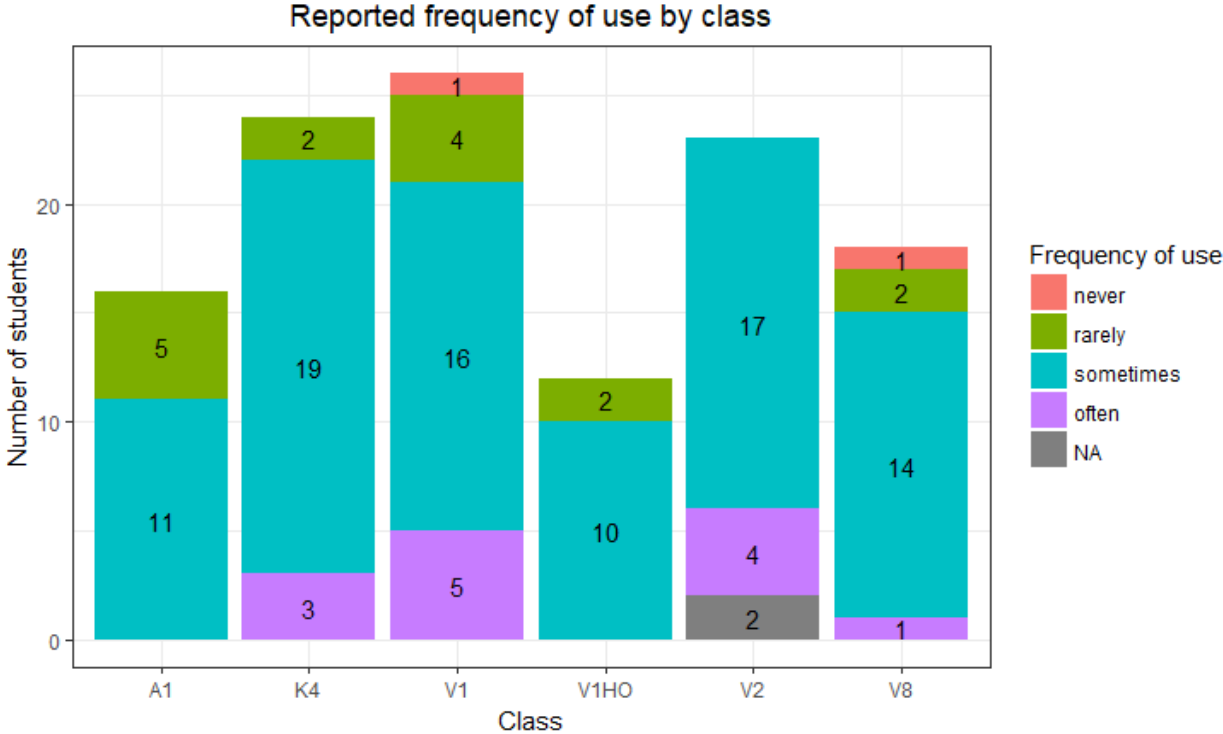Figure 24 illustrates the division of the students based on their reported frequency of use in each class:

Figure 24. Self-reported frequency of use of the intervention material by all intervention students divided by the class they attended.

Due to the vague formulation of the questions in the usage questionnaire, there might be a reason to doubt the validity of the self-reporting. Most students reported using the course materials *sometimes* which is a rather imprecise measurement. In order to better understand how much time students attribute to the answer *sometimes*, usage logs from the platform of the course were retrieved for 38 students (from two classes, one general studies class and one vocational class). These are analyzed below.

10.6.7 Students' actual use of the material

Those students who reported using the course material *often* (five students) on average spent 200.8 minutes on the course during the school year (median = 231 minutes, 9 modules used). Those who reported using the course *sometimes* (26 students) spent on average 91 minutes on the course during the school year (median = 79.25 minutes, 7.5 modules used) and those who reported *rarely* using the course (six students) spent on average 63.42 minutes on the course during the school year (median = 28.5 minutes, 3.5

modules used). There was only one student from this selected group who reported never using the course and this student actually used the course on four occasions spending altogether 18 minutes interacting with the materials. Since there was only one student in this category, the statistical tests, which involve comparing means and ranks, were performed only on the three categories with more than one participant. Table 16 summarizes the usage data:

Table 16

Summary of the actual use of the intervention material by a selection of the students from the intervention group

How often did you use the course?	Mean time (minutes)	Median time (minutes)	Max time (minutes)	Min time (minutes)	Modules used (median)
Often	200.80	231.00	279	111	9
Sometimes	91.02	79.25	179	6.5	7.5
Rarely	63.42	28.50	255	0	3.5
Never	18.50	18.50	-	-	4

There are significant differences in mean time spent with the course materials between the groups (tested with ANOVA run in R (Field et al., 2012, pp. 398-461) $F(3, 34) = 3.91, p < 0.05$). It seems that the self-reported data of the students roughly correspond to the time used. However, there is high variance in the *sometimes* and *rarely* groups (mean sometimes = 91.02, SD sometimes = 70.01; mean rarely = 63.42, SD rarely = 97.26) and the amount of time the students from these two groups used on the course does not significantly differ from each other. The only significant difference is between the students who reported using the course often and the other two groups.

10.6.8 Relevance of the time use for the intervention

It is clear that some students did not spend much time interacting with the course while others were more dedicated. The teachers were not specifically instructed how they should follow up students who did not do their homework. This factor needs to be controlled in potential follow-up studies. What is also relevant for the evaluation of the effect of the intervention on the agreement marking accuracy in particular, are the points in time when the students used the material and how much time they spent on the two modules which deal with agreement.

Most of the general studies students used the course for the last time in March and they submitted texts in November, December, February, and May. All of the vocational students used the course for the last time in January and they submitted texts in September, November, and May (see complete usage data in Appendix 5). This means that at least one of the premises of the intervention, i.e. that the materials should help the students revise before written tasks was not met. The course was last used 1-3 months before the final written exams. It is not clear whether this intention was not communicated clearly enough to the students or whether the students just did not follow the recommendations.

The second relevant factor for the evaluation of the improvement of accuracy in agreement marking is the time spent on the two modules which deal with agreement. When all 27 students with three measurements are considered together, they have spent on average 15.52 minutes on the two modules which deal with agreement (median = 7.5 minutes). Only the videos in these two modules take 20.27 minutes to watch so it is clear that many students only watched parts of the videos or only filled out the exercises without watching the videos at all. Only 10 students spent more than 20 minutes on these two modules (total time for both modules) and only eight of these watched the whole length of the videos in both modules. This goes against the basic premise of the intervention, i.e. that the students should get explicit grammar instruction in the problematic areas in addition to the traditional classroom teaching. Most of the students who did their homework only did the exercises and did not watch instructional videos. When less than one third of the students used the course as intended it is not expected that there would be any significant differences in the error scores of the intervention group and the comparison group.

11 Data analysis

This chapter focuses on the quantitative description of the similarities and differences between the comparison and intervention groups. The first part of the chapter discusses the general characteristics of these two groups and the process of data selection for statistical analysis (based on matching of the controlled variables such as school, track, and teacher). The second part describes the statistical procedures applied to test the distribution of the scores and the differences between these two groups. In the following, only the students with three measurements are considered since it is not possible to measure the effect of the intervention in the writing of those students who did not submit texts at all three measurement points.

11.1 Data type and background information

The basic characteristics of the data from the comparison group are discussed in detail in chapter 4. As with the comparison group, the intervention group data were collected in three high schools in the Agder region. There were altogether 119 students participating in the intervention (six classes, five teachers, four different tracks). The intervention corpus contains 348,380 words. Out of the 119 participating students, 100 have written texts at all three measurement points so their progress through the school year could be traced. The topics and format of the texts are similar as for the comparison group. Most of the contributions are tests written at school under controlled conditions, but there are also some home assignments with various sources used.

11.1.1 Controlled variables

The intention of the intervention was to match the students according to their school, study track, and teacher in order to avoid many independent variables which could influence the results. There were 134 students with three measurements in the control group and 100 in the intervention group. Out of these students, 69 in the control group and 64 in the intervention group are fully matched, i.e. they come from the same schools, tracks, and were taught by the same teachers. Only two of the three schools are represented in this selection due to a leave of absence of the participating teacher from the third school. The students come from three general studies classes and one vocational class (Healthcare, Childhood and Youth Development). They were taught by three

different teachers. Below follows a brief group-wise description of the background of these students.

11.1.2 Comparison group

There are 69 students from the comparison group who can be matched to students from the intervention group. Among these students there are 45 females, 22 males, and two students did not specify their gender. 57 claim to be monolingual Norwegian (no other languages used at home), four use Norwegian and an additional language at home (Turkish and Arabic; English and Danish; and two use English in addition to Norwegian), four students only use languages other than Norwegian at home (Chechen; Arabic and Tigrinya; and two use only Polish), and four students did not answer this question. One student reported having a learning-related diagnosis. Three students did not take their entire education in Norway (one starting in the eighth and two in the tenth grade).

11.1.3 Intervention group

There are 64 students from the intervention group who can be matched to the students from the comparison group. Among these students there are 39 females and 25 males. 51 claim to be monolingual Norwegian (no other languages used at home), six use Norwegian and an additional language at home (Romanian, Vietnamese, Persian, and three use English), and seven students only use a language other than Norwegian at home (Persian, Amharic, Romanian, and two each for Dutch and Polish). One student reported having a learning-related diagnosis. Eight of the students did not take their entire education in Norway (three started in the second grade, two in the third, and one each in the fourth, fifth, and eighth grade). See Appendix 6 for a full overview of the basic background data for both groups.

11.2 Quantitative description of the material

The background information described in the previous section shows that the comparison and the intervention groups are fairly similar when it comes to gender composition, exposure to additional languages and occurrence of learning-related diagnoses. This section focuses on the description of the quantitative measures for the two groups, such as the amount of text produced (in words per year and per measurement point), and the error scores of the individual students (total scores and scores per measurement points). The

development of the students from fall to spring is also evaluated. Only the fully matched students are included in this description.

11.2.1 Quantitative description of the comparison group

The matched students in the comparison group contributed 191,555 words to the corpus and made 1,296 subject-verb agreement errors in their texts. There are no distinctions made in this chapter regarding the subtypes of agreement errors (as in chapter 5), only the general ability to use the correct English agreement patterns is evaluated. Even though there is a high degree of individual variation among the students, the comparison group is treated as one population in the following in order to be able to evaluate the differences between the comparison and the intervention group and potential effects of the intervention.

11.2.1.1 Word production

The data were collected over one school year. The number of texts each student wrote differs, but each of the 69 students submitted at least one text at each of the measurement points. The mean word production for the comparison group is 2776.16 words per student and the median word production is 2672 words per student, which suggests that the distribution is close to normal. The Shapiro-Wilk test run in R (Field et al., 2012, pp. 182-185) confirms a normal distribution ($W = 0.99$, $p = 0.78$). The distribution is illustrated in Figure 25:

Histogram of words produced during the school year
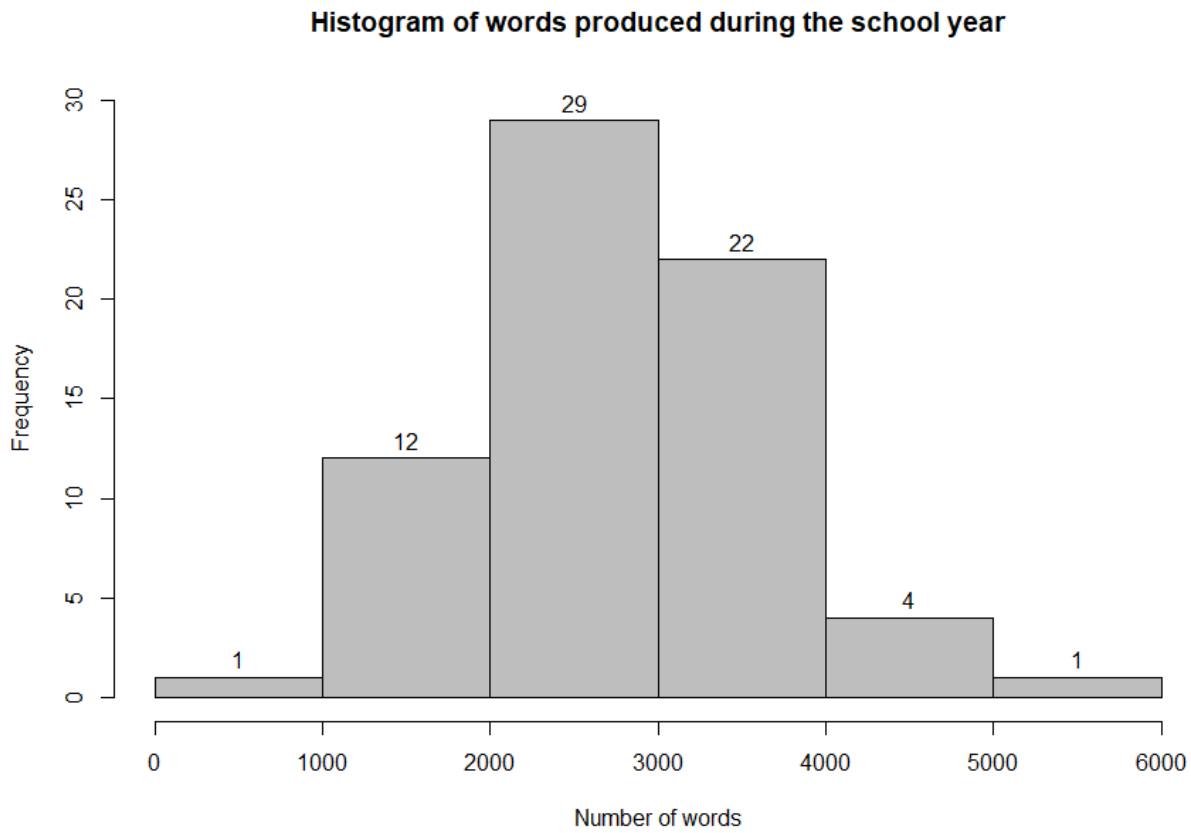
Figure 25. Histogram of the number of words produced by the individual students in the comparison group during the whole collection period.

There is a clear increase in the word production when the words produced are analyzed separately for the three measurement points. The students write longer texts as the school year progresses. The average amount of words written per student more than doubles (median Fall = 560, median Midterm = 901, median Spring = 1236). Histograms in Figure 26 – Figure 28 illustrate the distribution of words produced by individual students at the three measurement points:

Histogram of words produced at the Fall measurement point
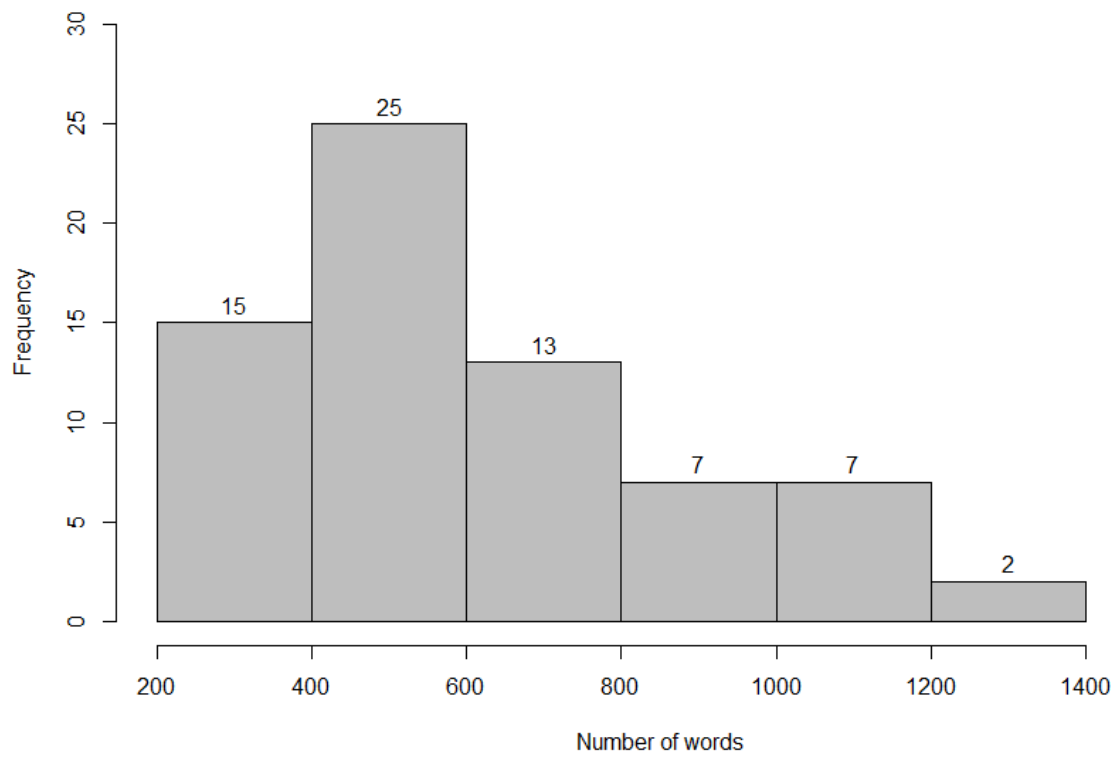
Figure 26. Histogram of the number of words produced by the individual students in the comparison group at the Fall measurement point.
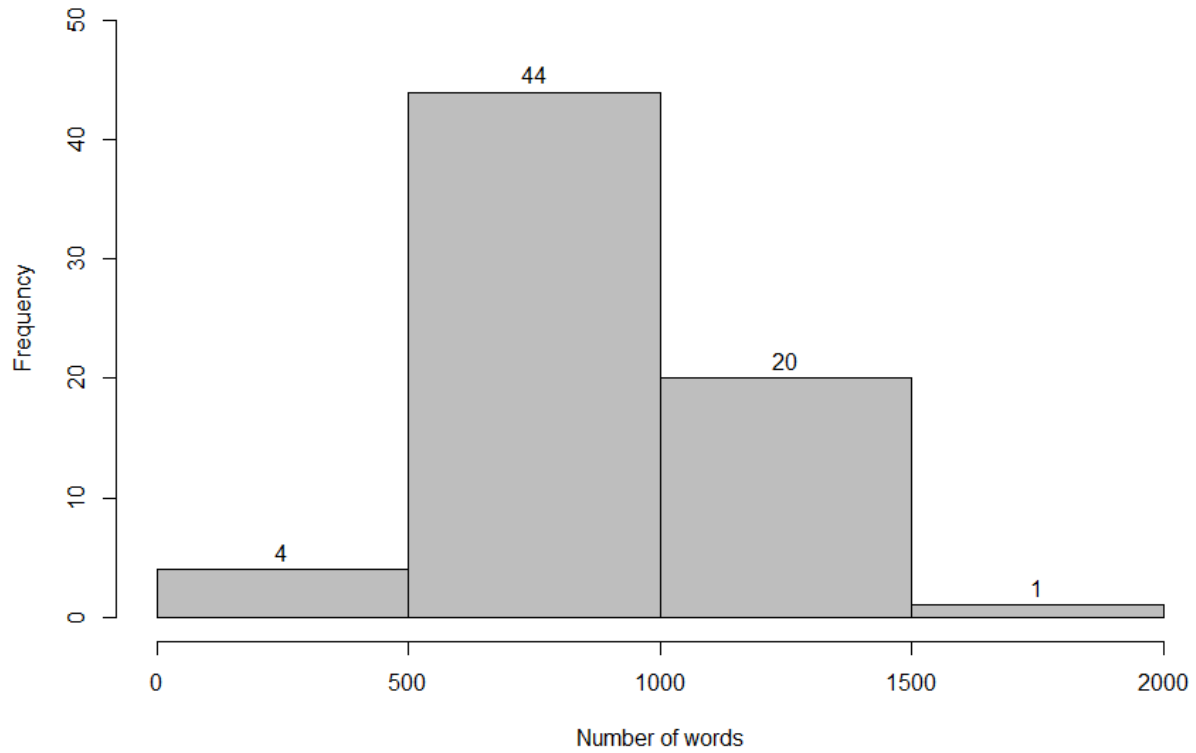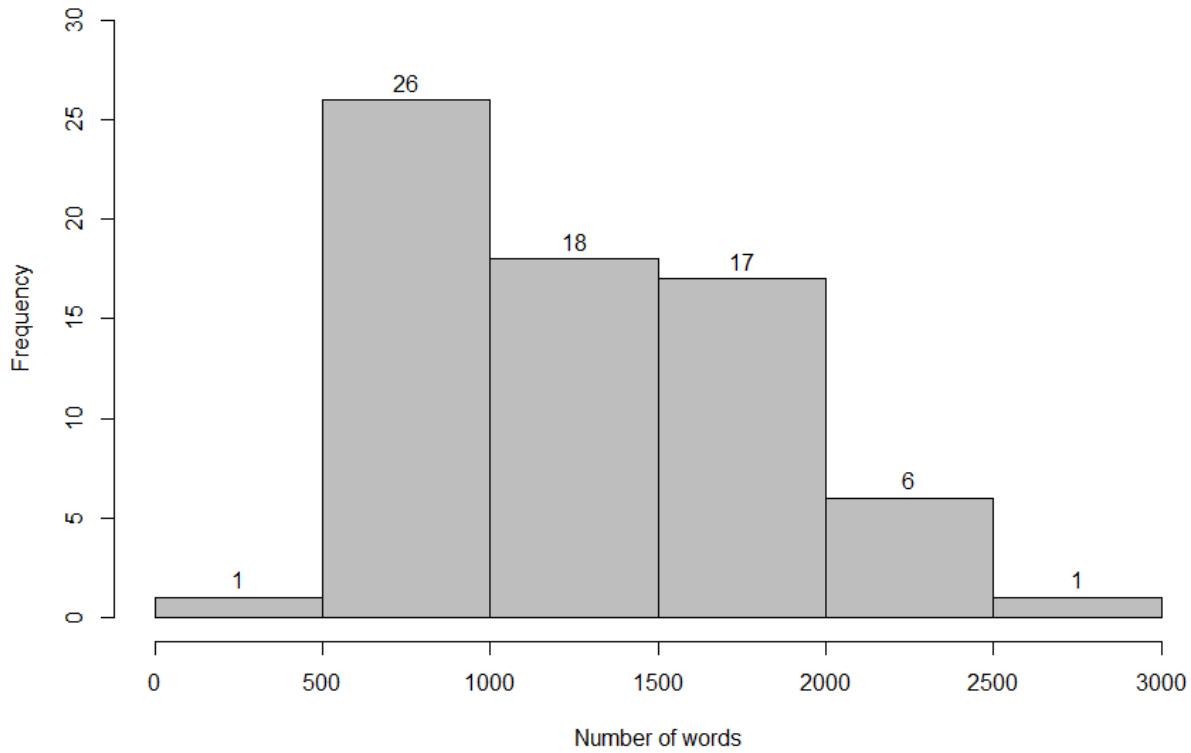
Figure 27. Histogram of the number of words produced by the individual students in the comparison group at the Midterm measurement point.

Histogram of words produced at the Spring measurement point

Figure 28. Histogram of the number of words produced by the individual students in the comparison group at the Spring measurement point.

11.2.1.2 Error scores

The mean error rate for the comparison group is 8.40% and the median is 8.00%. This means that the distribution has a slight positive skew, or in other words, there are only few students with very high error scores (four students score above 15%). The Shapiro-Wilk test run in R confirms that the distribution of total error scores is not normal ($W = 0.96$, $p = 0.02$). The total error scores are illustrated in a histogram in Figure 29:[21]

---

[21] Note that total scores are calculated across all texts each student submitted during the school year, so they obscure any potential development during the school year.
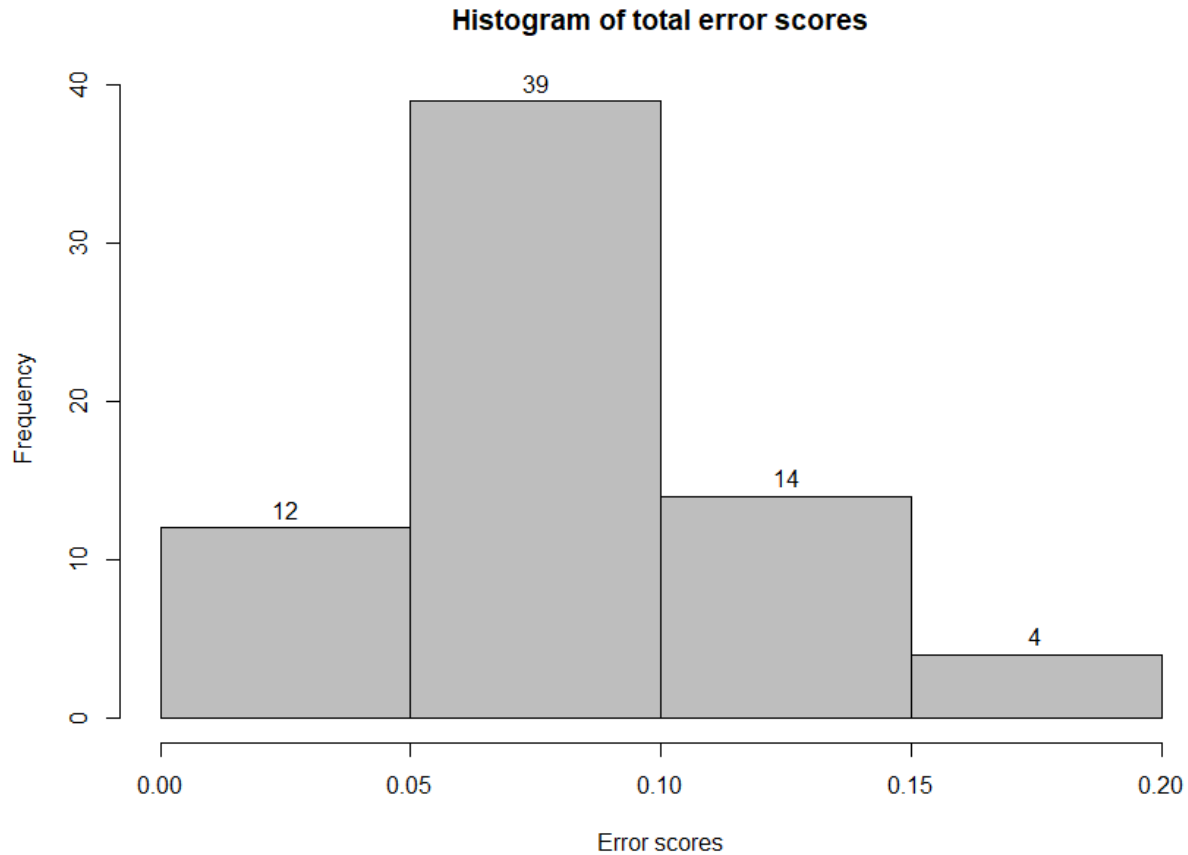
Figure 29. Histogram of the total error scores for the individual students.

As mentioned in chapter 10.4.1, in order to compare development over time and across students, the texts are divided into three measurement points for each student: Fall (all texts written from August to November), Midterm (all texts written from December to February), and Spring (all texts written from March to May). In this way, the progress of the students through the three measurement points can be compared regardless of the number of texts each student wrote in the given time period. Some students wrote only one text in each time period while others wrote two texts in some periods. Error scores are calculated after the texts were collapsed into the measurement point clusters, that is, as if all texts written in each period were only one file.

When the error scores are divided into the individual measurement points some changes in the distribution can be observed. The average error scores in the Fall appear to be lower than the error scores in the Midterm and the Spring. The distribution of the error scores in the Fall is not normal ($W = 0.88, p < 0.001$). Many students score under 5%

144

errors. The mean for the Fall is 8.67%, and the median is 6.25%. 11 students score above 15%:

Histogram of Fall error scores
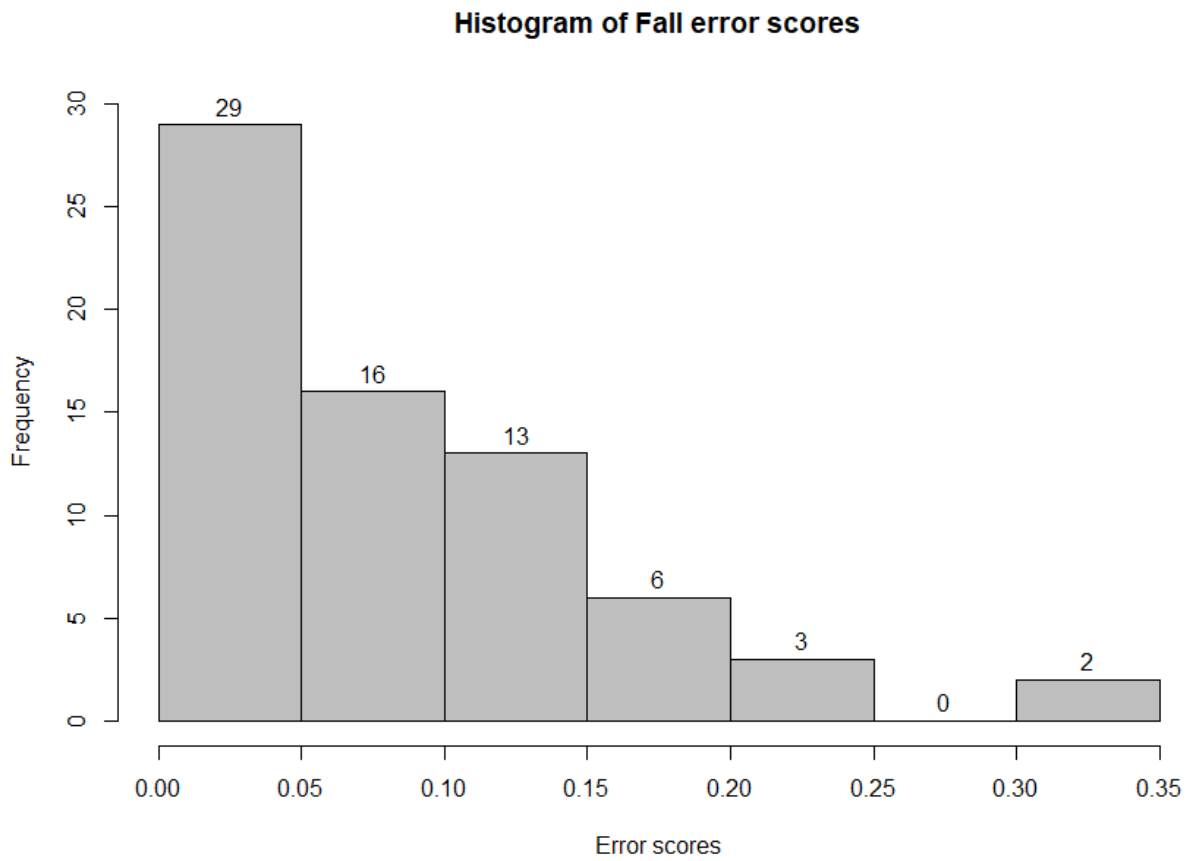
Figure 30. Histogram of the error scores of the comparison group at the Fall measurement point.

The distribution of the error scores changes in the Midterm as compared to the Fall. The mean for the Midterm is 9.04%, and the median is 8.00%. Eight students score above 15%. The distribution is normal ($W = 0.97, p = 0.18$):
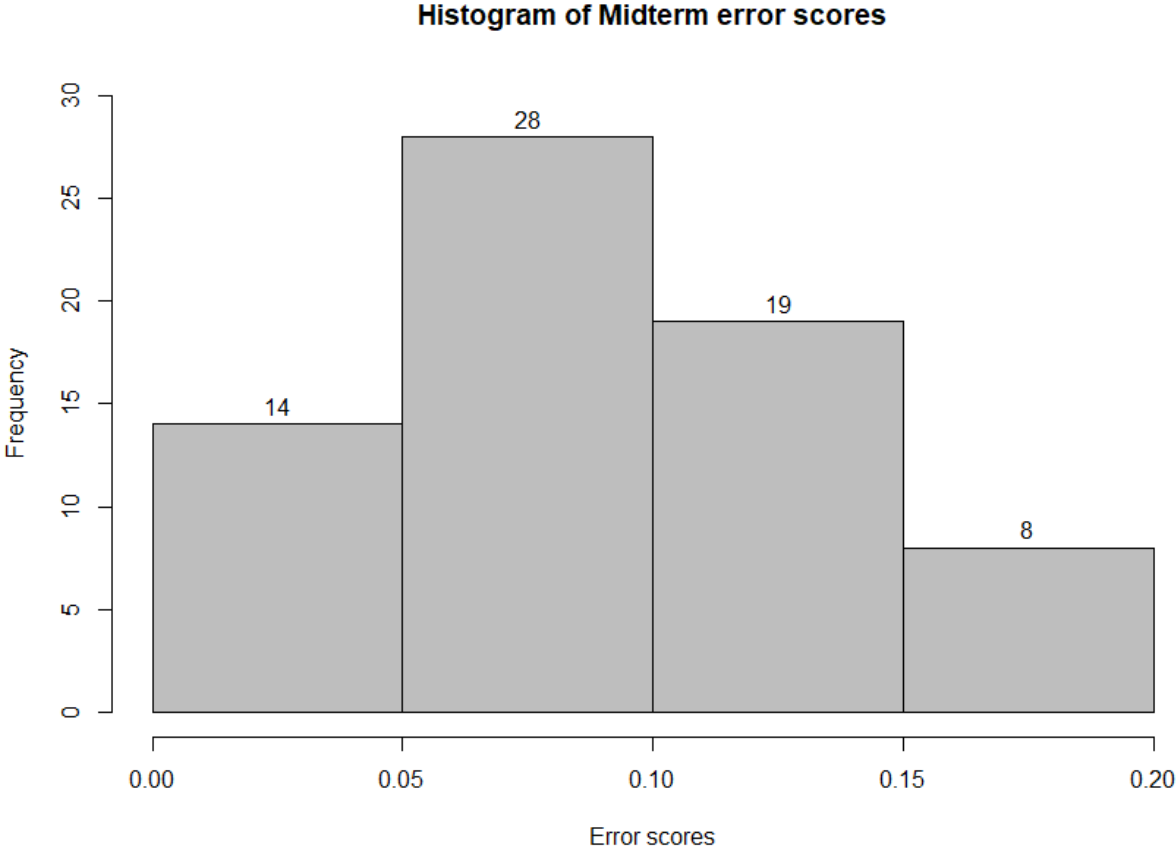
Histogram of Midterm error scores

Figure 31. Histogram of the error scores of the comparison group at the Midterm measurement point.

The distribution of the error scores in the Spring is not normal ($W = 0.89$, $p < 0.001$). The mean for the Spring is 7.73%, and the median is 7.32%. Four students score above 15%, as shown in Figure 32:
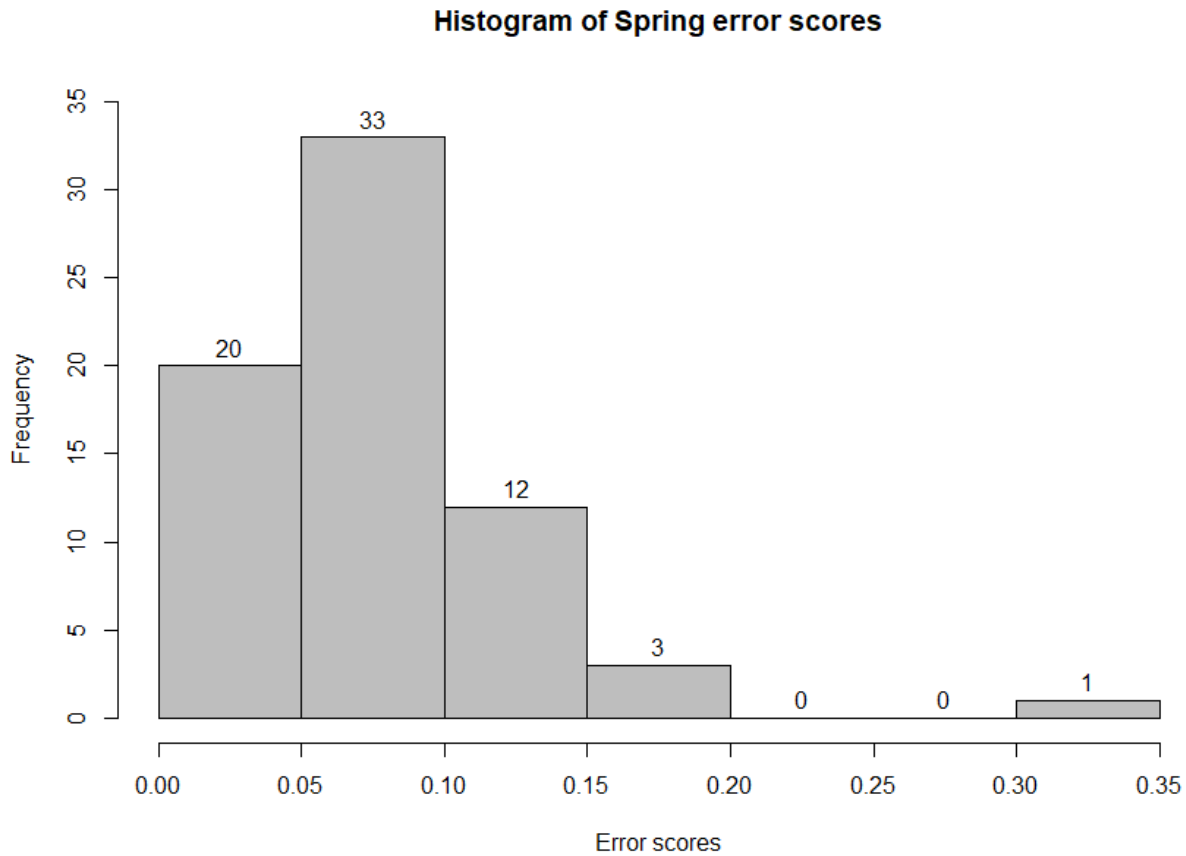
Histogram of Spring error scores

Figure 32. Histogram of the error scores of the comparison group at the Spring measurement point.

The distribution of error scores is heavily positively skewed in the Fall, it develops into a normal distribution in the Midterm, and back to a positive skew in the Spring. This means that both in the Fall and in the Spring, there were relatively many students who produced only few agreement errors while the distribution in the Midterm was relatively even. It is possible that the task in the Midterm was more complex than in the other two points, which caused the increase in agreement errors in the Midterm. Only one of the outliers at the three measurement points is the same student in more than one measurement point. This suggests that the outliers might be students who perhaps had a bad day when the test was written and who otherwise score better than what the

147

measurements might suggest. It also seems that many students who produce very few agreement errors in the fall (the 37 students in the first range of the histogram in Figure 30) produce more errors later in the school year. This could be explained by an increased length and complexity of their texts later in the school year.

There is, understandably, a high degree of individual variation in the error scores both among the students and between the measurement points for each student, ranging from 0% to almost 35% errors. There are three students with error scores above 30% at at least one of the measurement points and additional four students who score above 20% at at least one of the measurement points. When the individual scores are examined in detail, four of the seven outliers show high error rates throughout the collection period (14STA18, 15STV09, 15STV20, and 15STV49) – they all score above 10% in the Spring and around 15% in total scores. The plot in Figure 33 illustrates the error score distribution in the population between the three measurement points:

Student error scores in three measurement points
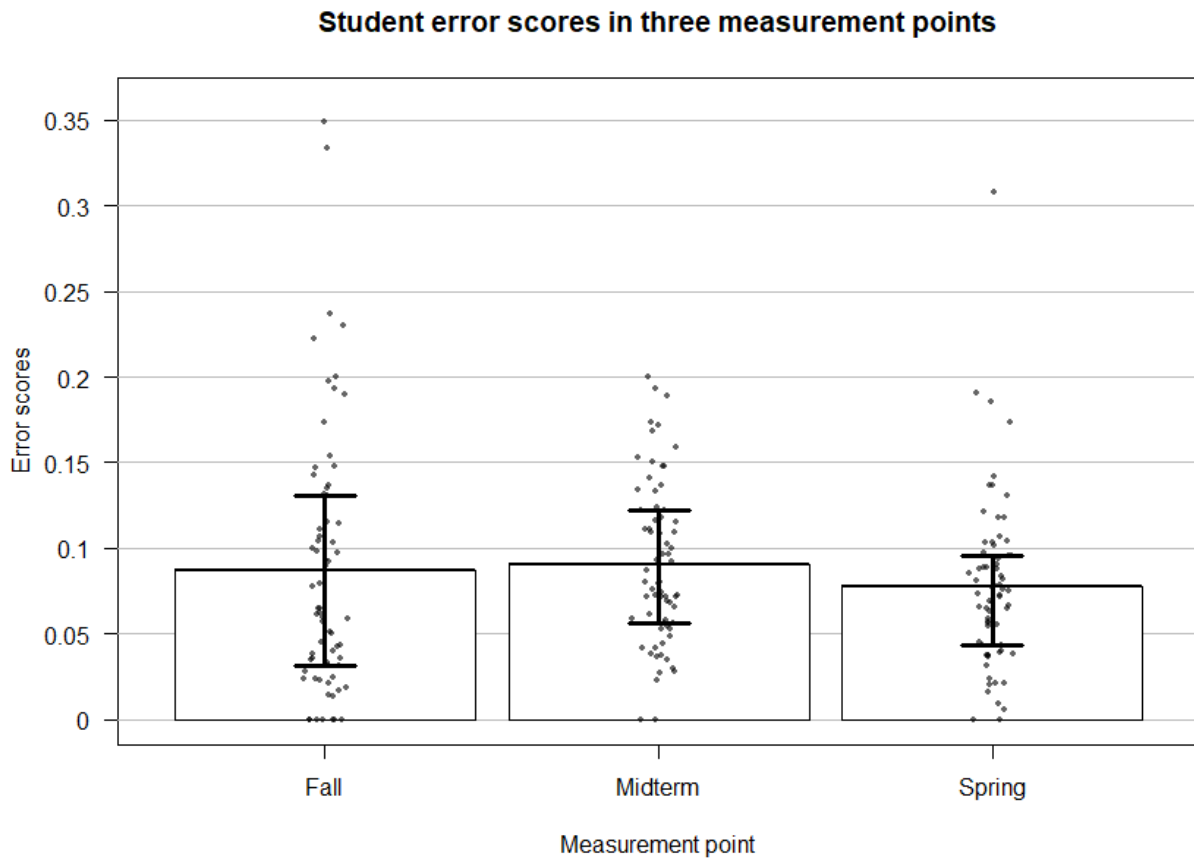
Figure 33. Distribution of the individual scores of the comparison group students at each of the three measurement points.

Even though some students improve their accuracy of agreement marking during the school year, and the spread of the data around the mean changes during the school year, it is not clear that there are any group-wise trends. Since the error scores are not normally distributed at two of the measurement points, Friedman's ANOVA was used in R (Field et al., 2012, pp. 686-692) to test the differences in scores between the measurement points. There are no significant differences between the scores when the whole group is considered as one population ($\chi^2(2) = 4.40, p = 0.11$). This means that there is no statistically significant improvement in agreement marking accuracy in this student group during the school year despite five hours of English instruction a week.

11.2.1.3 Developmental trajectories

It is also important to consider the individual developmental trajectories of the students in this population, i.e. whether and how much the students improved their scores during the school year. There are no discernible trends in the developmental trajectories when all students in the comparison group are considered together. Some students show clear progression through the school year towards lower error scores, while others actually finish the school year with worse scores than they had at the Fall measurement point. In order to see how the individual students' scores develop throughout the school year, the three measuremnt point scores of each student are compared to each other. There are altogether six options for developmental trajectories between the three measurement points. The trajectories are visualized in Figure 34 a-f. (In the following the letters N_F, N_M, and N_S represent the number of agreement errors produced by a single student at the Fall, Midterm, and Spring measurement point respectively):

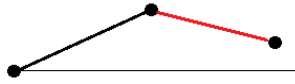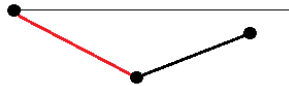a. The accuracy consistently improves from one checkpoint to the next (i.e. $N_F > N_M > N_S$)
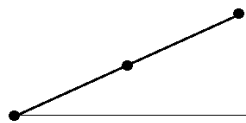
b. The student makes more errors at Midterm than in Fall, but the accuracy in Spring is higher than the start (i.e. $N_F < N_M > N_S \wedge N_F > N_S$)
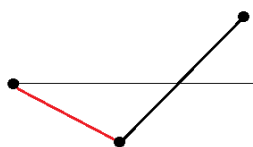
c. The student initially improves, then the accuracy goes down again, but the accuracy in Spring is higher than the start (i.e. $N_F > N_M < N_S \land N_F > N_S$)

d. The student makes more errors at Midterm than in Fall, then the accuracy improves, but the accuracy in Spring is lower than the start (i.e. $N_F < N_M > N_S \land N_F < N_S$)

e. The student initially improves, but then the accuracy goes down, and the accuracy in Spring is lower than the start (i.e. $N_F > N_M < N_S \land N_F < N_S$)
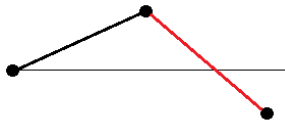
f. The student consistently makes more errors from one measurement point to the next (i.e. $N_F < N_M < N_S$)
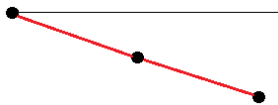
Figure 34. Possible developmental trajectories.

In the first three options (Figure 34 a-c) the students all in all score better at the end of the school year than when the school year started. In the second three options (Figure 34 d-f) the students score worse at the end of the school year than when the school year started. However, in option four (Figure 34 d) the accuracy after the Midterm measurement point is improving while in options five and six it is going down. The distribution of the students in these six trajectories is illustrated in the graph in Figure 35:

Developmental trajectories
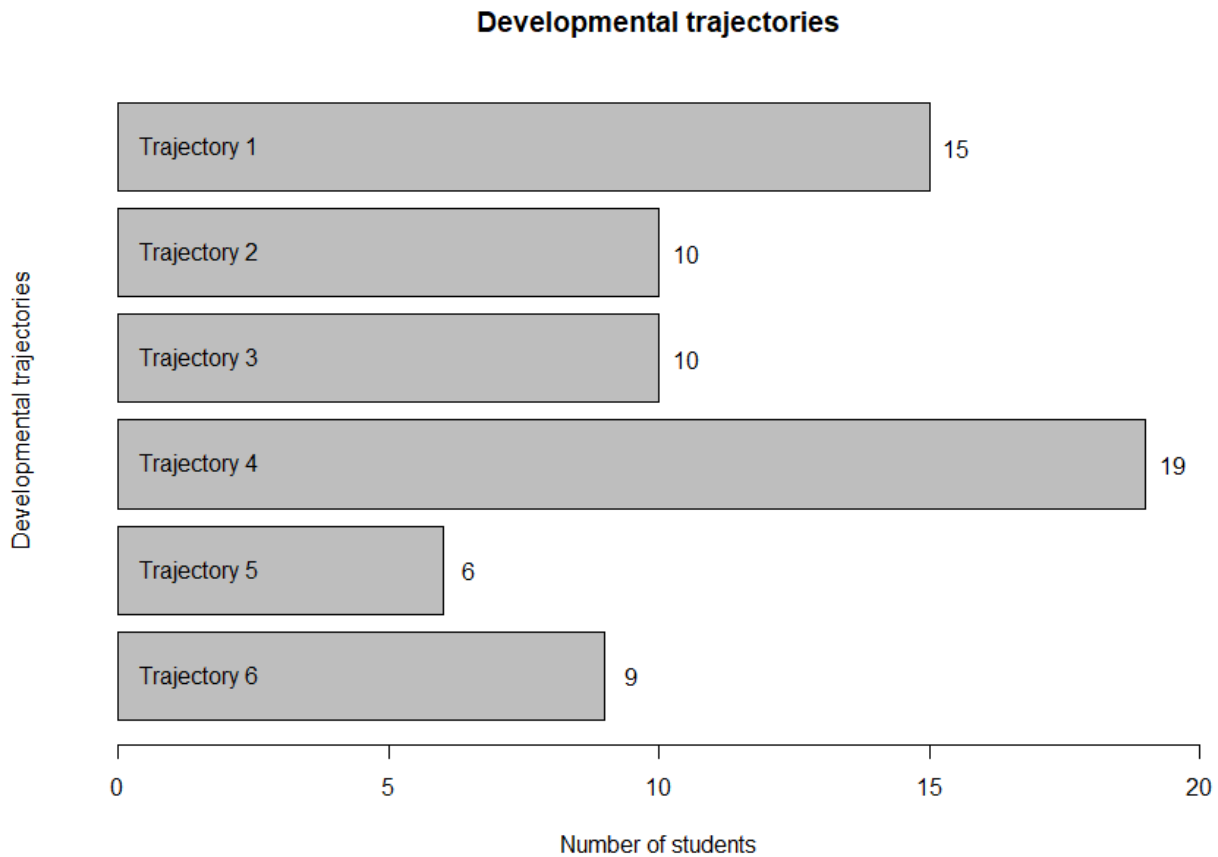
Figure 35. Developmental trajectories of the students in the comparison group.

It is clear from Figure 35 that approximately half of the students in the comparison group (35 students in trajectories 1-3) improve in their agreement marking accuracy during the school year, while the other half of the students (34 students in trajectories 4-6) have lower accuracy in the Spring than in the Fall. Again, part of the explanation for this trend could be that the students are writing more complex texts towards the end of the school year. This makes their language production more vulnerable to some types of agreement errors (e.g. distance-triggered errors or errors with infrequent, irregular nouns). Some evidence pointing in this direction is the fact that the length of their texts increases to more than double from Fall to Spring. The median length of the texts in words at each of the three measurement points in the comparison group is repeated here from section 11.2.1.1:

Median length Fall = 560 words

Median length Midterm = 901 words

Median length Spring = 1236 words

It is conceivable that the students who write longer and more complex texts make more errors in general. However, some students manage to improve their accuracy despite writing longer texts (trajectories 1-3), so the increasing length and complexity of texts cannot be the only explanation for the poor error scores and negative development of half of the students (34 out of 69). It seems that many of these students do not have good command of the rules of agreement in Standard English. Their scores fluctuate, and the agreement marking is not consistent even at the end of the school year (i.e. the end of their 11th year of instruction).

11.2.2 Quantitative description of the intervention group

The students in the intervention group contributed 203,159 words to the corpus and made 1,253 subject-verb agreement errors in their texts. As with the comparison group, only the general ability to use correct agreement patterns is evaluated in this section. The intervention group is treated as one population in the following in order to be able to evaluate the differences between the comparison and the intervention group and potential effects of the intervention.

11.2.2.1 Word production

The data from the intervention group were also collected over one school year. The mean word production for the intervention group is 3174.36 words per student and the median word production is 3050 words per student, which means that the intervention group students produced more words both totally and per student than the comparison group students. The distribution is normal (Shapiro-Wilk test run in R ($W = 0.98$, $p = 0.23$)):
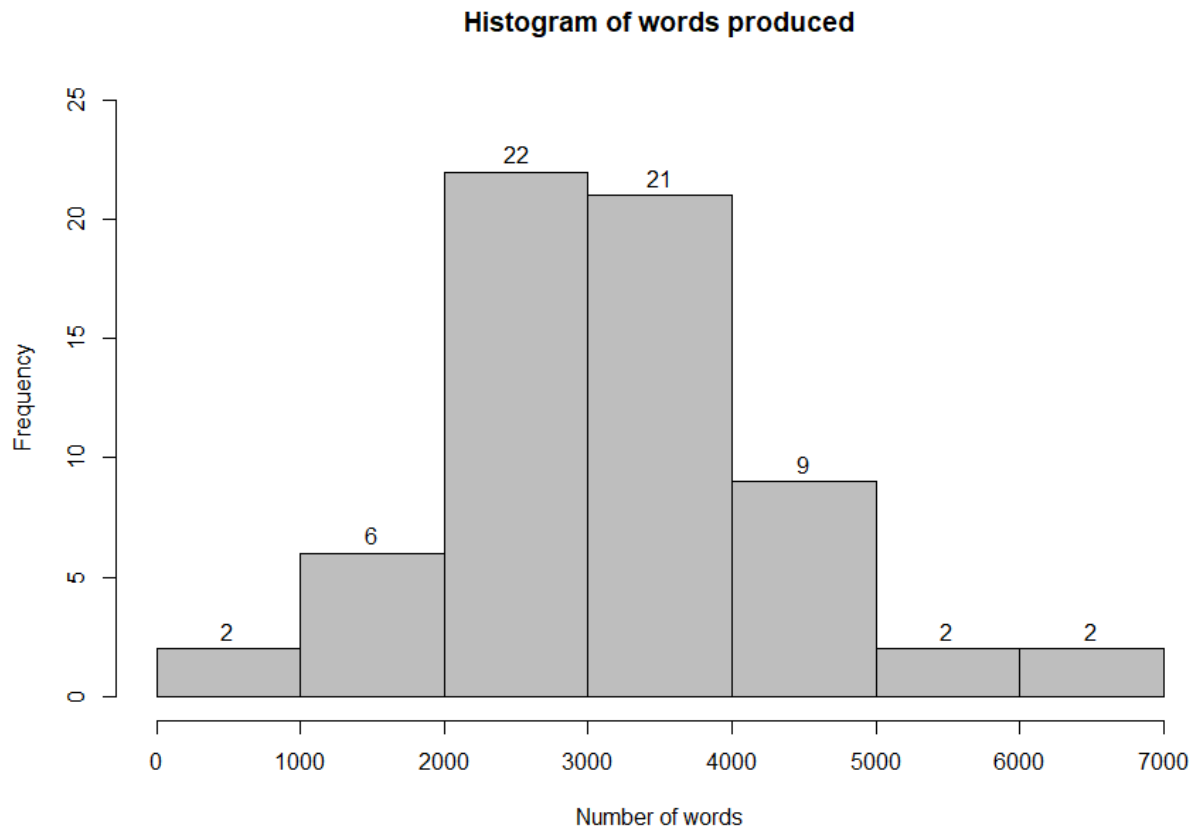
Figure 36. Histogram of the number of words produced by the individual students in the intervention group.

Like the students in the comparison group discussed in section 11.2.1, the students in the intervention group write longer texts later in the school year than in the beginning. The distribution of the word production at the individual measurement points is illustrated in *Figure 37 – 39*:

Histogram of words produced at the Fall measurement point
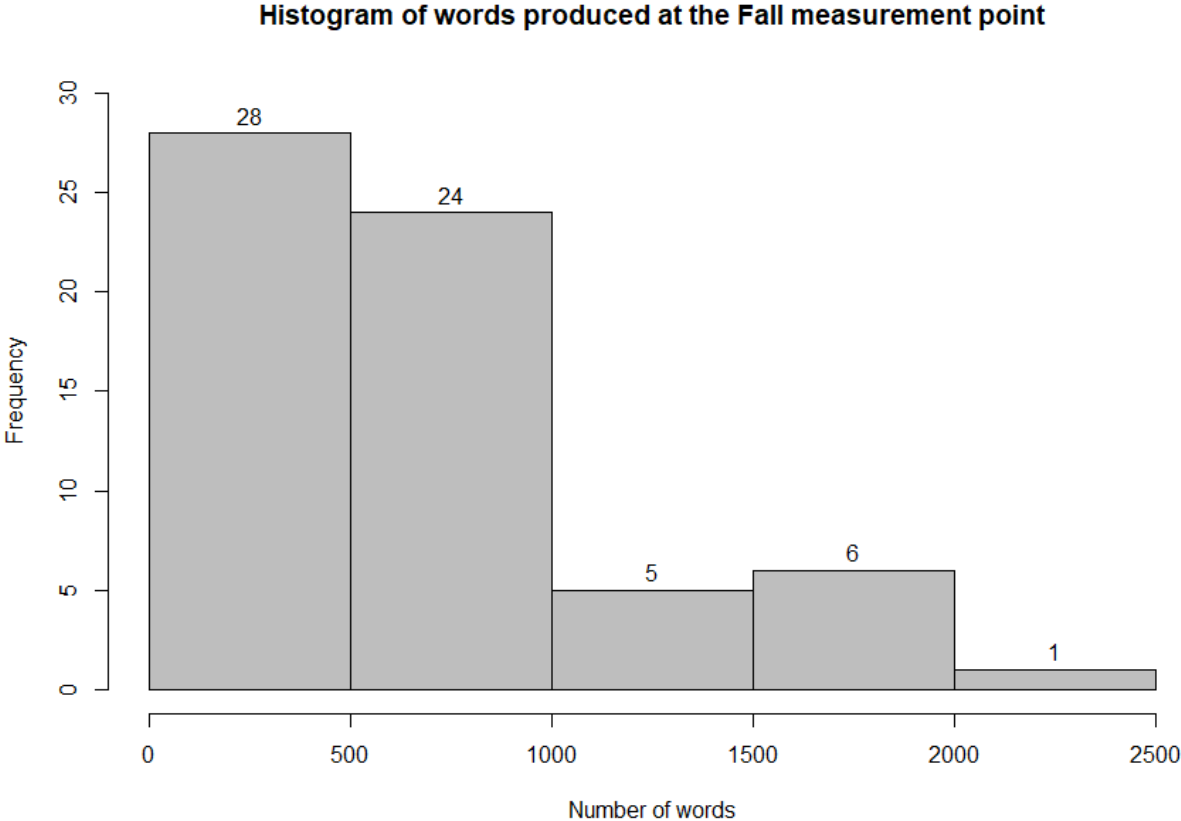
Figure 37. Histogram of the number of words produced by the individual students in the intervention group at the Fall measurement point.
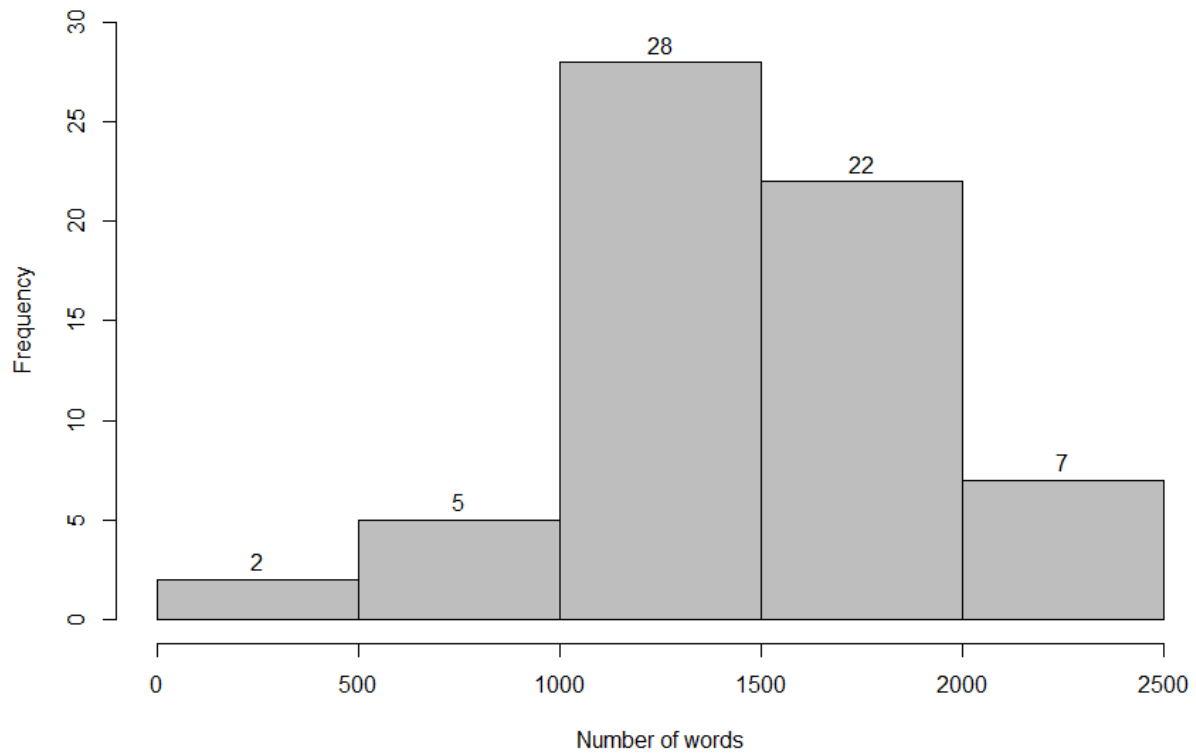
Figure 38. Histogram of the number of words produced by the individual students in the intervention group at the Midterm measurement point.

Histogram of words produced at the Spring measurement point
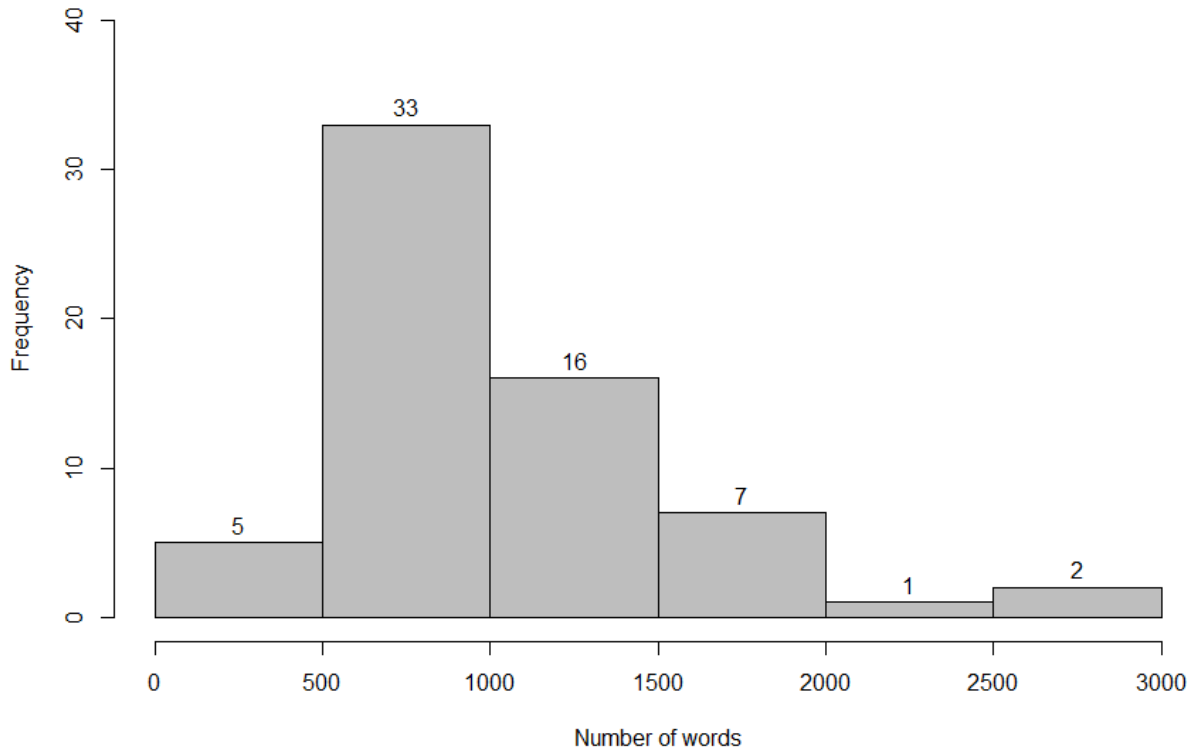
Figure 39. Histogram of the number of words produced by the individual students in the intervention group at the Spring measurement point.

11.2.2.2 Error scores

The mean error rate for the intervention group is 7.54% while the median is 7.47%, so the distribution should be close to normal (only two students score above 15%). The Shapiro-Wilk test run in R confirms that the total error scores are normally distributed ($W = 0.98$, $p = 0.51$):[22]

---

22 Note that total scores are calculated across all texts each student submitted during the school year, so they obscure any potential development during the school year.

Histogram of total error scores
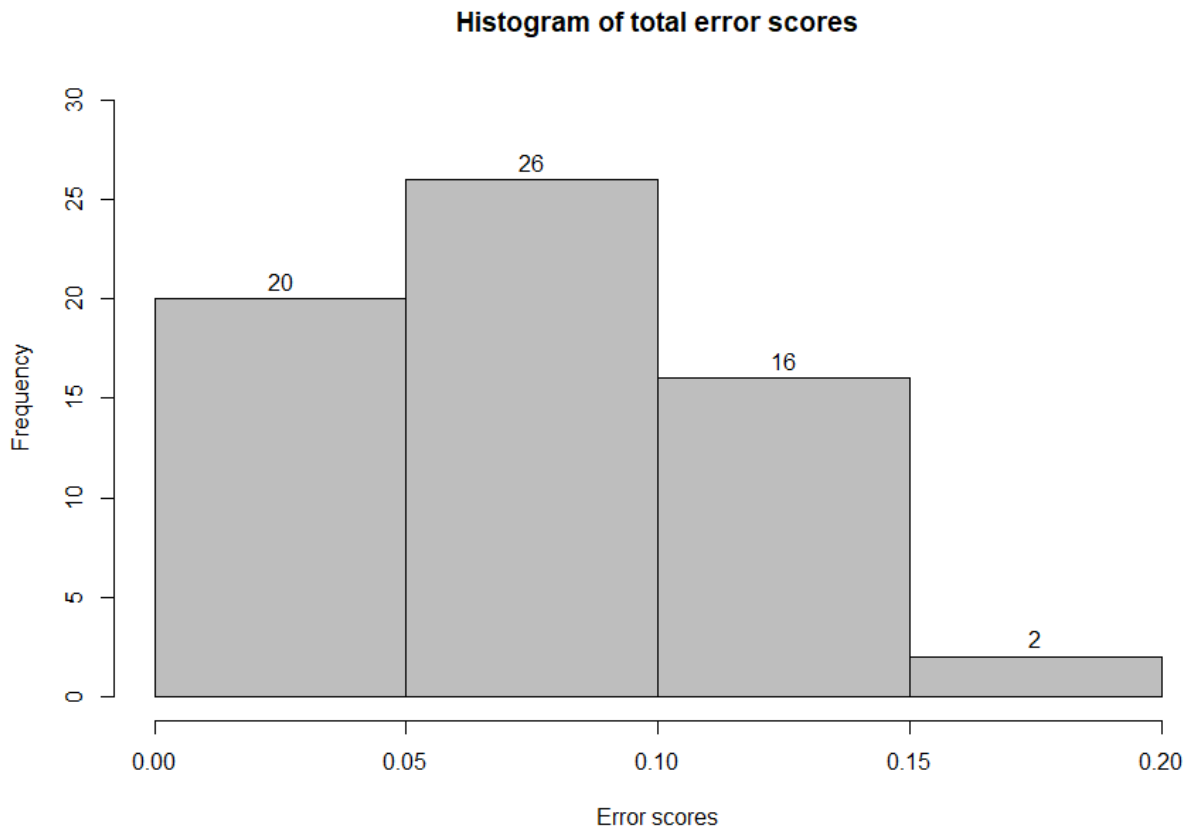
Figure 40. Histogram of the total error scores for the individual students in the intervention group.

As with the comparison group, the texts of the intervention group students are divided into three measurement points for each student: Fall, Midterm, and Spring. Error scores are calculated after the texts were collapsed into one file for each measurement point. When the error scores are divided into the individual measurement points it seems that the students from the intervention group behave differently than the students from the comparison group. The average error scores in the Fall and in the Spring appear to be similar with many students scoring under 5% of errors. The average scores in the Midterm are closer to a normal distribution and resemble the scores from the comparison group.

The distribution of the error scores in the Fall is not normal ($W = 0.91, p < 0.001$). Many students score under 5% errors. The mean for the Fall is 7.09%, and the median is 5.07%. Six students score above 15%:

Histogram of Fall error scores
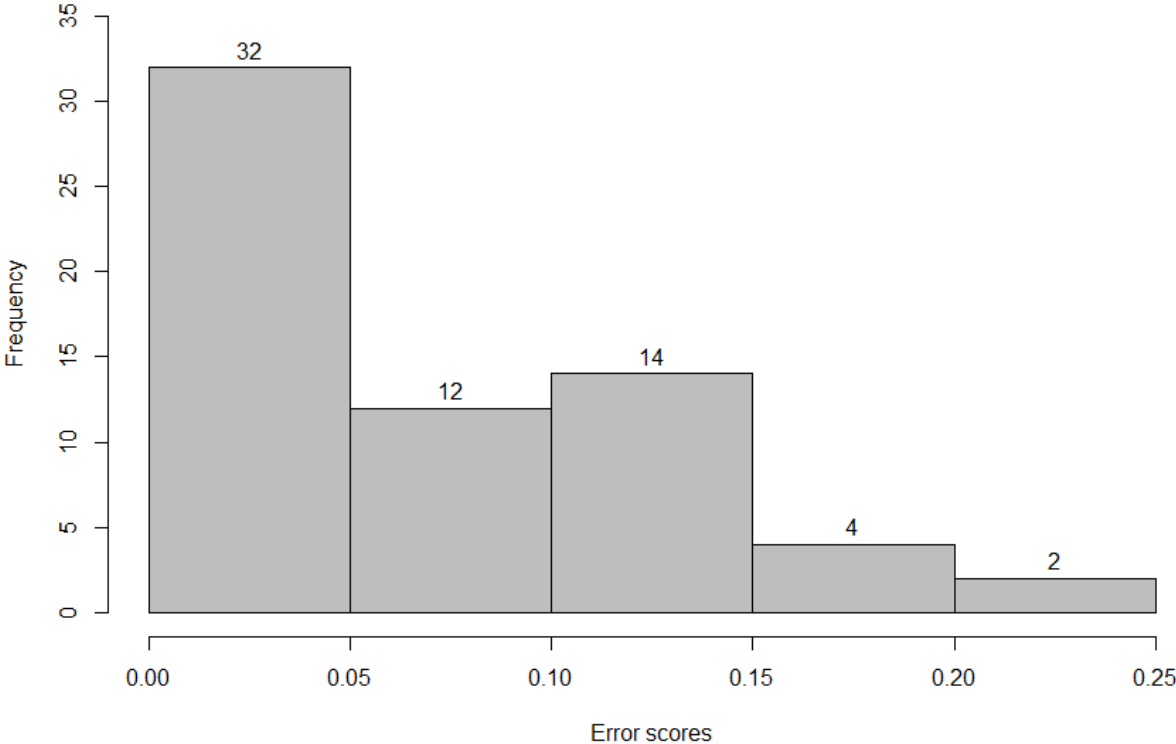
Figure 41. Histogram of the error scores of the intervention group at the Fall measurement point.

The distribution of errors in the Midterm changes compared to the Fall, and even though it appears positively skewed in the histogram, the Shapiro-Wilk test reveals that the error scores in the Midterm are normally distributed ($W = 0.97. p = 0.10$). The mean for the Midterm is 7.63%, and the median is 7.16%. Two students score above 15%:

Histogram of Midterm error scores
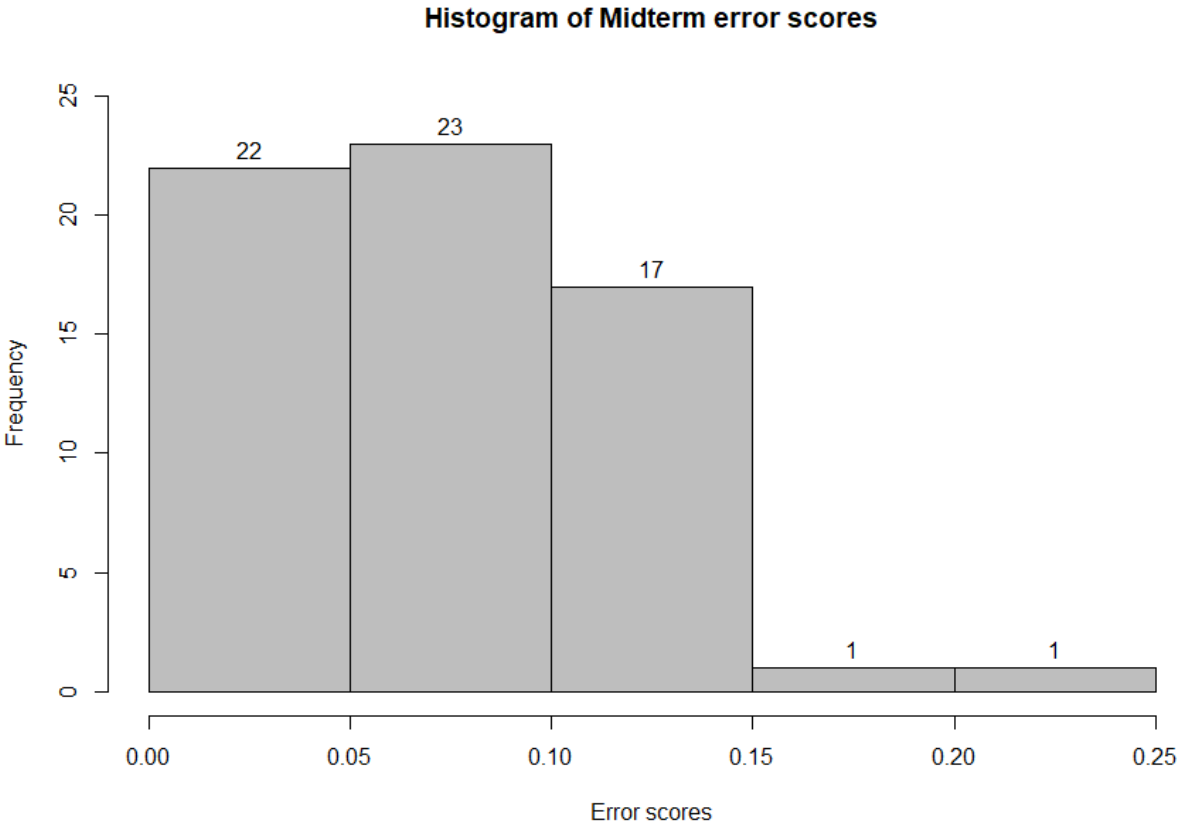
Figure 42. Histogram of the error scores of the intervention group at the Midterm measurement point.

The mean for the Spring is 7.58%, and the median is 6.59%. The distribution is not normal ($W = 0.93$, $p = 0.001$). There are many students scoring under 5%, but also an increased number of students with many errors. Eight students score above 15%:

Histogram of Spring error scores
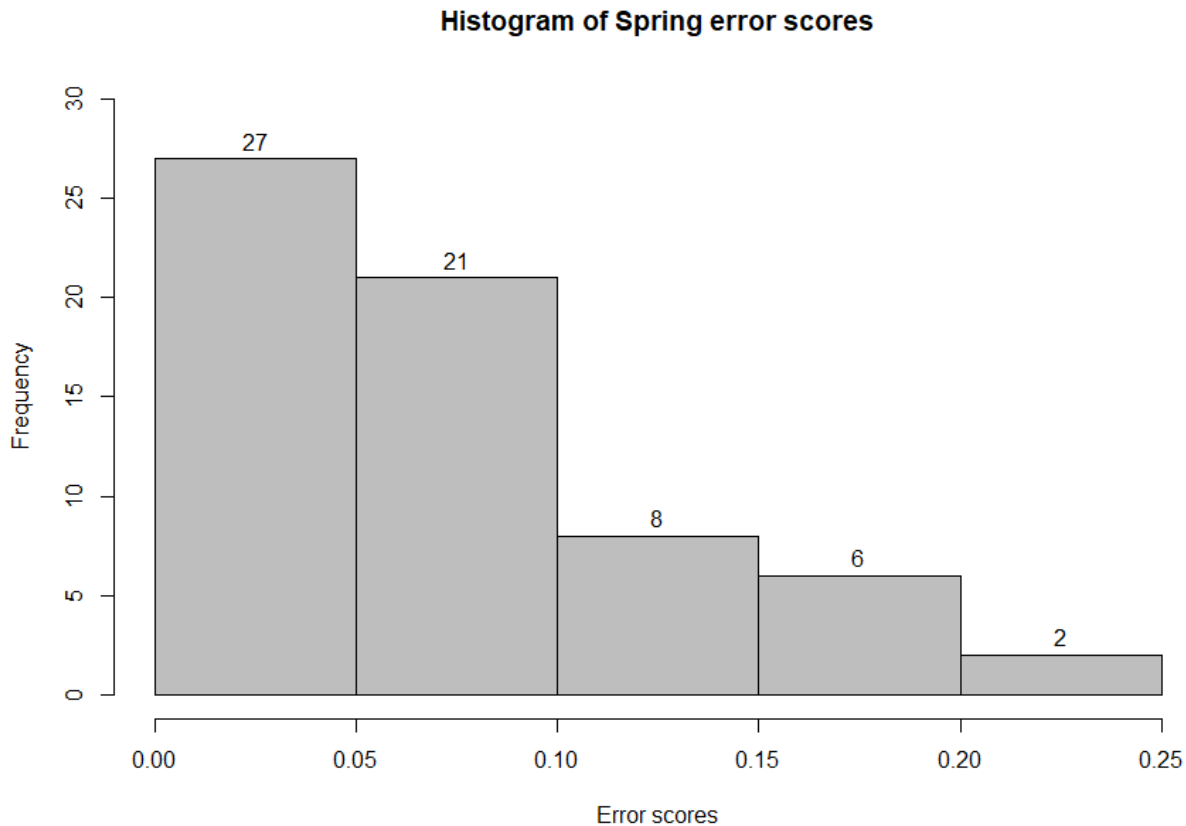
Figure 43. Histogram of the error scores of the intervention group at the Spring measurement point.

It seems that there are many students in the intervention group with very few agreement errors, but also several students with many agreement errors. Only one of the outliers at the three measurement points is the same student at more than one measurement point. This again, as with the comparison group, suggests that the outliers might be students who did not perform well the day when the test was written and who otherwise score better than what the data might suggest. When the individual scores are examined in detail, only one student (16STV33) has high scores throughout the collection period, but 16 students score above 10% in the Spring. It seems that many students produce more errors towards the end of the school year than in the beginning.

160

As is the case with the comparison group, the intervention group also has a high degree of individual variation in the error scores, ranging from 0% to 25% errors. However, it seems that there are more students scoring between 10% and 20%, but fewer scoring above 20% than in the comparison group. There are five students with error scores equal to or above 20% at at least one of the measurement points, but none who would score above 30% (as compared to three such students in the comparison group). The plot in Figure 44 illustrates the error score distribution in the intervention population between the three measurement points:

Student error scores in three measurement points
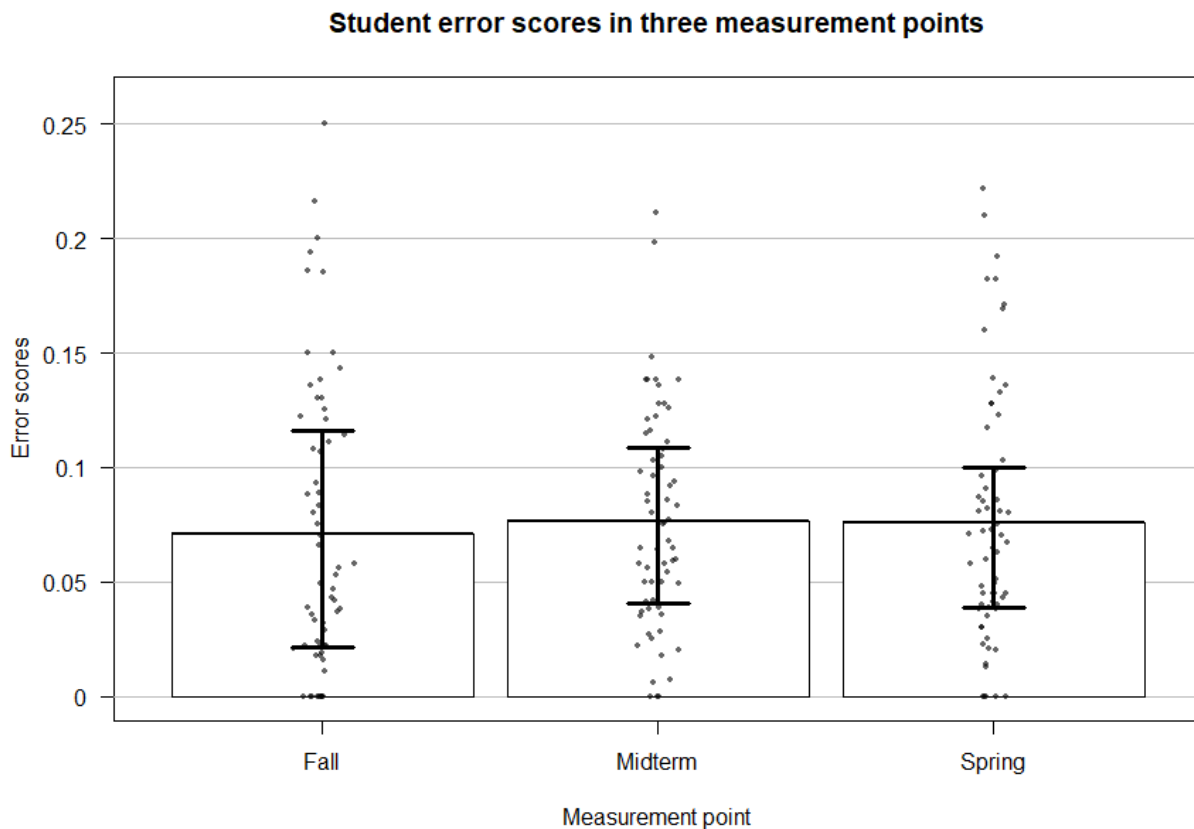
Figure 44. Distribution of the individual scores of the intervention group at each of the three measurement points.

As with the comparison group, the spread of the data around the mean changes during the school year, but there do not seem to be are any group-wise trends. Since the error scores are not normally distributed at two of the measurement points, Friedman's ANOVA was used in R to evaluate the differences between the three points. There are no significant differences between the scores at the three measurement points when the

161

whole group is considered as one population ($\chi^2(2) = 1.84, p = 0.4$). This means that, as in the comparison group, the students in the intervention group did not improve their accuracy in agreement marking to a statistically significant degree during the collection period.

11.2.2.3 Developmental trajectories

The developmental trajectories described in section 11.2.1.3 are the same also for the intervention group. Trajectories 1-3 represent students who finish the school year with higher accuracy than they had in the beginning, while trajectories 4-6 represent students who finish the school year with lower accuracy (see section 11.2.1.3 for more detail). The distribution of the students from the intervention group in the developmental trajectories:
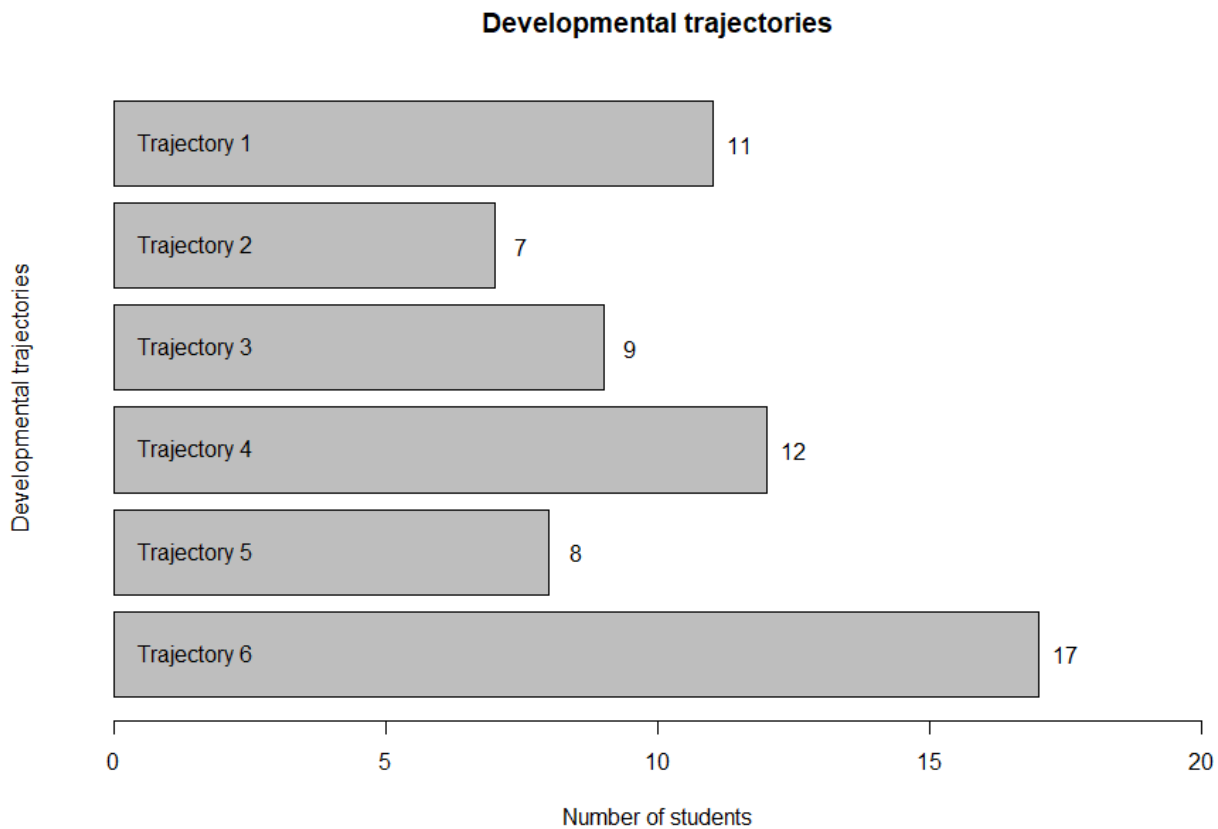
Developmental trajectories

Figure 45. Developmental trajectories of the students in the intervention group.

Figure 45 shows that the majority of the students make more agreement errors towards the end of the school year than when they started (37 students in trajectories 4-6 compared to 27 students in trajectories 1-3). Again, part of the explanation for this trend

162

could be that they are writing more complex texts towards the end of their school year. The median length of the texts in words at each of the three measurement points in the intervention group is as follows:

Median length Fall = 517.5 words

Median length Midterm = 1452 words

Median length Spring = 959 words

There is no continuous increase in the text length throughout the school year (as was the case in the comparison group), but the error scores at the three measurement points follow the development, i.e. the higher median length of the texts leads to higher error scores. Median error scores for the intervention group at the three measurement points are repeated below:

Median errror score Fall = 5.07%

Median error score Midterm = 7.16%

Median error score Spring = 6.59%

It is not surprising that the length of the text should have an influence on how many agreement errors the students produce. However, it is clear that the students in the intervention group, like the students in the comparison group, are not yet able to consistently mark subject-verb agreement in their written production at the end of their compulsory English intstruction.

11.3 Statistical evaluation of the results

As shown in chapter 10.3, the comparison group and the intervention group are comparable regarding their demographics, their exposure to additional languages, and their attitudes towards learning languages. As the school, study track, and in-class teacher are controlled for, the only factor influencing the development of the accuracy in agreement marking should be the extra exposure to explicit grammar teaching via the Inverted Classroom course. However, as discussed in chapter 10.6.7, the actual use of the material was unexpectedly low in the intervention group and the students did not use the course as intended. In addition, there is always some degree of individual variation, even in comparable populations, but this variation would influence both the groups and should not play major role as long as the starting scores of the students are comparable. Both the starting conditions and the development of the students' agreement marking accuracy are compared in the following sections.

11.3.1 Word production

The first condition for the validity of the comparison is an evaluation whether the students in the two groups produced comparable amount of text and whether their texts contained comparable amount of agreement marking occasions. The comparison group wrote marginally fewer words than the intervention group (191,555 words vs. 203,159 words), and they also produced fewer clauses where agreement can be overtly marked (15,632 vs. 16,804 occasions). The comparison group students thus produce an agreement marking occasion every 12.25 words while the intervention group students produce an agreement marking occasion every 12.09 words.

As shown in sections 11.2.1.1 for the comparison group and 11.2.2.1 for the intervention group, the total word production is normally distributed in both groups. An independent-means t-test run in R (Field et al., 2012, pp. 368-385) shows that there is no significant difference between the word production in the two groups ($t(121.84) = -2.33, p = 0.11$). Despite the small differences in total numbers, it is clear that both groups write approximately the same amount of text during the school year. Figure 46 illustrates the estimated word production density plotted for both populations:
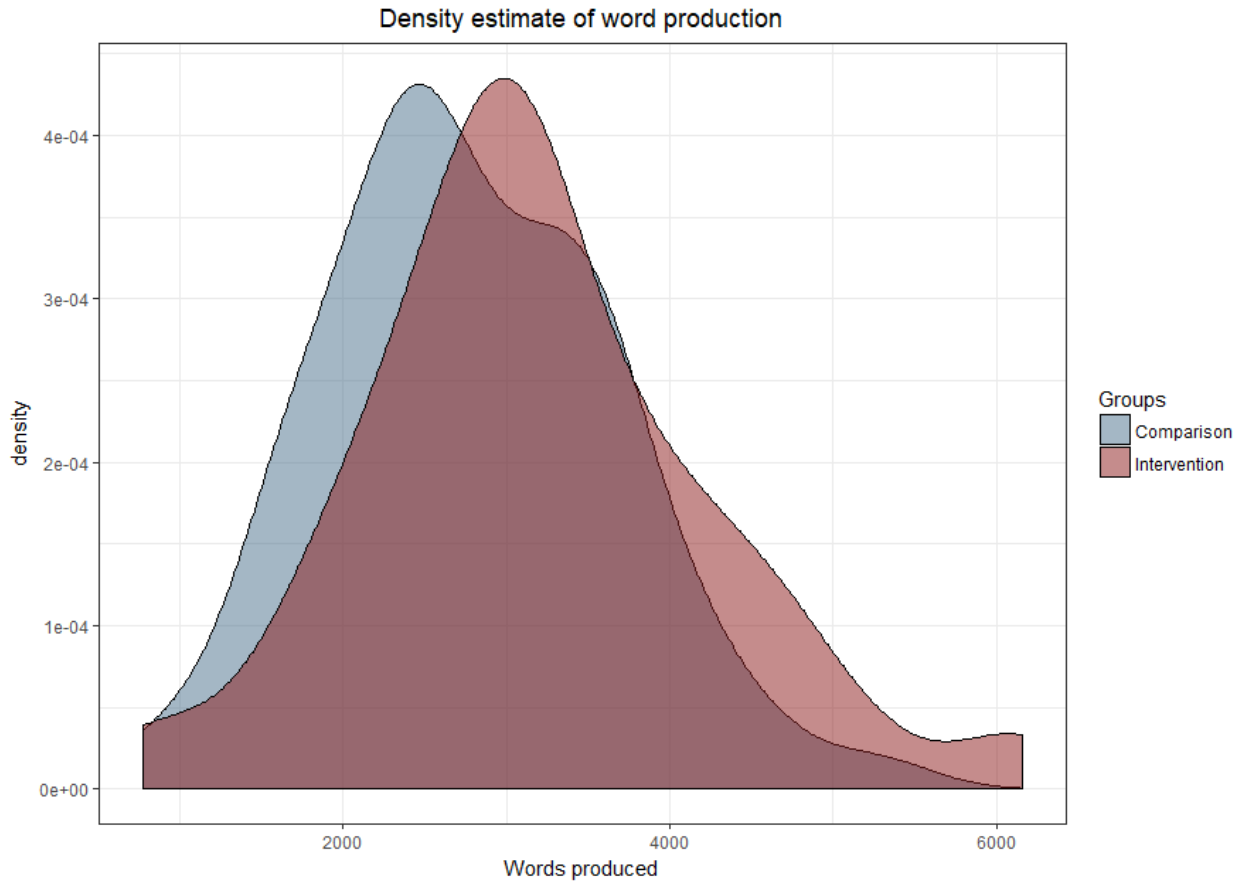
Figure 46. The estimated density of word production in the two groups.

11.3.2 Total scores

There are limitations to considering the total error score data, as such scores obscure the differences between the students who start with low error scores and do not improve much and the students who start with high error scores but improve significantly during the school year (see sections 11.2.1.2 and 11.2.2.2). The mean and median scores for the comparison and the intervention groups are summarized in Table 17:

Table 17

Total average error scores of the comparison and the intervention group

Group	Mean	Median
Comparison	8.41%	7.98%
Intervention	7.54%	7.47%

165

Both the mean and the median of the comparison group are higher than the mean and the median of the intervention group. However, these differences are not statistically significant. Since the total error scores of the comparison group are not normally distributed, a Wilcoxon's rank-sum test was run in R (Field et al., 2012, pp. 655-666): there are no significant differences in the total scores for the two groups ($W = 2476, p = 0.23$). Figure 47 illustrates the estimated error scores density plotted for both populations:
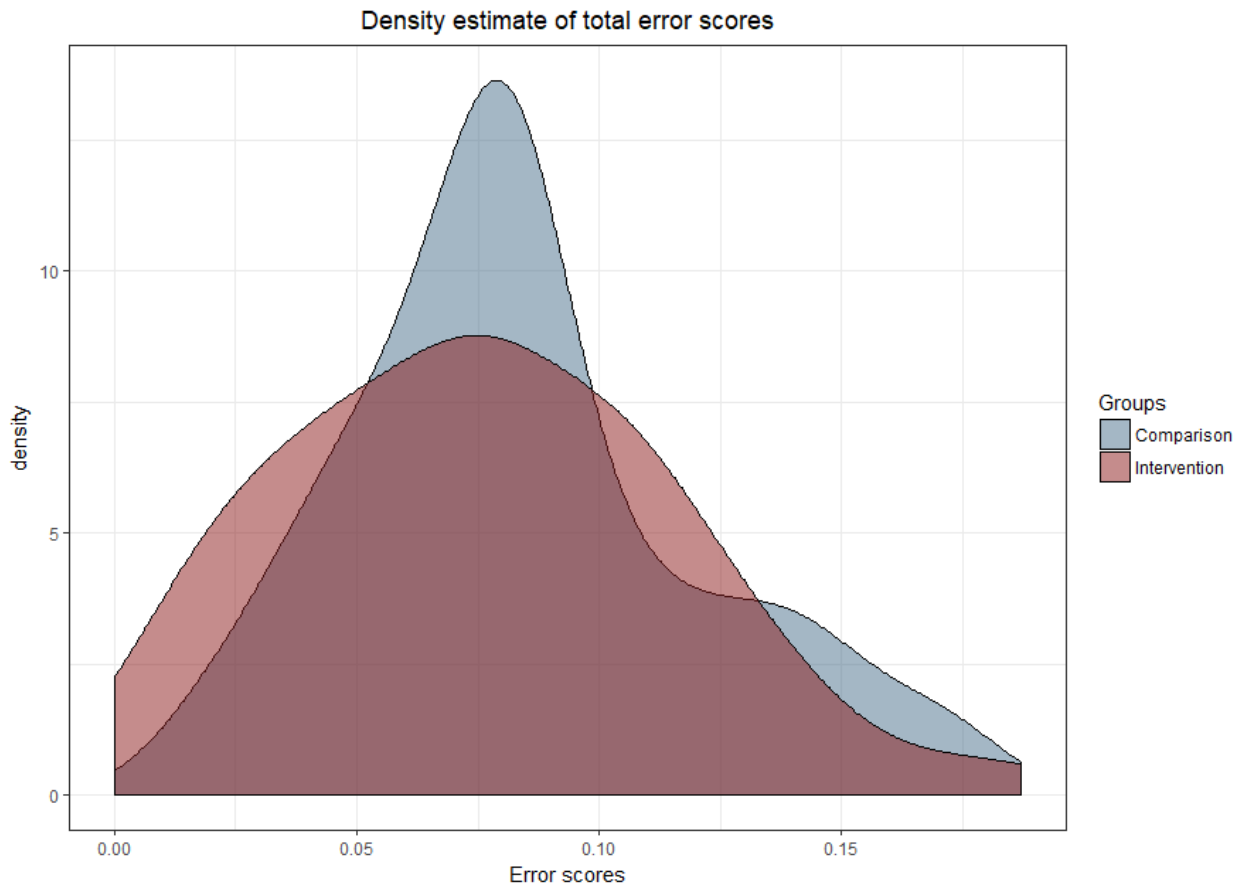
Figure 47. The estimated density of total error scores for both groups.

It seems that the error scores of the students in the intervention group show more individual variance than the error scores of the students in the comparison group. In other words, there are many students who score between 5% and 10% in the comparison group and only few who score under 5% or over 10%, while in the intervention group, the student error scores are more spread between 0% and 15%. However, there are no statistically significant differences in the distribution of total error scores.

11.3.3 Fall scores

An essential question for the evaluation of the differences between these two groups of students is whether their initial accuracy (at the beginning of the school year) is approximately the same. The scores at the Fall measurement point are summarized in Table 18 below. Both the mean and the median of the comparison group are higher than the mean and the median of the intervention group. In addition, the means are more than two percent points higher than the medians which suggests that there are many students with low error scores and only few outliers with high error scores in both groups:

Table 18

Average error scores of the comparison and the intervention group at the Fall measurement point

Group	Mean Fall	Median Fall
Comparison	8.67%	6.25%
Intervention	7.09%	5.07%

In the Fall, the students from the intervention group have lower average scores than the students from the comparison group. In addition, both groups show similar strong positive skew on their error scores, i.e. there are many students with low error scores and only few outliers with very high error scores. However, there are no statistically significant differences between these two groups when their Fall error scores are tested with a Wilcoxon's rank-sum test ($W = 2463.5$, $p = 0.25$). It is thus safe to conclude that both groups start at approximately the same level of accuracy in agreement marking. Figure 48 illustrates the estimated error scores density for the Fall measurement point plotted for both populations:

Density estimate of Fall error scores
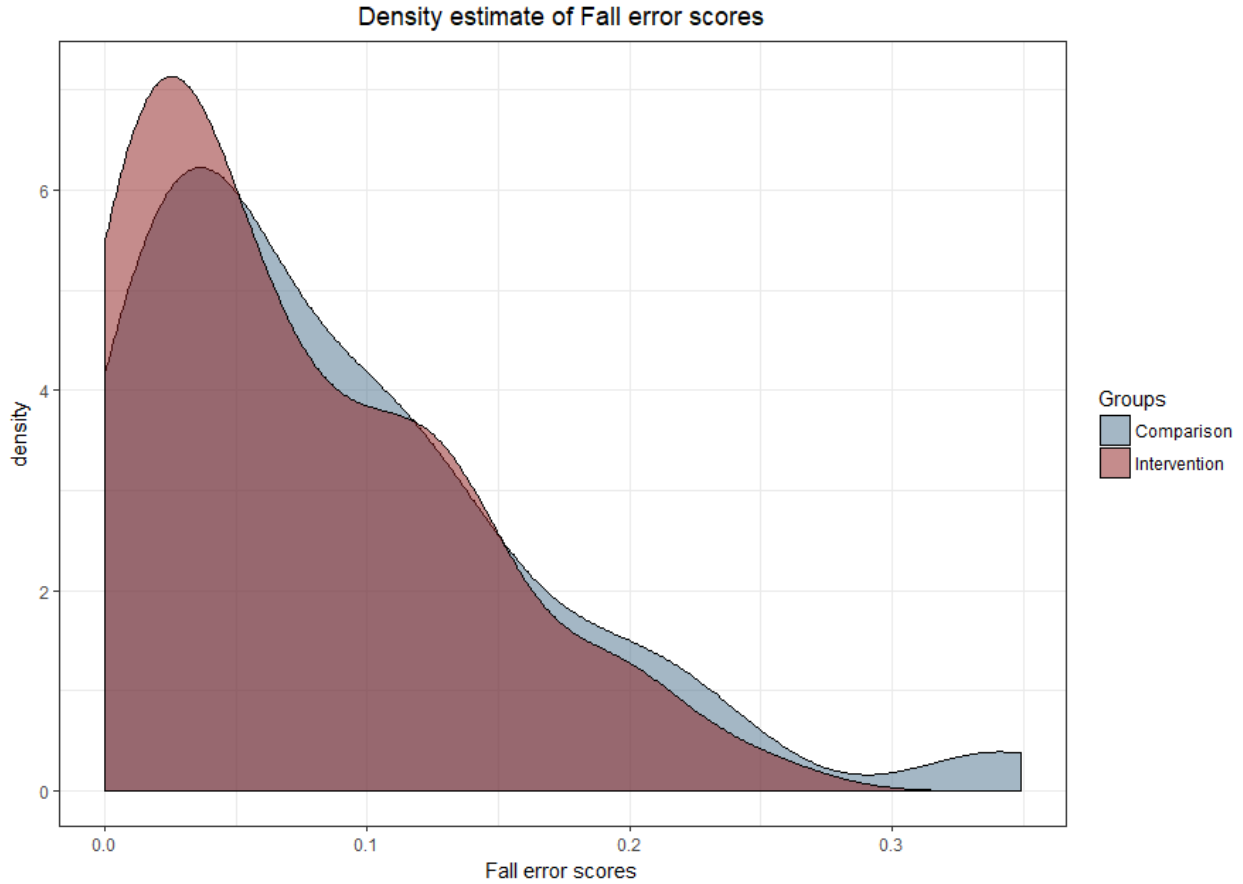
Figure 48. The estimated density of Fall error scores for the two groups.

11.3.4 Spring scores

Since the two groups start at a comparable level of accuracy in agreement marking, the next step is to compare their scores at the end of the school year. The Spring error scores for both groups are summarized in Table 19. The mean and the median of the intervention group are still lower than the mean and the median of the comparison group, but the differences are smaller than in the fall:

Table 19

Average error scores of the comparison and the intervention group at the Spring measurement point

Group	Mean Spring	Median Spring
Comparison	7.73%	7.32%
Intervention	7.58%	6.59%

The median scores for both groups increased from Fall to Spring. The mean for the intervention group increased as well, while the mean for the comparison group decreased, but the differences are small. The error scores in neither of the groups are normally distributed, but the scores in the intervention group are more positively skewed than in the comparison group, i.e. there are more students with very low error scores in the intervention group than in the comparison group. However, the Wilcoxon's rank-sum test shows that the differences are not significant ($W = 2343, p = 0.54$). Figure 49 illustrates the estimated error scores density for the Spring measurement point plotted for both populations:
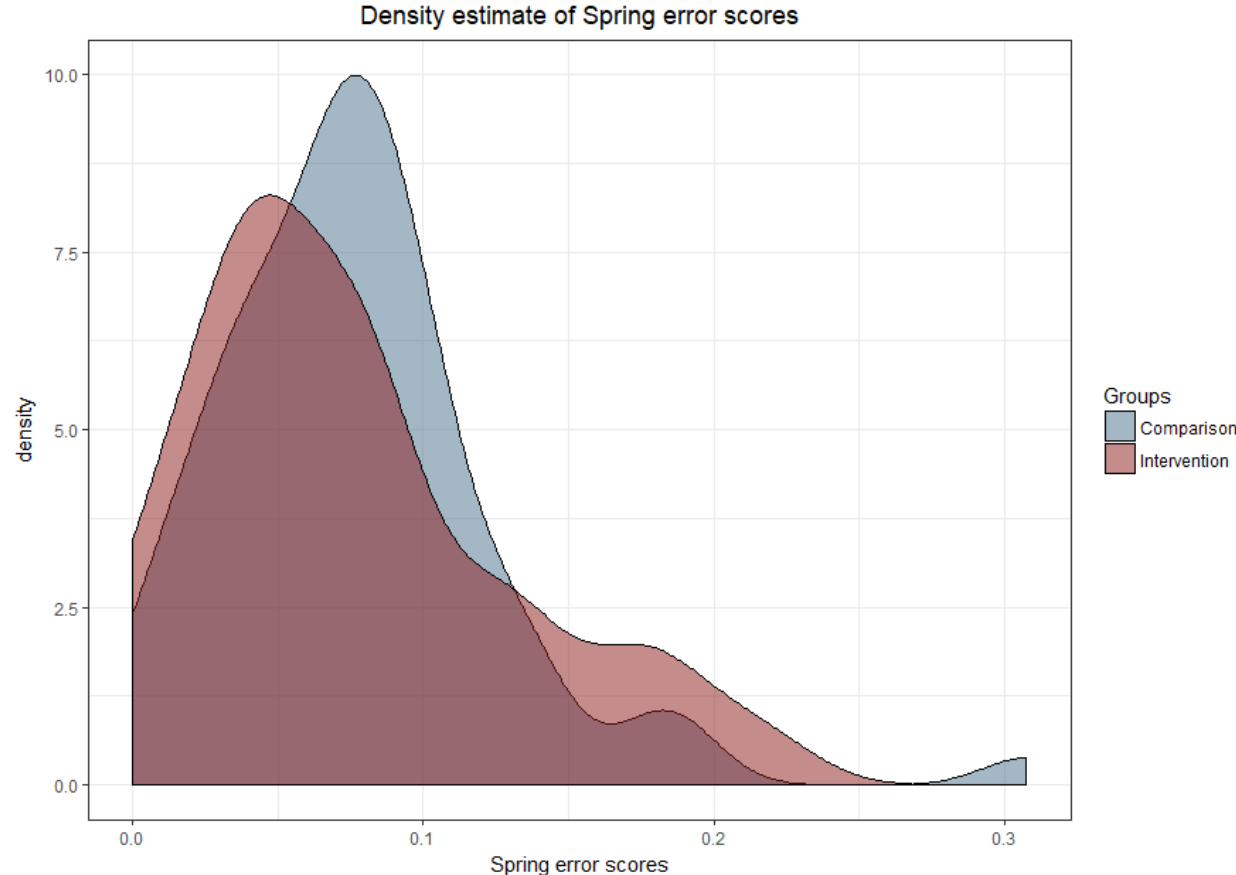
Figure 49. The estimated density of Spring error scores for both groups.

There are more students with low error scores (under 5%) in the intervention group than in the comparison group, but there are also more students with high error scores (over 15%) in the intervention group than in the comparison group. There is one outlier in the comparison group with a Spring error score over 30%.

11.3.5 Development during the school year

Finally, the difference in Fall and Spring scores is calculated for each group of students, i.e. how much has the accuracy of each student improved during the school year. The difference in scores is calculated as Fall score minus Spring score, so the students with positive difference made fewer errors in the Spring than in the Fall, while the students with negative difference scores made more errors in the Spring than in the Fall.

In the control group, 33 students (47.83%) have worse scores in the Spring than in the Fall, one has the same score, and 35 students (50.72%) have better scores in the Spring than in the Fall. In the intervention group, 36 students (56.25%) have worse scores in the Spring than in the Fall, two have the same scores, and 26 students (40.63%) have better scores in the Spring than in the Fall. Even though the intervention group has a marginally lower average score at all measurement points, fewer of the students improve during the school year. The Wilcoxon's rank-sum test (the difference scores for the comparison group are not normally distributed) reveals that the differences are not significant ($W = 2400.5$, $p = 0.39$). Figure 50 below shows the estimated density of the differences between the Fall and the Spring measurement point error scores plotted for both populations. The estimated densities of the differences between the Fall and Spring error scores in the two groups are almost identical:
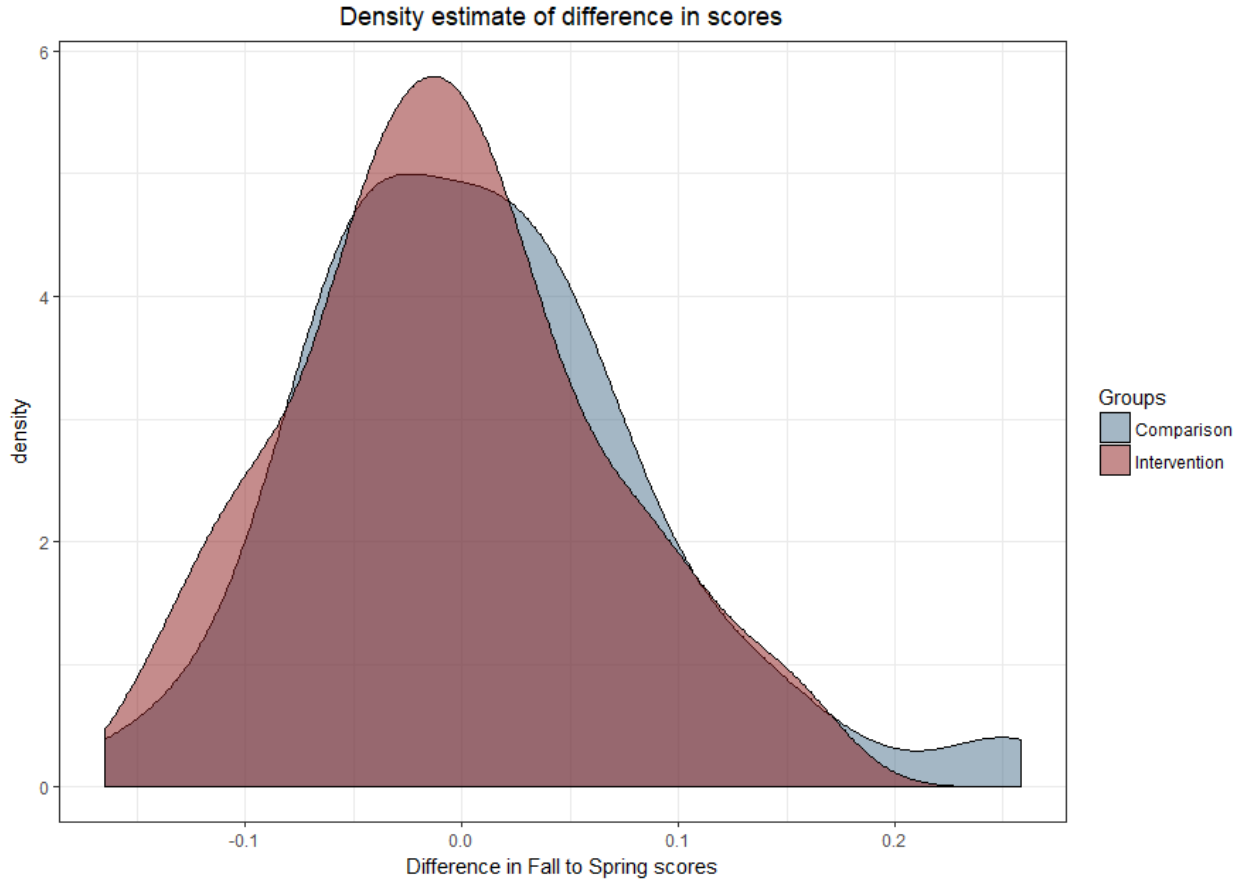
Figure 50. The estimated density of the difference in scores shows for both groups.

11.3.6 Summary of the statistical evaluation

Despite the differences in average scores, there are no statistically significant differences between the comparison group and the intervention group, neither in their Spring scores, nor when the difference in scores from Fall to Spring is calculated. In both groups, some students improve their accuracy during the school year, some stagnate, and some make more errors at the end of the school year than in the beginning (which might be caused by their attempt to write longer and more complex clauses). In conclusion, the intervention did not have any significant effect on the agreement marking accuracy of the students. On the other hand, as is mentioned in chapter 10.6.7, with such a low actual use of the intervention material (mean time spent on the modules dealing with agreement was 15.52 minutes, median time was only 7.5 minutes) it would be surprising if there were an effect. The shortcomings of the intervention design and possible remedies are discussed in the following chapter.

171

12 Evaluation of the methodology

There are several advantages of using the Inverted Classroom Methodology in a teaching intervention. First and foremost, there is a higher level of control over the teaching the students receive as compared to providing the in-class teacher with instructions or work sheets to work with. Second, the use of the Inverted Classroom allows the students to replay the 'lecture' as many times as they need instead of hearing the teacher's explanation only once in the class. Third, the use of the Inverted Classroom frees the teacher to actively help the students in the classroom instead of using the classroom time for blackboard teaching. The classroom time can be used for practice, discussion, and other activities in which the students can assume an active role. This distribution of activities breaks with the traditional PPP model since the presentation and partially also the controlled practice parts are moved into the homework domain and only the production part is left in the classroom. Students can then receive assistance when they need it most: when they are supposed to use their newly acquired knowledge in practice – be it written or spoken production. Last but not least, this methodology can be used to teach students who struggle and students who are on a more advanced level than their peers within the same classroom. The teacher can create adapted lectures or direct students to other lectures that meet their needs.

The Inverted Classroom Methodology promotes cooperation and sharing of materials among teachers across schools and areas as the lectures with exercises are accessible online and can be made accessible with or without user authentication. Campus Inkrement (Thue, 2016) currently offers several public courses in math, physics, natural science, physical education, and economics for students in elementary, junior high, and high schools in Norway. The English grammar course developed in this project will be added to the public course bank and made available for all users under the creative commons license. In this chapter, I first evaluate the inverted English grammar course from the perspective of the author, the teachers, and the students. As the results of the statistical analysis are not significant the final part of this chapter focuses on the shortcomings of the collected data and the intervention procedure and suggests improvements for future research.

12.1 User-friendliness

The general usability of the platform was perceived as good by both the teachers and the students. The teachers could choose to either follow the order of topics in the course or pick topics which were relevant for their year plan. Both approaches were used by the participating teachers. The Campus Inkrement platform also enables the users to adjust the content of the courses to their needs. One can use a ready-made course or clone, change, and add content to it. The teachers participating in the intervention were asked not to change the content beyond rearranging the order of the topics. After the data collection was finished, the course was also presented at a workshop for teachers and the cloning and adjusting process was perceived as complicated. It is thus expected that a wide implementation of this or similar course would require some initial training for the teachers. On the other hand, the teachers who do not want to invest the time into adjusting the course or creating content themselves can use the course in its current form as a supplement for their own teaching with only minimal initial effort.

12.1.1 Usage by teachers

All teachers who expressed interest in the project were informed about the methodology and the process of the intervention at a common meeting at each participating school. Those who chose to participate were then offered help with the course setup. Some teachers struggled initially with the course setup and with adding the students to the course room. I have created two instructional screen capture videos which explained the procedure and visited two of the teachers to help them set up the course. After the initial setup, there were no technical issues reported by any of the teachers. The teachers were contacted again after one semester in order to collect consent forms from the students. At this occasion the teachers were informally asked about any issues or comments they have. At the end of the second semester the teachers were contacted again for collection of the student papers. They were again asked about the experience with the course. The log data were acquired retrospectively via the teachers after the discrepancy between the reported use and actual use became evident.

All the teachers reported asking the students to work with the course at home as intended. However, it seems that most of them did not check whether the students completed the homework assignments (see chapter 10.6.7). One teacher expressed this concern also before the intervention. According to this teacher, his students never do their

homework and thus he wanted to use the videos in the classroom instead (this teacher chose not to participate in the intervention after the procedure was explained). While there is certainly room for adjustment to the individual needs of the students, using the videos in the classroom instead of at home removes one of the advantages of the Inverted Classroom approach, i.e. the option of using classroom time for more student-centered activities instead of teacher-centered lectures. Handke (2013a, p. 15) suggests that as with any other teaching and learning scenario, also with the Inverted Classroom Methodology, there is limited control over what the students do with the material that is provided to them. In other words, the degree to which the students actually do their homework is always an uncontrollable factor.

Another issue discovered during the data analysis is the lack of a built-in software structure which would hinder the students from proceeding if they have not watched the video or have not attempted to do the exercises. Some Inverted Classroom course contain such features and Handke (2013b, p. 66) considers them as essential to the success of the inverted method. The review of the log data revealed that several students did not watch the lecture or the exercise feedback videos. Some of these students were clearly strong students who completed the exercises correctly despite not watching the video, but there were also cases of students who had many errors in the integrated exercises and nevertheless did not watch the feedback videos to find out what was the problem with their answers. Such an approach, of course, compromises the intention of the feedback videos. With the current version of the Campus Inkrement software, the only remedy would be a thorough follow-up of the homework by the teacher. Only one of the involved teachers reported that she regularly discussed the homework in class with her students.

12.1.2 Usage by students

The students were not interviewed directly, but some of them chose to leave comments in the user logs and some comments were sent to me from their teachers. They have also answered questions about the course in the second part of the background questionnaire (see Appendix 4). The comments from the logs and student comments sent from teachers were mostly about exercises which did not work properly. These were corrected as soon as they were reported. There are certain limitations for self-correcting exercises and the experience shows that students often do not follow instructions carefully enough and then get frustrated when their answers are marked as incorrect due to typographical issues such

as missing full stop and similar. However, careful reading of instructions is a useful skill to learn and with the option of redoing each exercise as many times as each user wants the rigidness of the self-correcting system should not be a problematic feature.

One of the advantages of the Inverted Classroom Methodology is the option of self-pacing and re-watching the lectures, which is not possible with a traditional teaching model where the teacher explains new material in the classroom. There is only little research done on the actual use of self-pacing within the Inverted Classroom Methodology. Owston, Lupshenyuk & Wideman (2011, p. 265) report that low-achieving students use the recordings often while high-achieving students use them rarely and often fast-forward only to some sections. There is no such correlation detected in the survey data from my intervention. If the students are divided into four quartile groups according to their error scores in the spring, i.e. the error rate at the end of their obligatory English instruction, three of the groups report almost identical usage frequency: three to four students report using the course often, 13-21 students report using the course sometimes, three to five report using the course rarely, and zero to one never. Only among the students whose error scores fall within the range of 5%-10% errors, there is a majority who reports using the course rarely or never (13 and three students respectively) and a minority who reports using the course often or sometimes (one and seven students respectively). It is of course not given, that the students who score low on agreement errors can be defined as high-achieving students and vice versa. The overall grades in English were not collected in the project so Owston et al.'s conclusion cannot be further validated. However, there might be a potential in testing tailored versions of the course on different learner groups based on their initial knowledge of the subject as is suggested by the Cognitive Load Theory (see R. C. Clark et al., 2006, pp. 247-273). According to this approach the strategies which work best for beginners can have only small or even reversed effect on advanced learners (see also Abeysekera & Dawson, 2015, p. 9). If this would be the case, the Inverted Classroom course could be used for specific groups of students according to their needs.

Lai & Hwang (2016) report that the group which received specific instruction in how to use the Inverted Classroom material efficiently, e.g. individualized plans and goals, performed better than the group which only received the materials and was only asked to use them. The students in my project did not receive any additional instructions on how to use the materials. Only five students report to have re-used the lectures often

and additional 21 report having re-used the lectures sometimes. The user logs reveal that only two students re-watched the whole lecture on agreement. In the light of the results, an additional instruction in how to use the Inverted Classroom course effectively seems to be one of the factors which could improve the overall effect of the method.

The students were also asked whether they liked the Inverted Classroom Method and whether they felt that they learned more in this way as compared to the traditional blackboard teaching (see questions three and four in Appendix 4). The majority of the students enjoyed the new teaching form (100 out of 118 answers) but they were uncertain whether it helped them learn or remember better (only 8 students felt they learned much more, 49 students answered they felt they learned a little bit more, 50 were uncertain, and 11 felt they did not learn more using the inverted course). Out of the 118 students who answered this part of the questionnaire, 45 stated that they have tried Inverted Classroom in other subjects before (mainly math), but for 73 students this was the first encounter with this methodology. 45 of these 73 students expressed that they would like to try the Inverted Classroom method in other courses.

12.2 Evaluation of the intervention procedure

As discussed in chapter 11, the intervention did not have the expected effect on the student error scores. However, the importance of linguistically informed ISLA practices and research-based didactic innovations should not be underestimated. The results from the corpus exploration (chapter 5) show that Norwegian students have problems with agreement marking in English and that the non-standard agreement pattern is perhaps fossilized for many of the students. This finding suggests that subject-verb agreement should be targeted more or differently in the English instruction in Norwegian schools. Results from several studies in different disciplines and in various educational settings show that the Inverted Classroom methodology can be beneficial both for student satisfaction and their results. The question which remains is whether the procedure applied in this project can be improved in such a way that would yield the desired effect in the high school English classrooms.

Kim, Kim, Khera & Getman (2014) review Inverted Classroom practices in three disciplines offered at the same university (within engineering, social studies, and humanities) and suggest the following nine design principles of a successful inversion:

1. Provide an opportunity for students to gain first exposure prior to class.

2. Provide an incentive for students to prepare for class.
3. Provide a mechanism to assess student understanding.
4. Provide clear connections between in-class and out-of-class activities.
5. Provide clearly defined and well-structures guidance.
6. Provide enough time for students to carry out the assignments.
7. Provide facilitation for building a learning community.
8. Provide prompt/adaptive feedback on individual or group works.
9. Provide technologies familiar and easy to access.

Even though the study was conducted at a higher educational level, the findings can be applicable also in the further work with the English grammar course for Norwegian high school students. Most of the principles from Kim et al. were observed in the creation of the course, but there is definitely room for improvement in the intervention design regarding points four, five, and seven.

The main weaknesses of the procedure applied in this study seem to be the lack of homework follow-up in the classroom and the lack of connection between the in-class and out-of-class activities (Kim et al.'s Point 4). Only one of the teachers reported regularly building on the video lectures in the classroom. Spannagel & Spannagel (2013, p. 113) stress that quality video lectures are only one part of the Inverted Classroom methodology and a student-centered use of the classroom time is equally important to the success of the method. This aspect was unfortunately underestimated in my project as I relied on the teachers to engage in language-focused students-centered activities in the classroom without a specific instruction from me. Better implementation of additional language-focused instruction in the classroom would probably also facilitate building of a learning community among the students (Kim et al.'s Point 7).

Lai & Hwang's (2016, see the summary in chapter 9.5.2) study suggests that students may benefit from a clearly defined and well-structured guidance (Kim et al.'s Point 5) in addition to having access to the materials. Lai & Hwang's subjects were elementary school students and my assumption was that perhaps high school students are mature enough learners to be able to create learning strategies for themselves. However, this was not the case and the recommendation for further research is to incorporate the grammar course more into the year plan of the class and require that homework is attended to by all students. The comparison of the reported use and actual use (chapters 10.6.6 – 10.6.8) clearly shows that only few students worked with the course as intended.

The comparison of student-reports and user logs is also recommended by Owston et al. (2011) to improve the validity of any distance learning procedure so this feature needs to be retained in further research.

A creation of a grammar course to test the Inverted Classroom methodology was chosen partially because grammar teaching is often perceived as theory-heavy and too teacher-centered. Grammar topics thus seem good candidates for video lectures. However, Prodoehl (2016) in her study suggests that the Inverted Classroom methodology in a second/foreign language classroom does not need to be limited to the technicalities of the language itself. She suggests that video lectures can be used to introduce any information-heavy topic and serve as a preparation for a student-centered discussion in the classroom. The English grammar course can thus be seen as a first step in the development of a complex Inverted English course for Norwegian high school students in which also mini video lectures on topics in history, culture, and literature could be included.

13 Summary and suggestions for further research

Part II of this thesis described a didactic intervention which aimed at increasing the accuracy of agreement marking in L2 English by enhancing the explicit knowledge of some grammar concepts and raising the metalinguistic awareness of L2 English students. An Inverted Classroom course in English grammar was developed and tested on a group of first-year high school students following the last year of their obligatory English instruction. The intervention did not yield a significant increase in accuracy in the intervention group because, as was later discovered, a majority of the students did not use the materials as intended.

The Inverted Classroom Method has previously been tested in higher education and usually within the STEM subjects. Despite reported use, there are only few studies which have tested the effects of the Inverted Classroom Method in lower education and in different subjects. This study targeted such a population: high school students in a second language class, which contributes to widening the scope of the method. Even though the intervention did not lead to significant results, there are good indications regarding the problematic areas which need to be targeted in a potential follow-up study. High school students require closer supervision than was expected and perhaps also an introductory information session about the method in order to increase their motivation to use the materials. It might also be beneficial to monitor the use of the course and guide the participating teachers during the intervention.

Nevertheless, the intervention study resulted in a functional didactic tool which can be developed further both by individual teachers and other researchers. The value of creating a linguistically informed didactic tool lies mainly in the possibility to target specific problematic areas of the target language. The inverted English grammar course was developed based on a corpus of authentic interlanguage produced by a specific learner group. The course is thus tailored to the attested usage problems in this group. In addition, it can be freely used by any teacher with only minimal training. This course could therefore compensate for a potential reluctance of some teachers to teach explicit grammar topics. It can also be used to adjust the level of instruction to different competence levels within the same classroom. In the future, the course can be tailored to different learner groups or expanded to include also other information-heavy topics in the second language classroom.

Conclusion

This thesis combined a corpus-based interlanguage analysis with a didactic tool development. Part I focused on the analysis of subject-verb agreement errors in a corpus of learner English produced by young Norwegian learners. The Norwegian data were then compared to learner data produced by speakers of other L1s. Part II detailed how the data from the corpus were used to develop a didactic tool which targeted the areas of English grammar and discourse structures identified as problematic in the learners' texts. The resulting Inverted Classroom course in English grammar was tested on a group of young Norwegian learners to evaluate whether the increased explicit grammar instruction would increase the learners' accuracy in agreement marking.

The agreement errors analysis revealed that the Norwegian students produce higher number of overgeneralization errors than has previously been reported for learners with other L1 backgrounds. It was suggested that the Norwegian learners use the marked verb form as a default instead of the unmarked form which is identical with the non-finite form in English. This pattern was especially prominent in complex contexts such as after post-modified NP subjects. The Norwegian learners also tend to overuse the plural forms of the verb *BE* when they make errors in suppletive agreement. It was suggested that this may be due to the phonological similarity of the Norwegian forms of *BE* and the English plural forms *are* and *were*. The Norwegian error patterns were compared to data from Swedish and German speaking learners. The Swedish students, whose L1 is syntactically similar to Norwegian, produced similar errors as the Norwegian students. The data from the German students were too limited to draw any conclusion, but, unlike in the Norwegian and Swedish data, a tendency to omit the 3rd person marker was noted. It was therefore suggested that the non-standard agreement marking pattern in the Norwegian data may stem from crosslinguistic influence.

The learner corpus was also used as a basis for the development of a 12-module Inverted Classroom English grammar course. This course was tested in Norwegian schools and its impact on agreement marking accuracy in L2 English production was evaluated. There were no significant differences in accuracy detected between the comparison and the intervention group. The didactic intervention was thus not successful in increasing the accuracy of agreement marking in the production of young Norwegian learners. However, a post-hoc analysis of log data of a selected group of students from the intervention group revealed that most of them did not use the material as intended. It was

therefore suggested to adjust the methodology of the intervention and provide a closer follow-up both to the students and the participating teachers in future research.

The combination of an interlanguage analysis with linguistically informed didactic intervention offered in this thesis showed that there is great potential in linking theoretical second language studies and practical classroom applications. The corpus analysis provided both new data on subject-verb agreement development in L2 English and allowed for a L1 specific didactic tool development. Even though the intervention did not yield the expected results, it offers a valuable starting point for further didactic material development.

Bibliography

Abe, M., & Tono, Y. (2005). Variations in L2 spoken and written English: Investigating patterns of grammatical errors across proficiency levels. *Proceedings from the Corpus Linguistics Conference Series, 1*(1).

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development, 34*(1), 1-14.

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Adger, D., & Svenonius, P. (2011). Features in minimalist syntax. In C. Boeckx (Ed.), *The Oxford handbook of linguistic minimalism* (pp. 27-51). Oxford: Oxford University Press.

Åfarli, T. A., & Eide, K. M. (2003). Norsk generativ syntaks [Norwegian generative syntax]. Oslo: Novus.

Anderson, L. W., & Krathwohl, D. R. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Baker, J. W. (2000). The "Classroom Flip": Using Web course management tools to become the guide by the side. In J. A. Chambers (Ed.), *Selected papers from the 11th International Conference on College Teaching and Learning*. Jacksonville, FL: Florida Community College at Jacksonville.

Barber, C. L., Beal, J. C., & Shaw, P. (2009). *The English language: A historical introduction* (2nd ed.). Cambridge: Cambridge University Press.

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Eugene: International Society for Technology in Education.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bishop, J., & Verleger, M. A. (2013). *The flipped classroom: A survey of the research*. Paper presented at the 2013 ASEE Annual Conference & Exposition, Atlanta, Georgia. https://peer.asee.org/22585

Bley-Vroman, R. (1983). The comparative fallacy in interlangauge studies: The case of systematicity. *Language Learning, 33*(1), 1-17.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*(1-2), 3-45.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393.

Breiteneder, A. (2005). The naturalness of English as a European lingua franca: The case of the 'third person -s'. *Vienna English Working Papers, 14*(2), 3-26.

Chalmers, A. F. (1999). *What is this thing called science?* (3rd ed.). Buckingham: Open University Press.

Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.

Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. New York: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass: MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalistic syntax in honor of Howard Lasnik* (pp. 89-156). Cambridge, MA: MIT Press.

Clahsen, H., & Hong, U. (1995). Agreement and null subjects in German L2 development: New evidence from reaction-time experiments. *Second Language Research, 11*(1), 57-87.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. San Francisco: Pfeiffer.

Crouch, C. H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics, 69*(9), 970-977.

Curzan, A. (2014). *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press.

Davies, M. (2004-). BYU-BNC. (Based on the British National Corpus from Oxford University Press). https://corpus.byu.edu/bnc/

Davies, M. (2008-). The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. https://corpus.byu.edu/coca/

De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol: Multilingual Matters.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94-112). New York: Routledge.

Della Porta, D., & Keating, M. (2008). How many approaches in the social sciences? An epistemological introduction. In D. Della Porta & M. Keating (Eds.), *Approaches and methodologies in the social sciences: A pluralist perspective* (pp. 19-39). Cambridge: Cambridge University Press.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Drew, I., & Sørheim, B. (2009). English teaching strategies: Methods for English teachers of 10 to 16-year-olds (2nd ed.). Oslo: Samlaget.

Dröschel, Y. (2011). Lingua franca English: The role of simplification and transfer. Bern: Peter Lang.

Egbert, J., Herman, D., & Chang, A. (2014). To flip or not to flip? That's not the question: Exploring flipped instruction in technology supported language learning environments. *International Journal of Computer-Assisted Language Learning and Teaching, 4*(2), 1-10.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit langauge learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35-47). Washington: Georgetown University Press.

Faarlund, J. T. (2004). The syntax of Old Norse: With a survey of the inflectional morphology and a complete bibliography. Oxford: Oxford University Press.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles: SAGE.

Forsøksrådet for skoleverket. (1960). Læreplan for forsøk med 9-årig skole [Curriculum for the 9-year compulsory school trial]. Oslo: Aschehoug.

Franck, J., Frauenfelder, U. H., & Rizzi, L. (2007). A syntactic analysis of interference in subject–verb agreement. *MIT Working Papers in Linguistics*(53), 173-190.

Franklin, M. (2008). Quantitative analysis. In D. Della Porta & M. Keating (Eds.), *Approaches and methodologies in the social sciences: A pluralist perspective* (pp. 240-262). Cambridge: Cambridge University Press.

Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition, 19*(2), 101-131.

Fromkin, V. A. (1973). The non-anomalous nature of anomalous utterances. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 215-269). The Hague: Mouton.

Fylkenes informasjonstjeneste for søkere til videregående opplæring. (2014). Gratis læremidler og utstyr [Free education materials and equipment]. Retrieved from https://www.vilbli.no/nb/nb/no/gratis-laremidler-og-utstyr/a/025439

Garshol, L. (2014). Agreement errors in Norwegian high school students' English: Pilot project.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Mahwah, N.J: LEA.

Gaustad, M., & Horn, G. (Unpublished). *Kvalitet i videoforelesning: Sluttrapport*. Norgesuniversitet.

Genesee, F. (2005). The capacity of the language faculty: Contributions from studies of simultaneous bilingual acquisition. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 890-901). Somerville, MA: Cascadilla Press.

Gilboy, M. B., Heinerichs, S., & Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *Journal of Nutrition Education and Behavior, 47*(1), 109-114.

Graedler, A.-L. (2002). Norwegian. In M. Grölach (Ed.), *English in Europe* (pp. 57-81). New York: Oxford University Press.

Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 7-29). Malden: Blackwell.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International corpus of learner English* (2nd ed.). Louvain: Presses universitaires de Louvain.

Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on computer* (pp. 199-209). London and New York: Longman.

Handke, J. (2013a). Beyond a simple ICM. In J. Handke, N. Kiesler, & L. Wiemeyer (Eds.), *The Inverted Classroom Model: The 2nd German ICM-Conference - Proceedings* (pp. 15-21). München: Oldenbourg Wissenschaftsverlag.

Handke, J. (2013b). The VLC video strategy. In J. Handke, N. Kiesler, & L. Wiemeyer (Eds.), *The Inverted Classroom Model: The 2nd German ICM-Conference - Proceedings* (pp. 59-76). München: Oldenbourg Wissenschaftsverlag.

Handke, J., Kiesler, N., & Wiemeyer, L. (2013). *The Inverted Classroom Model, The 2nd German ICM-Conference – Proceedings*. München: Oldenbourg Wissenschaftsverlag.

Harley, B. (1993). Instructional strategies and SLA in early French immersion. *Studies in Second Language Acquisition, 15*(2), 245-259.

Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Harlow: Longman.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics, 42*(01), 25-70.

Hasselgård, H., & Johansson, S. (2011). Learner corpora and contrastive interlanguage analysis. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 33-61). Amsterdam/Philadelphia: John Benjamins Pub.

Hasselgreen, A., & Sundet, K. T. (2017). Introducing the CORYL corpus: What it is and how we can use it to shed light on learner language. *Bergen Language and Linguistics Studies, 7*, 197-215.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London: Routledge.

Haznedar, B. (1997). L2 acquisition by a Turkish-speaking child: Evidence for L1 influence. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston*

University Conference on Language Development (pp. 245-256). Sommerville, MA: Cascadilla Press.

Horverak, M. O. (2016a). English writing instruction in Norwegian upper secondary schools – a linguistic and genre-pedagogical perspective. (PhD thesis), University of Agder,

Horverak, M. O. (2016b). An experimental study on the effect of systemic functional linguistics applied through a genre-pedagogy approach to teaching writing. *Yearbook of the Poznan Linguistic Meeting, 2*(1), 67-89.

Housen, A. (2002). A corpus-based study of the L2 acquisition of the English verb system. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 77-116). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Information Resources Management Association (Ed.) (2016). *Blended learning: Concepts, methodologies, tools, and applications*: Information Science Publishing.

Ionin, T., & Wexler, K. (2002). Why is 'is' easier than '-s'?: Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research, 18*(2), 95-136.

Izumi, E., & Isahara, H. (2004). Investigation into language learners' acquisition order based on an error analysis of a learner corpus. In L. Anthony, S. Fujita, & Y. Harada (Eds.), *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning* (pp. 63-71). Tokyo: Waseda University.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Jensen, I. N. (2016). *The Bottleneck Hypothesis in L2 acquisition: Norwegian L1 speakers' knowledge of syntax and morphology in English L2.* (Master's thesis), UiT The Arctic University of Norway, Tromsø, Norway. Retrieved from https://munin.uit.no/handle/10037/9727

Jensen, I. N., Slabakova, R., & Westergård, M. (2017). The Bottleneck Hypothesis in L2 acquisition: Norwegian L1 speakers' knowledge of syntax and morphology in English L2. In M. LaMendola & J. Scott (Eds.), *BUCLD 41: Proceedings of the 41st annual Boston University Conference on Language Development* (pp. 333-346). Somerville, MA: Cascadilla Press.

Johansson, S. (2008). *Contrastive analysis and learner language: A corpus-based approach*. University of Oslo. Retrieved from http://www.hf.uio.no/ilos/forskning/grupper/Corpus_Linguistics_Group/papers/contrastive-analysis-and-learner-language_learner-language-part.pdf

Julien, M. (2002). Syntactic heads and word formation: A study of verbal inflection. Oxford: Oxford University Press.

Kachru, B. B. (1992). *The other tongue: English across cultures* (2nd ed.). Urbana, Ill: University of Illinois Press.

Kiesler, N. (2013). Using videos in the linguistics classroom. In J. Handke, N. Kiesler, & L. Wiemeyer (Eds.), *The Inverted Classroom Model: The 2nd German ICM-Conference - Proceedings* (pp. 77-92). München: Oldenbourg Wissenschaftsverlag.

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education, 22*, 37-50.

King, A. (1993). From sage on the stage to guide on the side. *College Teaching, 41*(1), 30-35.

Kirke- og undervisningsdepartementet. (1974). Mønsterplan for grunnskolen: M 74 [Model plan for elementary school: M74]. Oslo: Aschehoug.

Kirke- og undervisningsdepartementet. (1987). Mønsterplan for grunnskolen: M 87 [Model plan for elementary school: M87]. Oslo: Aschehoug.

Kirke- undervisnings- og forskningsdepartementet. (1996). Læreplanverket for den 10-årige grunnskolen [The curriculum for the 10-year elementary school]. Oslo: Nasjonalt læremiddelsenter.

Krashen, S. (1985). The input hypothesis: Issues and implications. London: Longman.

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education, 31*(1), 30-43.

Lai, C.-L., & Hwang, G.-J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education, 100*, 126-140.

Lardiere, D. (1998a). Case and tense in the 'fossilized' steady state. *Second Language Research, 14*(1), 1-26.

Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14*(4), 359-375.

Lardiere, D. (2008). Feature assembly in second language acquisition. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106-140). New York: Lawrence Erlbaum Associates.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research, 25*(2), 173-227.

Lardiere, D. (2012). Linguistic approaches to second language morphosyntax. In A. Mackey & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (pp. 106-126). London: Routledge.

Leclercq, P., & Edmonds, A. (2014). How to assess L2 proficiency? An overview of proficiency assessment research. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perscpectives from SLA* (pp. 3-23). Bristol: Multilingual Matters.

Lee, G., & Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: Outcomes and perceptions. *TESOL Quarterly, 52*(1), 62-84.

Leis, A., Tohei, A. A., & Cooke, S. (2015). The effects of flipped classrooms on English composition writing in an EFL environment. *International Journal of Computer-Assisted Language Learning and Teaching, 5*(4), 37-51.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365.

Loewen, S. (2015). Introduction to instructed second language acquisition. New York: Routledge.

Long, M. H. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives*. Cambridge, MA: Newbury House Publishers.

Loviscach, J. (2013). The Inverted Classroom: Where to go from here. In J. Handke, N. Kiesler, & L. Wiemeyer (Eds.), *The Inverted Classroom Model: The 2nd German ICM-Confrence - Proceedings* (pp. 3-14). München: Oldenbourg Wissenschaftsverlag.

Mazur, E. (1997). Peer Instruction: Getting students to think in class. *AIP Conference Proceedings, 399*(1), 981-988.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology, 7*.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Nassaji, H., & Fotos, S. (2011). Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context. New York: Routledge.

Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., Rica, J.-P., Díez, M., & Prieto, R. (2007). A contrastive functional analysis of errors in spanish EFL university writers' argumentative texts: Corpus-based study. In E. Fitzpatrick (Ed.), *Corpus linguistics beyond the word: Corpus research from phrase to discourse* (pp. 203-225). Amsterdam: Rodopi.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. M. Sinclair (Ed.), How to Use Corpora in Language Teaching (pp. 125-152). Philadelphia, Netherlands: John Benjamins Publishing Company. Retrieved from http://ebookcentral.proquest.com/lib/agder/detail.action?docID=622666.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528.

Norwegian Department of Education. (2006). *English subject curriculum*. Oslo: Norwegian Department of Education. Retrieved from https://www.udir.no/kl06/ENG1-01.

Norwegian Department of Education. (2012). *Sensorveiledning ENG1002/ENG1003 og ENG1005/1006 [Sensor instructions ENG1002/ENG1003 og ENG1005/1006]*. Oslo: Norwegian Department of Education. Retrieved from https://docplayer.me/1204239-Eksamensveiledning-om-vurdering-av-eksamensbesvarelser.html.

Norwegian Department of Education. (2013). *English subject curriculum*. Oslo: Norwegian Department of Education. Retrieved from https://www.udir.no/kl06/ENG1-03?lplang=http://data.udir.no/kl06/eng.

Ocampo, S. (2013). *Morphological variability in long-distance subject-verb agreement: A study of native and nonnative processing.* (Master's thesis), University of Kansas, Retrieved from https://kuscholarworks.ku.edu/handle/1808/11712?show=full

Opplæringslova. (1998). *Lov om grunnskolen og den vidaregåande opplæringa (LOV-1998-07-17-61) [Elementary and comprehensive education act]*. Retrieved from https://lovdata.no/lov/1998-07-17-61.

Østerlie, O. (2018). Can flipped learning enhance adolescents' motivation in physical education? An intervention study. *Journal for Research in Arts and Sports Education, 2*(1), 1-15.

Owston, R., Lupshenyuk, D., & Wideman, H. (2011). Lecture capture in large undergraduate classes: Student perceptions and academic performance. *The Internet and Higher Education, 14*(4), 262-268.

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition, 18*(1), 1-25.

Piaget, J. (1968). *Six psychological studies* (A. Tenzer, Trans. D. Elkind Ed.). New York: Vintage Books.

Pollock, J.-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry, 20*(3), 365-424.

Prodoehl, D. E. (2016). Flipping first-year English: Strengthening teacher-student conferencing through online modules. In Information Resources Management Association (Ed.), *Blended learning: Concepts, methodologies, tools, and applications* (pp. 74-96): Information Science Publishing.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Radford, A. (2004). *Minimalist syntax: Exploring the structure of English*. Cambridge: Cambridge University Press.

Roach, T. (2014). Student perceptions toward flipped learning: New methods to increase interaction and active learning in economics. *International Review of Economics Education, 17*, 74-84.

Roberts, I. (1993). Verbs and diachronic syntax: A comparative history of English and French. Dordrecht: Kluwer.

Roberts, I. (1998). *Have/Be* Raising, Move F, and Procrastinate. *Linguistic Inquiry, 29*(1), 113-125.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*(2), 147-163.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*(1), 40-72.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*(2), 165-179.

Sheen, R. (2002). 'Focus on form' and 'focus on forms'. *ELT Journal, 56*(3), 303-305.

Sheen, R. (2005). Focus on formS as a means of improving accurate oral production. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition:* (pp. 271-310). Berlin/New York: Mouton de Gruyter.

Simo Bobda, A. (1998). *On the indigenization of English in Cameroon and New Englishisms*. Paper presented at the LAUD, Essen. http://www.linse.uni-due.de/laud-downloadliste.html?articles=on-the-indigenization-of-english-in-cameroon-and-new-englishisms

Slabakova, R. (2000). L1 transfer revisited: the L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics, 38*(4), 739–770.

Sobin, N. (1994). An acceptable ungrammatical construction. In S. D. Lima, R. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 51-66). Amsterdam: J. Benjamins.

Sobin, N. (1997). Agreement, default rules, and grammatical viruses. *Linguistic Inquiry, 28*(2), 318-343.

Sørensen, C.-F. (2016). *Omvendt klasserom som strategi for et inkluderende læringsmiljø - En rapport om erfaringer fra et stort emne ved NTNU*. Retrieved from http://www.universell.no/fileshare/filarkivroot/Universell/Inkluderende%20l%C3%A6ringsmilj%C3%B8/Nyheter/Universellrapport%203_2016%20Omvendt%20klasserom%20og%20UUL.pdf

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*(2), 181-207.

Spannagel, C., & Spannagel, J. (2013). Designing in-class activities in the *Inverted Classroom Model*. In J. Handke, N. Kiesler, & L. Wiemeyer (Eds.), *The Inverted Classroom Model: The 2nd German ICM-Conference - Proceedings* (pp. 113-120). München: Oldenbourg Wissenschaftsverlag.

Statistics Norway. (2017a). *ICT usage in households*. Retrieved from https://www.ssb.no/en/teknologi-og-innovasjon/statistikker/ikthus.

Statistics Norway. (2017b). Upper secondary education. Retrieved 09.06. 2018, from Statistics Norway https://www.ssb.no/en/statbank/list/vgu?rxid=c982f187-74fb-4ec4-9ef4-d0f05858dd2a

Steen, C. (2013). Omvendt undervisning i matematikk: En studie av elevers oppfatning av undervisningsmetoden [Inverted instruction in mathematics: A study of students' experience with the teaching method]. (Master thesis), University of Agder, Kristiansand. Retrieved from http://hdl.handle.net/11250/193726

Strayer, J. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research, 15*(2), 171-193.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110*(16), 6313-6317.

Thagg Fisher, U. (1985). The sweet sound of concord: A study of Swedish learners' concord problems in English. CWK Gleerup, Lund.

Thewissen, J. (2015). *Accuracy across proficiency levels: A learner corpus approach*. Louvain-la-Neuve: Presses universitaires de Louvain.

Thue, B. O. (2016). Campus Inkrement. Retrieved from http://campus.inkrement.no/

Tono, Y., & Díez-Bedmar, M. B. (2014). Focus on learner writing at the beginning and intermediate stages: The ICCI corpus. *International Journal of Corpus Linguistics, 19*(2), 163-177.

Trudgill, P. (2002). *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.

Vanderbilt University Center for Teaching. (2016). Bloom's Taxonomy. Retrieved from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

VanPatten, B. (2011). Stubborn syntax: How it resists explicit teaching and learning. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 9-21). Washington: Georgetown University Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Westergård, M. R. (2009). The acquisition of word order: Micro-cues, information structure, and economy. Amsterdam: John Benjamins.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Interlanguage studies bulletin, 7*(2), 133-161.

White, L. (1992). Long and short verb movement in second language acquisition. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 37*(2), 273-286.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly, 28*(2), 377-389. doi:10.2307/3587438

Williams, J. (1995). Focus on form in communicative language teaching: Research findings and the classroom teacher. *TESOL journal, 4*(4), 12-16.

Wong, S.-M. J., & Dras, M. (2009). Contrastive analysis and native language identification. *Australasian Language Technology Association Workshop: Proceedings of the Workshop*, 53-61.

Appendices

Appendix 1

Background questionnaire

Spørreskjema – bakgrunnsopplysninger

Kode   **Kjønn**   **Fødselsår**

Mann ☐   Kvinne ☐   _____

Hvilke språk bruker dere hjemme?

_____

Har du gått på skole i Norge fra første trinn?   Ja   Nei

Hvis nei, i hvilket trinn begynte du i Norge?   _____

Hvilke andre språk har du hatt på skolen eller kurs?

_____

Bruker du et annet språk enn norsk i ditt daglige liv (utenfor skolen)?   Ja   Nei

Hvis ja, hvilke språk?   _____

Hva bruker du fremmedspråket til?   _____

_____

I hvilken grad tror du du kommer til å ha bruk for fremmedspråk i ditt fremtidige arbeid?
veldig liten   liten   usikker   stor   veldig stor
☐   ☐   ☐   ☐   ☐

Tror du det er viktig å være en selvstendig bruker av fremmedspråket?
ikke viktig   litt viktig   usikker   ganske viktig   veldig viktig
☐   ☐   ☐   ☐   ☐

Hvor viktig er det for deg å bruke språket (utenfor skolen) på en grammatisk korrekt måte MUNTLIG?
ikke viktig   litt viktig   usikker   ganske viktig   veldig viktig
☐   ☐   ☐   ☐   ☐

Hvor viktig er det for deg å bruke språket (utenfor skolen) på en grammatisk korrekt måte SKRIFTLIG?
ikke viktig   litt viktig   usikker   ganske viktig   veldig viktig
☐   ☐   ☐   ☐   ☐

Har du noen gang fått en diagnose knyttet til læring (f.eks. dysleksi, ADHD)?   Ja   Nei

190

Appendix 2

Screenshots from the course platform

Appendix 3

Overview of the progress through the course – whole class (screenshot)

Elev	Progresjon	Resultat		Tid
Student names removed	67%		64% rett	38:24
	53%		45% rett	01:28
	72%		75% rett	26:46
	30%		32% rett	12:28
	80%		84% rett	20:36
	64%		53% rett	33:54
	62%		64% rett	1:13:31
	80%		57% rett	24:51
	27%		29% rett	03:26
	100%		98% rett	05:19
	66%		55% rett	16:21
	0%		0% rett	NA
	71%		66% rett	10:24
	56%		40% rett	31:41
	80%		84% rett	42:37
	62%		52% rett	05:16
	34%		29% rett	10:02
	64%		66% rett	22:33
	5%		1% rett	NA
	67%		58% rett	51:14
	60%		51% rett	59:14

192

Overview of the students' work with one module (screenshot)

Emne:	8.1 Simple or Continuous? ▼	Dato:	fra	til

Kompaktvisning

Elev	Oppgaver					Evaluering			Video		Tid
Student names removed	1	2	3	4	5	1	2	If you have any questions or comments, you can write them here.	1	2	
	✓	✓	✓	✓	✗	🙂	☹		📹	📹	10:52
	✓	✓	✗✓	✓	✓	🙂	😐		📹	📹	09:59
	✓	✓	✓	✓	✗	🙂	☹		📹	📹	04:50
	✓	✓	✗	✓					📹		01:24
	✓	✗✓	✗✓	✓	✓	🙂	🙂		📹	📹	13:39
	✓	✓	✓	✓	✓	🙂	☹		📹	📹	02:17
	✗✓	✗✓	✓	✓	✗✓	🙂	🙂	På oppgave 5 glemte dere "N" i "kitchen".	📹	📹	02:17
	✓	✓	✓	✗	✗	🙂	😐		📹	📹	16:19
	✓	✓	✗✓	✓	✗✓	🙂	🙂		📹	📹	12:28
	✗..✓	✗✓	✗	✓	✓	😐	😐		📹	📹	1:19:44
	✗✓	✓	✓	✓	✗✓	😐	😐		📹	📹	12:33
	✗✓	✓	✗✓	✓	✗✓	🙂	🙂		📹	📹	18:52
	✓	✓	✗	✓	✗	🙂	😐		📹	📹	08:51
	✓	✗✓	✗✓	✓	✗✓	😐	😐		📹	📹	02:14
	✓	✗✓	✓	✓	✗	🙂	😐		📹	📹	1:11:55
	✗✓	✓	✗✓	✓	✗✓	🙂	🙂		📹	📹	01:28
	✓	✓	✗	✓	✓	🙂	🙂		📹	📹	03:07
	✓	✗	✓	✓	✗	😐	🙂		📹	📹	18:58
	✓	✓	✗✓	✓	✗✓	🙂	😐		📹	📹	13:11

Appendix 4

Questionnaire about the use of the Inverted Classroom course

Spørreskjema – bruk av omvendt undervisning

Kode

Hvor ofte har du brukt omvendt undervisningskurset i engelsk i løpet av det siste året?

ofte	noen ganger	sjeldent	ikke i det hele tatt
☐	☐	☐	☐

Har du kommet tilbake til/spilt på nytt noen moduler?

ja, mange ganger	ja, noen ganger	sjeldent	ikke i det hele tatt
☐	☐	☐	☐

Syns du at du lærte/husket bedre på denne måten enn ved vanlig tavleundervisning?

ja, mye bedre	ja, litt bedre	usikker	nei
☐	☐	☐	☐

Likte du denne formen av teoriundervisning?

ja, veldig	det var ok	nei, ikke så godt	ikke i det hele tatt
☐	☐	☐	☐

Har du erfaring med omvendt undervisning fra andre fag?

ja	nei
☐	☐

Hvis ja, hvilke?

_____

Hvis nei, har du lyst å prøve det ut (i f.eks. matte, historie, osv.)?

ja	nei
☐	☐

194

Appendix 5

Usage data

Student	Frequency of use	Frequency of reuse	Nr. of modules used	Time spent on module 3 (in minutes)	Total time spent (in minutes)	Last log-in
16HOV07	sometimes	sometimes	5	6	15.5	11.1.
16HOV08	sometimes	never	5	7.5	21	9.1.
16HOV09	rarely	rarely	6	8.5	35	23.1.
16HOV19	sometimes	sometimes	6	3.5	18.5	23.1.
16STV10	sometimes	rarely	8	3.5	160.5	13.3.
16STV24	sometimes	rarely	9	30	187	25.4.
16STV28	sometimes	rarely	9	34	141	17.3.
16STV29	often	never	6	0	119	15.3.
16STV30	rarely	never	1	0	22	11.11.
16STV31	rarely	never	8	14	68.5	17.3.
16STV32	sometimes	never	9	23	120	16.3.
16STV33	sometimes	rarely	8	4.5	58.5	16.3.
16STV34	sometimes	rarely	9	4	51.5	16.3.
16STV35	often	never	10	39	264	13.3.
16STV37	sometimes	rarely	7	0	89.5	30.3.
16STV39	sometimes	never	6	0	85	16.3.
16STV40	sometimes	sometimes	8	51	236	19.12.
16STV41	sometimes	sometimes	8	52.5	107.5	16.3.
16STV42	often	never	9	23	111	16.3.
16STV43	sometimes	never	8	3	58	17.3.
16STV44	rarely	never	10	51	255	16.3.
16STV45	often	rarely	10	23	231	13.3.
16STV46	sometimes	never	8	12	133.5	17.3.
16STV47	sometimes	sometimes	4	0	120	25.4.
16STV48	sometimes	never	10	22	179	13.3.
16STV49	never	never	4	4	18.5	16.3.
16STV50	rarely	rarely	0	0	0	26.9.

Appendix 6

Background data – matched groups only

	Comparison group	Intervention group
Total students	69	64
Females	45	39
Males	22	25
Multilinguals	8	13
Education partially abroad	3	8
Learning-related diagnosis	1	1
Use of English outside of school	22	25
English for communication	10	14
English online	10	11
Travel, culture, hobby	7	6
Positive answer for:		
Usefulness of additional languages for work	32	47
Importance of being independent user	41	58
Importance of correctness orally	33	42
Importance of correctness in writing	35	49