UiT

THE ARCTIC
UNIVERSITY
OF NORWAY

Department of Language and Linguistics

The Bottleneck Hypothesis in L2 acquisition

Norwegian L1 speakers' knowledge of syntax and morphology in English L2

—

Isabel Nadine Jensen
Master thesis in English Linguistics May 2016

Acknowledgements

First of all, I would like to thank my supervisor, Marit Westergaard, for support and guidance throughout this process. I would also like to thank Roumyana Slabakova for valuable advice and feedback.

Thanks to the LAVA group at the Faculty of Humanities, Social Science and Education, Department of Language and Linguistics for receiving the master's grant, and for letting me take part in interesting and fruitful seminars and discussions.

Thanks also to Björn Lundquist for helping me with R and the analysis of the data, to all of the participants in the pilot study and the main experiment, and to the schools and teachers who let me test their students.

I would further like to thank Tone, Sten Ove, Marte, and Christian for proof reading and moral support, and finally, thank you Nikolai, for your technical and moral support, motivation and inspiration during this last year.

1. Introduction

In this thesis, Roumyana Slabakova's (2006; 2008; 2013) Bottleneck Hypothesis is tested in second language (L2) acquisition of English by Norwegian speakers. The issue of why something is difficult or easy to acquire in an L2 is a widely discussed topic (see e.g. Sorace and Filiaci 2006; Sorace 2011). The reason why I find it interesting to address this is that it provides insight about the cognitive process involved in language acquisition, and it may point towards implications for language teaching. The former notion is the main purpose of this thesis: to contribute to current knowledge of the cognitive process of L2 acquisition.

In short, the Bottleneck Hypothesis argues that functional morphology is the bottleneck in L2 acquisition, and thus more difficult to acquire than linguistic domains such as syntax, semantics, and internal and external interfaces. In this thesis, I only focus on the comparison of syntax and functional morphology. Based on this, the hypothesis in the current thesis is that English L2 learners' performance on functional morphology is weaker than their performance on syntax. In order to investigate the Bottleneck Hypothesis, I pose the following research questions:

RQ1: Is morphology more difficult than narrow syntax in L2 acquisition?
RQ2: Is morphology a more persistent problem than narrow syntax?

In order to test research questions 1 and 2, I compare acquisition of subject-verb agreement, which represents functional morphology, and verb movement, which represents narrow syntax. The reason for this is that there is a mismatch between Norwegian and English with respect to these two constructions. That is, there is no overt agreement morphology in Norwegian, whereas English marks present tense verbs when the subjects are 3rd person singular. In addition, Norwegian is a V2 (verb second) language, which means that the verb always occurs in the second position of a declarative main clause. English, on the other hand, is an SVO language (subject-verb-object), which means that the verb stays in VP in these types of sentences. Example of the mismatches are provided in (1) and (2):

(1) *Subject-verb agreement*

     a.   Mary drinks wine   [English]

     b.   Mary and John drink wine   [English]

     c.   Mari   drikker   vin   [Norwegian]
         Mari   drinks   wine
         'Mary drinks wine'

     d.   Mari og   Jon   drikker vin   [Norwegian]
         Mari and   Jon   drink   wine
         'Mari and Jon drink wine'

(2) *Verb movement*

     a.   Yesterday Mary drank wine   [English]

     b.   I går   drakk Mari vin   [Norwegian]
         Yesterday   drank Mari wine
         'Yesterday Mari drank wine'

It is expected that accurate functional morphology is not only more challenging, but also that it is a more persistent problem as the learners become more advanced speakers of English. This is based on previous research on acquisition of syntax and functional morphology which has laid the foundation of the Bottleneck Hypothesis (see e.g. Ionin and Wexler 2002; Haznedar 2001; Lardiere 1998a,b). These studies have found that syntax comes before functional morphology in L2 acquisition.

Furthermore, previous research has shown that the difficulty of a construction depends on the sentence structure. To exemplify, Håkansson and Collberg (1994) found that acquisition of the target word order *negation<modal* in more difficult than *negation<lexical verb* in Swedish embedded clauses. In addition, Ocampo (2013) found that long-distance subject-verb agreement is more challenging when the structural distance in a sentence is increased, as well as when the number feature on the noun that is immediately preceding the verb is [plural] (see section 2.4.2 and 2.4.3). For that reason, it is necessary to test and discuss different structures of subject-verb agreement and verb movement. Morphology is tested by means of six different experimental conditions: long-distance agreement and local agreement, and singular and plural subjects. In the sentences with long-distance agreement, prepositional phrases are used, and

there is an asymmetry between the number of the subject noun and the noun that is closest to the verb. Narrow syntax is tested by two other conditions: Non-subject-initial declarative clauses with lexical verbs and auxiliary verbs. The different types of sentences are exemplified in (3) and (4).

(3) *Subject-verb agreement*

    a.   The girl drinks wine   [3rd person sg, local agreement]

    b.   The girls drink wine   [3rd person pl, local agreement]

    c.   The girl with the heavy books drinks coffee   [3rd person sg, long-distance]

    d.   The girls in the red car drink coffee   [3rd person pl, long-distance]

(4) *Verb movement*

    a.   Yesterday the students drank wine   [Lexical verbs]

    b.   Tomorrow the students will drink wine   [Auxiliary verbs]

This gives rise to the third research question, which is thus not directly related to the comparison of performance in narrow syntax and functional morphology:

RQ3: Which of the syntactic and morphological conditions are more difficult?

For research question 3, there are three expectations regarding functional morphology: First, it is expected that long-distance agreement is more difficult than local agreement. The reason for this is the study by Ocampo (2013) mentioned above, which found that learners are less sensitive to agreement errors in more structurally complex sentences, in comparison to less complex sentences. This has also been found in several other studies (see section 2.4.3). It is thus expected that adding an element between the subject and the verb will cause more problems for the establishment of agreement than when the subject immediately precedes the verb. Secondly, it is expected that long-distance agreement is more difficult when the subject is singular than when it is plural. This prediction is also based on findings in previous studies, such as Ocampo (2013), who found that learners are significantly less sensitive to long-distance agreement errors when the subject is singular and the intervening noun is plural. Similarly, it is predicted that local agreement is more challenging when the subject is plural, as previous studies have found that the most common agreement error is to drop the 3rd person singular suffix *–s* in environments in which there should be an *–s,* i.e. with plural subjects, rather than inserting a superfluous *–s.*

With respect to the syntactic structures, the prediction is that verb movement is more challenging when sentences contain an auxiliary verb. The reason for this is Pollard's (1996) argument that auxiliary verbs are unspecified for the feature [inverted] in English, whereas lexical verbs have the feature [-inverted]. Consequently, learners of English receive input which tells them that the auxiliary may move in some cases in English and that lexical verbs never move out of the VP. This suggests that there might be more variability with word order in sentences which contains auxiliary verbs, as the learners are more inclined to move these verbs in English. Nevertheless, both syntactic constructions are expected to cause few problems for the learners.

In order to investigate the research questions, an acceptability judgement test is used to test the participants' intuitions about syntax and functional morphology. This is a widely used quantitative method in generative linguistics, which makes it possible to gather data from a large number of speakers. From this, it is possible to look for descriptive and causal patterns in the speakers' L2 grammar. As mentioned, Norwegian learners of English L2 are used in this experiment. There are 60 participants in total, and their age ranges from 11-12 and 15-18. In addition, they are split up into four proficiency groups: low intermediate, intermediate, high intermediate and advanced speakers. These groups are determined on the basis of a subset of a Standardized Oxford Proficiency test with 40 multiple-choice questions. The participants' performance on morphology and syntax is considered both across proficiency levels, and within each level.

Furthermore, following the theoretical framework of the Bottleneck Hypothesis, the current study is set within generative linguistic theory (see e.g. Chomsky 1957; 1965). This view argues that grammar consists of mental representations which are limited by linguistic universals. These are referred to as Universal Grammar (UG) and are part of the innate language faculty, with which we are all born. A salient issue in generative second language research is whether or not L2 learners can still access UG. The Bottleneck Hypothesis assumes the Full Transfer/Full Access Hypothesis (Schwartz and Sprouse 1994; 1996), which argues that L2 learners have direct access to UG, i.e. that it is accessible to the same extent as in L1 acquisition. Furthermore, as the name suggests, full transfer of the L1 grammar is also argued for. That is, learners initially assume that the L2 linguistic system is the same as the L1 system. They then restructure their L2 grammar by accessing UG when they are exposed to L2 input which illustrates that the target linguistic system differs from their native system. In other words,

acquisition can happen by means of positive transfer or access to UG, whereas those representations that are not transferable and thus not processed by the means of universal operations must be lexically learned.

The main finding in the current thesis is that subject-verb agreement is significantly more difficult than verb movement in acquisition of English L2 by Norwegian speakers. This is not only seen at the lower proficiency levels, but also among the more proficient speakers, who have been exposed to English L2 for several years. Moreover, whereas performance in verb movement makes a developmental jump from the intermediate stage to the high intermediate stage, acquisition of subject-verb agreement seems to stagnate. These findings lend support the Bottleneck Hypothesis and may contribute to the discussion of why something is difficult or easy to acquire in a second language.

This thesis is divided into the following sections: Section 2 describes the theoretical background, section 3 discusses the methodology, research questions and the predictions, section 4 illustrates the results from the current experiment, section 5 discusses the research questions and predictions, and finally, a conclusion is provided in section 6.

2   Theoretical background

In this chapter, I first discuss the field of second language acquisition (also referred to as SLA or L2 acquisition), with focus on the generative linguistic tradition. Secondly, in section 2.2, I discuss the differences between Norwegian and English with respect to the two constructions that are tested in this experiment, namely, subject-verb agreement and verb movement. In section 2.3, I discuss the Bottleneck Hypothesis (Slabakova 2006; 2008; 2011), as well as the contrastive Interface Hypothesis (Sorace and Filiaci 2006; Sorace 2011). Finally, in section 2.4, I address previous research on L2 acquisition of subject-verb agreement and verb movement.

2.1 Second language acquisition

The field of second language acquisition aims to investigate the cognitive process involved in creating a new language system (Gass 2013: 1). Several authors use the term *second language* to refer to all languages that are acquired after the native language (e.g. Berggreen and Tenfjord 1999; Gass 2013). Rothman et al. (2013:372-373), on the other hand, argue that types of adult language learning must be differentiated, as their initial states[1] differ. The definition of the initial state is widely discussed topic. In this thesis, I use the definition put forward by White (2003b) and Hermas (2014), who argue that the initial state refers to those representations that the learners bring to the L2 learning task, i.e. the grammar that is at the outset of the L2 acquisition (White 2003b:19; Hermas 2014:2). This means that monolinguals and multilinguals have different starting points when they acquire a new language. For that reason, the current study uses the term *second language* solely to describe learners who acquire their second language, and not their third, fourth, or fifth, etc.

Furthermore, as mentioned in the introduction, the thesis is set within the generative linguistic framework (see Chomsky 1957; 1965). Generative linguistics is based on the nativist approach, and more specifically, on special-nativism, which holds that we are born with an innate language faculty (Gass 2013: 160; Rothman et. al. 2013: 373). Grammar is viewed as mental representations, constrained by universal principles (White 2003b: 19). These linguistic universals are called Universal Grammar (UG), and are part of the language faculty (White

---

[1] In addition, Hamas (2014:2) distinguishes between the *initial state* and *initial stages.* The latter refers to the stages on which the speakers have received some input from the L2, whereas the former refers to the representations that the learners have before they are exposed to the L2 (Hermas 2014:2).

2003b: 20). In other words, human beings already know something about grammar when they are born. These arguments build on cognitive psychology, which argued against the behaviourist claim that the brain is initially an empty box, and that learning comes from forming habits acquired by mimicking others (Gass 2013: 81).

The reason for the generativist claim about UG comes from the *logical problem of language acquisition,* also referred to as the problem of the *poverty of the stimulus* (see e.g. Crain and Nakayama 1987). This idea is the core of Chomsky's generative linguistics, and is based on the so-called Plato's Problem: "[h]ow do we come to have such rich and specific knowledge, or such intricate systems of belief and understanding, when the evidence available to us is so meagre?" (Cook and Newson 2007: 55). More specifically, this problem concerns the fact that L1 learners illustrate knowledge about linguistic properties that they cannot have received from the linguistics input to which they are exposed. The generativist explanation for why this happens is that these properties are part of UG, and for that reason, they do not need to be learned, as they are already built into the speakers' knowledge (White 2003b: 20 and 22). Instead, the learners need exposure to the language in order to hypothesize and then test their hypotheses about the target grammar.

Crain and Nakayama (187:525) tested the logical problem by investigating whether or not 3-5 year-old speakers of English L1 have a rule governed grammar that is based on internal structures of the language. For this purpose, they tested if the speakers' subject/auxiliary verb inversion was governed by structure-independent hypotheses or structure-dependent hypotheses. An example of the former is to say that *yes/no* interrogatives are formed by verb movement to the front of the sentence in declaratives like (5a)-(5c), resulting in the interrogative counterparts in (6a)-(6c) (Crain and Nakayama 1987:525).

(5)

     a. The man is tall
     b. The book is on the table
     c. I can go

(6)

     a. Is the man tall?
     b. Is the book on the table?
     c. Can I go?

However, this rule cannot account for all instances of *yes/no* interrogatives in English, as exemplified in (7) and (8). Here, movement of the leftmost verb to the front of the sentence results in a non-target like structure, as seen in (8a). In the target structure, on the other hand, the *is* in the relative clause in (8b) is not affected when the interrogative is formed, which argues against the structure-independent hypothesis.

(7) The man who is tall is in the other room

(8)

      a.   *Is the man who __ tall is in the other room?

      b.   Is the man who is tall __ in the other room?

Instead, this indicates that a rule about language structure must be based on the internal structure of a sentence, i.e. it requires a syntactic analysis of the sentence that takes, for instance, NPs, VPs, main clauses and subordinate clauses into consideration. This is called a structure-dependent hypothesis (Crain and Nakayama 1987: 525).

The main finding in this study was, according to Crain and Nakayama (1987: 530) that none of the participants moved the verb that was in the relative clause, i.e. they never produced sentences like (8a). The fact that they did not do this suggests that L1 learners do not make structure-independent hypothesis about their target language, but rather, that they engage structure-dependent grammatical rules (Crain and Nakayama 1987: 533). This is taken as an indication of the existence of a logical problem in English L1 acquisition.

Similarly, Newson and Cook (2007) argue for the logical problem in L1 acquisition by referring to English L1 speakers' knowledge of sentences like (9). In an acceptability judgment test, child L1 speakers of English rejected these types of sentences 99.6 % of the time (Cook and Newson 2007: 56).

(9) *Is Sam is the cat that black?

The question that is posed is thus how these native speakers have managed to learn that sentences like (9) are incorrect in English. It is likely that they have never been exposed these structures by adult English L1 speakers, and it is unlikely that their parents corrected them if they uttered a sentence like this, as parents rarely correct their children's grammar (Cook and Newson 2007: 56; Guasti 2004: 3). In other words, the L1 speakers have no explicit knowledge of the fact that a sentence like (9) is ungrammatical, but they still know that it is an unacceptable

sentence in English. As stated above, the answer to this question is, according to the Chomskyan view, that the speakers have innate knowledge that constrains their grammar and children use this to fill in the gaps left open by insufficient input (Cook and Newson 2007: 57).

Furthermore, what is characteristic about L1 acquisition is that the learners always end up with the same grammar, and they always become native speakers (under normal circumstances). Bley-Vroman (2009:175) refers to this as *reliability* and *convergence* in the L1, i.e. that it is always uniform and successful. In L2 acquisition, on the other hand, the acquisition process is not uniform, the outcome usually differs considerably from learner to learner, and the learners rarely, if ever, end up with native-like competence (see e.g. White and Genesee 1996; Birdsong 1992; Coppieters 1987 for discussions about native-like competence in the L2).

Based on these differences, White (2003b:22) poses two questions with respect to the nature of L2 acquisition. The first one is whether there is a logical problem in L2 acquisition, and the second questions is whether UG is still available. The reason why a logical problem in the L2 does not automatically suggest that UG is involved is the fact that the knowledge that goes beyond the input may come from sources like the speakers' L1 (i.e. transfer), explicit language instruction or input in combination with general learning principles (White 2003b:22). In other words, the questions are if the participants' knowledge of the L2 goes beyond their input, and if so, whether UG provides this knowledge.

Evidence for both of these questions is put forward by Felix (1988), who investigated the intuitions of 48 adult German learners of English on grammatical and ungrammatical English sentences. The sentences tested a number of constructions, and one example is the superiority effect, as exemplified in (10) (Bley-Vroman 1990: 37). In short, the superiority effect means that in these types of sentences, *what* cannot move over *who* (Falk 2012:2). Felix (1988: 290) points out that the constructions that are tested are not taught in classrooms, and are not represented in the German grammar. For instance, with respect to the example in (10), there is no superiority effect in German (Bley-Vroman 1990: 38). In other words, the participants cannot have received knowledge about these structures from positive transfer from their German L1, nor from classroom instruction of English.

(10)

    a. I don't know who did what

    b. *I don't know what who did

Nevertheless, the results from the study show that the majority of the participants judged the sentences correctly (Felix 1988: 285). Felix (1988: 279 and 285) argues that this clearly indicates that the participants involve UG in L2 acquisition, as they illustrate knowledge that must come from the universal operations of the language faculty.

Based on evidence from studies like this (see also e.g. White 2003b: 24), several hypotheses about the initial state and stages argue for access to UG, such as the Minimal Tree Hypothesis (Vainikka and Young-Scholten 1994; 1996) and the Full Transfer/Full Access Hypothesis (Schwartz and Sprouse 1994; 1996). As the Bottleneck Hypothesis assumes the Full Transfer/Full Access Hypothesis, I only focus on this view in the following sections.

In short, the Full Transfer/Full Access Hypothesis holds that L2 learners initially transfer their whole L1 end-state grammar to the L2, i.e. they assume that the target grammar is the same as their native grammar. In other words, the L1 is the initial state in L2 acquisition. In addition, it argues that learners have full, or direct, access to UG, i.e. that they are able to access UG to the same extent as L1 learners. In order to elaborate on this, I first discuss the notion of transfer, and secondly, the notion of access to UG in L2 acquisition.

2.1.1   Transfer

Transfer in the Full Transfer/Full Access Hypothesis is defined as the process in which underlying mental representations, rather than surface structures, are transferred from the native language to the L2 (White 2000 136; Hermas 2014:2). To exemplify, the syntactic consequences of functional categories and feature values are considered universal, and can thus be transferred, whereas lexical words are not transferable (functional categories and formal features are discussed in more detail in section 2.2).

There are two possible outcomes of transfer: positive transfer (facilitation) and negative transfer (interference) (Odlin 2003: 438). Negative transfer leads to non-target like sentences. To exemplify, transfer of the Norwegian word order in non-subject-initial declarative main clauses leads to non-target like English sentences, as the word order differs in Norwegian and English.

This is illustrated in (11).

(11)   Norwegian L1:   I går   gikk Per   til   butikken

            English L2:   *Yesterday went Per   to   shop.DEF

            Intended: 'Yesterday Per went to the shop'

Positive transfer, on the other hand, results in target like utterances in the L2. This is exemplified in (12), in which transfer of the Norwegian word order in declarative main clauses leads to target like English, because the word order in Norwegian and English is the same in these sentences.

(12)   Norwegian L1:   Per   snakker   engelsk

            English L2:   Per   speaks   English

In other words, similarities between transferable representations in the L1 and the L2 lead to positive transfer, whereas differences lead to negative transfer (Odlin 2003: 348).

2.1.2   Access to UG

The Full Transfer/Full Access Hypothesis holds that there are no differences between L1 acquisition and L2 acquisition with respect to how involved UG is (Schwartz and Sprouse 1996: 41; White 2003b:30). To specify, L2 learners access UG in order to restructure their interlanguage[2] system (Schwartz and Sprouse 1996: 40). This happens when the learners are exposed to L2 input, from which they determine if the target language has the same underlying structures as their native language. That is, if a representation differs, they access UG to make their interlanguage more target-like (Schwartz and Sprouse 1996: 41). This may happen immediately after exposure to the L2, or, if differences between the L1 and the L2 grammar are not illustrated in the surface representations, the learners may assume that the two grammars are the same, and for that reason, not change their interlanguage (White 2000: 132). In other words, it is impossible to pinpoint when the restructuring happens. To exemplify, the sentences in (12) above show that English and Norwegian has the same word order in main clauses. However, the underlying word order is not the same, as English is an SVO language, and

---

[2] The term *interlanguage* refers to the mental grammar of a language learner. That is, the learners' L2 grammar is a natural language system described by linguistic rules and principles. In other words, non-target like representations are not viewed as faulty imitations of the target language, but rather, part of the L2 interlanguage system (White 2003b:19). As Odlin (2003:348) points out, it is clear that the concept of interlanguages can be used in the majority of instances of L2 acquisition, although the question of whether or not it can be used all cases has been discussed.

Norwegian is a V2 language. This is visible in the sentences in (11). In other words, there are different requirements with respect to verb movement (this is explained in more detail in section 2.3), but these are not always visible in the surface structure. In other words, the Norwegian learners of English must be exposed to the structures in which the differences in word order is seen, i.e. sentences like (11), in order to change their L2 grammar. If the learners are not exposed to these structures, they continue to assume that the English word order is the same as the Norwegian one. The sooner the learners are exposed to the mismatch, the sooner they are able to acquire the English word order (White 2000: 132).

2.2 The Bottleneck Hypothesis

The Bottleneck Hypothesis is proposed by Roumyana Slabakova (2006; 2008;2013), and concerns the developmental and final stages of L2 acquisition (Slabakova 2008:84). The hypothesis argues that functional morphemes and their features are the bottleneck of L2 acquisition, and thus the most challenging part for L2 learners to acquire. In other words, functional morphology is predicted to be more difficult than other linguistic domains, such as syntax and semantics, as well as both internal and external interfaces (internal and external interfaces are discussed in more detail in section 2.2.1). As mentioned, the current thesis compares the difficulty of narrow syntax[3] and functional morphology, and for that reason, I focus on the Bottleneck Hypothesis' predictions about these two domains in the following text.

Slabakova (2013:23) argues that knowledge about narrow syntax comes before accurate knowledge of functional morphology. A number of studies on child and adult L2 acquisition are used to illustrate this. These are summarized by White (2003a), and shown here in figure 1. The studies measured accurate production of the phenomena seen in figure 1 in obligatory contexts in English L2, produced by speakers with different L1s. For instance, Lardiere (1998b) investigated the fossilized, i.e. end-state, English L2 grammar of a native speaker of Chinese, Ionin and Wexler (2002) looked at children with Russian as their L1, and Haznedar (2001) investigated children with Turkish as their L1. As the table illustrates, the accuracy rates for the morpho-syntactic phenomena (subject-verb agreement on lexical verbs, past tense and suppletive forms) vary from 4.5 % to 90 %, whereas syntactic phenomena such as the verb staying in the VP, case, and overt subjects have accuracy rates between 98 and 100 %. In other

---

[3] A narrow, or core, syntactic operation means that no other domains than syntax are involved (Holmberg 2010: 39-40; Cook and Newson 2007: 48)

words, despite the fact that all phenomena are all related to the same functional category (IP), the accuracy rates are higher and more consistent with syntax (Slabakova 2013: 10).

These results are in support for the *syntax-before-morphology* view, which argues against the *morphology-before-syntax* view. That is, the latter view argues that acquisition of functional morphology drives acquisition of functional categories (see e.g. Clahsen, Penke and Perodi 1993/1994). According to Slabakova (2013: 10), this cannot be true, as the results from the studies illustrated in figure 1 show that speakers engage knowledge of syntactic phenomena related to the functional category IP, despite a low accuracy rate with functional morphological phenomena related to the same category. To exemplify, Lardiere (1998b: 359) argues that his results indicate that L2 speakers are able to determine feature strength, although verbal morphology is not acquired. In other words, knowledge of syntactic phenomena seems to have fallen in place before accurate knowledge of functional morphology. Based on this, one of the main predictions in the Bottleneck Hypothesis is that knowledge of syntax comes before knowledge of functional morphology (Slabakova 2013: 23).

Figure 1: Accuracy rate of phenomena related to the functional category IP in obligatory contexts

	3sg agreement on lexical verbs	Past tense	Suppletive forms: *Be* (aux/copula)	Overt subjects	Nom. case	V in VP (no raising)
Haznedar (2001)	46.5 %	25.5 %	89 %	99 %	99.9 %	–
Ionin and Wexler (2002)	22 %	42 %	80.5 %	98 %	–	100 %
Lardiere (1998a,b)	4.5 %	34.5 %	90 %	98 %	100 %	100 %

(Slabakova 2008: 102).

Furthermore, it is argued that the reason why acquisition of functional morphology is more difficult than acquisition of narrow syntax is related to the learning tasks involved. To specify, the argument is that narrow syntactic operations, unlike functional morphology, are transferable. Following the Full Transfer/Full Access Hypothesis (see section 2.1), this suggests that narrow syntax can be acquired by positive transfer or access to UG, whereas functional morphology must be lexically learned (Slabakova 2013: 14 and 25). Lexical learning is

considered the most challenging learning task, and for that reason, functional morphology is predicted to cause more problems in L2 acquisition than narrow syntax. This is supported by Ullmann (2001: 717), and the declarative/procedural model, who argues that acquisition of the Functional Lexicon, which is expressed by functional morphology in most cases, involves the declarative (explicit) memory, and acquisition of narrow syntax involves the procedural (implicit) memory[4]. These are both parts of long-term memory, and the difference is that whereas the declarative memory consists of memories that are explicitly, or consciously, recollected, the procedural memory consists of unconscious, internalized, memories (Ullmann 2001: 718). In the following paragraphs, I discuss the acquisition process of narrow syntax and functional morphology in more detail, including the Functional Lexicon.

It is a Minimalist assumption (see Chomsky 2000; 2001;2004;2005) that most language variation is encoded in the Functional Lexicon, which is a part of the language faculty, as illustrated in figure 2. Because this lexicon is usually expressed through functional morphology, it follows that most language variation is found in functional morphological forms (Slabakova 2013:10 and 14-15).

Figure 2: The language faculty
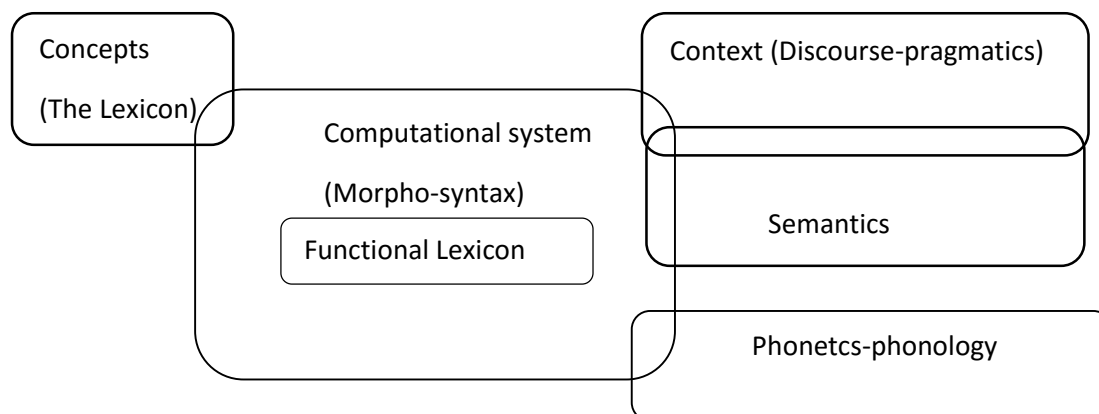
Note: the Functional Lexicon is a sub-module of the computational system, which is where syntactic operations take place (Slabakova 2013: 8).

---

[4] Here, it is important to note that implicit and explicit memory is not the same as implicit and explicit learning (Dekeyser 2003: 315). That is, it is possible that something that is explicitly acquired may become part of the implicit memory over time, as the learner may lose awareness of the structures that they have acquired Similarly, something that is implicitly acquired may become explicit, as the learner may be made aware of the structures (Dekeyser 2003: 315)

More specifically, the Functional Lexicon is formed by functional categories. Each category is associated with lexical items that are specified for formal features (Slabakova 2013: 8). Two formal features are relevant: uninterpretable features and interpretable features. The interpretable ones are semantic. That is, they contribute to grammatical meaning, such as gender, tense and aspect, which means that they cannot be eliminated before Spell-Out, i.e. they survive into the semantic system for interpretation, and are thus pronounced (Slabakova 2006: 305; 2008:9 2013: 8). Uninterpretable features, on the other hand, are deleted before Spell-Out, which means that they are purely formal in nature, and serve to establish syntactic dependencies rather than contributing to meaning (Slabakova 2006: 305; 2009: 61; 2013: 8). In other words, uninterpretable features are only relevant to the morpho-syntax of a sentence, such as case or agreement (Slabakova 2008: 9; Pesetsky and Torrego 2001: 3). To exemplify, the morphology in (13) does not only carry information about the interpretable feature, which is [singular], but also about the uninterpretable features, which are the features that ensure subject-verb agreement (Slabakova 2013: 8).

(13)   The girl walks to school

    Interpretable feature: [singular]
    Uninterpretable feature: subject-verb agreement

Consequently, when acquiring an L2, one must learn the formal features of a set of lexical entries in the Functional Lexicon, i.e. figure out which formal features are encoded in the target functional morphology. This is predicted to be problematic for L2 learners, as the combination of features varies from language to language, and thus, cannot be transferred from the L1 to the L2 (Slabakova 2013: 8).

Krashen (1981) supports this, as he claims that transfer of bound morphology is weaker than transfer of syntactic operations such as, for instance, verb movement. This argument is based on Duskova's (1969) investigation of Czech learners of English L2, which found only 19 out of 166 errors with morphology can be traced back to the influence of the L1. For instance, with respect to subject-verb agreement, the study found that although the speakers mark agreement for person and number on the finite verb in their native Czech grammar, there is a high number of errors with subject-verb agreement in English. This indicates that the influence from Czech is not the cause of errors in bound morphology in English L2 by Czech speakers (Krashen 1981: 65 and 66).

The above also suggests that functional morphology carries a high level of syntactic information, as the differences between languages are captured by the formal features encoded in functional morphology (Slabakova 2013: 15). This suggests that once the learners have acquired target morphology, complex L2 syntax should be less problematic (Slabakova 2013: 14). In other words, functional morphology is the bottleneck through which the learners must pass in order to acquire the target grammar. An example of complex syntax is long distance *wh*-movement, as illustrated in (14), in which the *wh*-phrase moves to a higher position than CP, which is where it is born (Santorini and Kroch 2007).

(14)   [$_{CP}$ What$_i$ did she say [$_{CP}$ that she was doing t_i]]?

To conclude this section, the Bottleneck Hypothesis argues that the reason why acquisition of functional morphology is challenging is that it encodes all of the grammar's non-transferable formal features. Rather than being able to acquire these through UG or positive transfer, the formal features must be lexically learned (Slabakova 2013: 14). Narrow syntax, on the other hand, is easier to acquire because it is processed by the means of universal operations, and is thus transferable. If the result of transfer is negative, the learners access UG to restructure their interlanguage grammar (Slabakova 2013: 14).

2.2.1   A contrasting view: the Interface Hypothesis

In this section, I describe the Interface Hypothesis (Sorace and Filiaci 2006; Sorace 2011), which is concerned with the final stage of L2 acquisition (Sorace 2011: 9). In comparison to the Bottleneck Hypothesis, it puts forward different predictions about what the most difficult parts of language acquisition are. More specifically, the Interface Hypothesis argues that the syntax-pragmatics interface (also referred to as the syntax-discourse interface), which is an external interface[5], is the most difficult part of L2 acquisition, and possibly unacquireable for L2 learners. These types of structures involve both syntactic and pragmatic constraints. An example of this is seen in the Italian sentences in (15), in which the discourse determines whether or not the subject can be dropped. The example is taken from Sorace (2011:2).

---

[5] Sorace distinguishes between internal and external interfaces (Sorace 2011: 9): the internal ones refer to the interface between sub-modules of language, i.e. two linguistic modules, like syntax-semantics for instance, whereas external interfaces refer to the interface between a linguistic module and a cognitive, non-linguistic system, such as the syntax-pragmatics interface (Sorace 2011: 6).

(15)

    a. Perchè Giovanna non è venuta?
       'Why didn't Giovanni come?'

    b. Perchè ___ non ha trovato un taxi
       'Because she couldn't find a taxi'

    c. Perchè lei non ha trovato un taxi
       'Because she couldn't find a taxi'

More specifically, Italian is a Null Subject (NS) language, which means that the subject is overtly expressed or dropped from the surface structure depending on pragmatic constraints. The subject is overtly expressed if it has not been previously mentioned, as exemplified in (15a) above. If the subject has already been mentioned, on the other hand, i.e. if there is a continuing topic rather than a topic shift, the subject pronoun is dropped, as seen in (15b) (Sorace 2011: 2).

Furthermore, Sorace and Filiaci's (2006) investigated how these pronominal subjects are interpreted by near-native speakers of Italian L2, in comparison to monolingual native speakers of Italian (Sorace 2011:2). The study found that whereas monolinguals only accept sentences like (15b), the near-natives accept both sentences like (15b) and (15c) (Sorace 2011: 2). In other words, the near-natives optionally express the overt subject although its antecedent is already mentioned. This suggests that there is an over-extension of the overt anaphora in the Italian L2 grammar (Sorace and Filiaci 2006: 340; Sorace 2011: 2).

However, a difference between near-native and native speakers was only found with respect to the interpretation of overt subject pronouns, and not for that of null subject pronouns, i.e. the near-native speakers did not differ significantly from the monolingual native speakers with respect to how they interpreted sentences with null subjects. To exemplify, in embedded structures like (16), the subject of the matrix clause ('the old woman') is interpreted as the antecedent of the null subject (Sorace 2011: 2). Sorace and Filiaci's (2006) argue that the interpretation of these structures involves purely syntactic properties, which is taken as evidence for the fact that syntactic operations are acquireable.

(16)   La vecchiettai saluta la ragazzaj quando PRO attraversa la strada

               'The old woman greets the girl when Ø crosses the road'

(Sorace 2011: 2)

Two propositions have been put forward in order to identify the sources of optionality in the structures that involve the syntax-pragmatics interface: the representational account and the processing resources account. In short, the representational account argues that optionality is caused by cross-linguistic influence from one grammar to the other, and thus that native monolinguals and near-native L2 speakers differ from each other at the level of knowledge representations (Sorace and Filiaci 2006: 340; Sorace 2011: 12). The result is under-specification of the constraints that determine whether a structure should have an overt subject pronoun, or a null subject (Sorace and Filiaci 2006: 340; Sorace 2011: 12). The processing resources account, on the other hand, argues that native monolinguals and near-native L2 speakers differ at the level of processing. That is, the fact that there are different types of information involved in external interfaces (non-linguistic and linguistic) makes it difficult to process these types of sentences, which results in optionality in the L2 grammar (Sorace and Filiaci 2006: 340; Sorace 2011: 12). In the following text, I show that the latter view most efficiently accounts for the patterns of optionality.

The representational account argues that the language with the most economical syntax-pragmatic system affects the language with the more complex system (Sorace 2011: 13). To exemplify, this suggests that the English grammar of near-native speakers of Italian affects the Italian grammar, irrespective of which one is the L1 and L2, and the result is an over-extension of overt subject anaphora in Italian L2 (Sorace 2011: 13). In more technical terms, the reason for this is that the interpretable feature [+Topic Shift] is underspecified in the L2 grammar (Sorace 2011: 13). As (17) illustrates, the overt anaphora receives the [+Topic Shift] feature in native monolingual Italian grammar, whereas it receives both the negative and the positive value in the near-native grammar, as illustrated in (18) (Sorace 2011: 13).

(17)   *Native monolingual grammar:*

               NULL → [-Topic Shift] OVERT → [+Topic Shift]

(18)   *Near-native L2 grammar:*

               NULL → [-Topic Shift] OVERT → [+Topic Shift] OVERT → [-Topic Shift]

However, this does not account for the fact that the same over-extension of overt subjects is observed in the L2 grammar of speakers of two NS languages, i.e. two languages with complex systems, such as Spanish-Italian (see Bini 1993), Greek-Spanish (see Margaza and Bel 2006; Lozano 2006), or Spanish-Brazilian Portuguese (Guido Mendes and Iribarren 2007), for instance (Sorace 2011: 14). For that reason, it has been argued that the representational view cannot account for optionality in L2 grammar (Sorace 2011: 14).

The processing resources account holds that inadequate processing resources are the source of the optionality (Sorace and Filiaci 2006: 341). More specifically, the argument is that native monolingual speakers are more efficient than near-native speakers when it comes to integrating information from different domains in language use (Sorace 2011: 14 and 20). Consequently, it is easier for L2 learners to acquire target-like structures that involve purely syntactic operations, as these are more economical to process than those involving interfaces (Sorace 2011: 15). This approach manages to account for the observed patterns of optionality in near-native L2 grammar regardless of language combination, as the structure of languages involved is irrelevant. As Sorace (2011: 20) states, the observed L2 patterns are the result of the fact that near-native L2 speakers compensate "... for occasional failure to compute the correct syntax-pragmatics mappings in real time".

2.2.2   The Bottleneck Hypothesis versus the Interface Hypothesis

Based on the above description, I now turn to discuss the Bottleneck Hypothesis and the Interface Hypothesis. As already described, the Interface Hypothesis holds that structures involving the syntax-pragmatics interface are possibly unacquirable in L2 acquisition (Sorace and Filiaci 2006: 341). Slabakova, on the other hand, argues that these structures are not the most problematic ones for L2 speakers, and refers to the fact that, although some studies find that this is extremely difficult, other studies find that the syntax-pragmatics interface is acquirable (Slabakova 2013: 21). One example is Valenzuela's (2006) study of the clitic-left dislocation (CLLD) in Spanish. The CLLD refers to the fact that a clitic is added to a structure only if the fronted object is specific (Valenzuela 2006: 286; Slabakova 2013: 20). This is exemplified in (19), in which the subject is specific in (19a), and non-specific in (19b). The examples are taken from Valenzuela (2006:286).

(19)

    a. El   libro,   lo   leí

       The   book,   CL   read.1sg

       'The book, I read'

    b. *Un   libro,   lo   leí

       A   book,   CL   read.1sg

       Intended: 'A book, I read'

In this study, near-native speakers of Spanish, with English as their L1, were tested in their knowledge of the discourse constraint that determines whether or not the clitic should occur (Slabakova 2013: 20). The results show that the near-natives have acquired the CLLD structure, however, they do not distinguish between the constructions in (19) to the same degree as the monolingual speakers do (Valenzuela 2006: 300). Nevertheless, the participants' use of clitics points in the same direction as the monolinguals' use. The fact that they cannot have acquired this knowledge from their L1, as there are no clitics in English, thus suggests that it is acquireable, according to Slabakova (2013: 20 and 21). In other words, Slabakova (2013) argues against the Interface Hypothesis with respect to how difficult the syntax-pragmatics interface is.

Sorace (2011: 26), on the other hand, argues that the Bottleneck Hypothesis is unclear with respect to what kind of bottleneck the functional morphology poses. This critique is based on the fact that the Bottleneck Hypothesis has not been tested on a near-native level, and not with various experimental methods. Based on this, Sorace argues that future research must attempt to tease apart the difficulties that are caused by the acquisition of representation, and those that are caused by processing of morphological exponents, as well as to test speakers at the near-native level (Sorace 2011: 26).

In other words, the Bottleneck Hypothesis and the Interface Hypothesis have different views on what the most difficult parts of L2 acquisition are. In this thesis, I only focus on the predictions that the Bottleneck Hypothesis makes, and for that reason, the Interface Hypothesis is not mentioned in the following text.

2.3 Constructions

In the current study, subject-verb agreement is used to test knowledge about functional morphology, and verb movement is used to test knowledge about narrow syntax. As previously mentioned, the purpose is to test if functional morphology is more challenging than narrow syntax in English L2 by native speakers of Norwegian. These constructions have been chosen on the basis of previous studies which are discussed in more detail in section 2.4. In short, Westergaard (2003) illustrates that acquisition of English word order is challenging for native speakers of Norwegian at an early state, whereas Slabakova and Gajdos (2008) show that subject-verb agreement is problematic, not only for beginners, but also for intermediate speakers of German L2. The latter has also been suggested by Ionin and Wexler (2002), Haznedar (2001) and Lardiere (1998a,b), as discussed in section 2.3. These observations indicate that both verb movement and subject-verb agreement should be challenging for Norwegian learners of English L2. However, which one of them that is more challenging has not yet been investigated, and for that reason, this is the purpose of the current study. In the following sections, I elaborate on the mismatches between Norwegian and English with respect to these two constructions.

2.3.1   Verb movement

As mentioned in section 2.1, Norwegian is a verb second (V2) language, which means that the finite verb occurs in the second position of main clauses. The exception is *yes/no*-interrogatives, and there is also some dialectal variation with respect to *wh*-interrogatives and so-called *om*-questions (see Vangsnes 1996; Westergaard and Vangsnes 2005; Westergaard 2009; Rognes 2011; Bentzen 2014; Larsson and Johannessen 2015:162). In main declarative clauses, on the other hand, the finite verb always occurs in the second position. This is illustrated in the declarative main clauses in (20) and (21), in which the finite verb is always in the second position.

(20)   *Finite lexical verbs*

    a. Kari **misforstod**   oppgaven   [Main clause]

        Kari misunderstood assignment.DEF

        'Kari misunderstood the assignment'

    b. Per **gikk** ofte til   butikken   [Main clause with adverb]

        Per went often to   shop.DEF

        'Per often went to the shop'

    c. I går   **misforstod** Kari oppgaven   [Non-subject-initial]

        Yesterday misunderstood Kari assignment.DEF

   'Yesterday Kari misunderstood the assignment'

(21)   *Finite auxiliary verbs*

    a. Kari **hadde** misforstått   oppgaven   [Main clause]

        Kari had   misunderstood assignment.DEF

        'Kari had misunderstood the assignment'

    b. Per **har** ofte gått   til   butikken   [Main clause with adverb]

        Per has often gone to   shop.DEF

        'Per has often gone to the shop'

    c. I går   **hadde** Kari misforstått   oppgaven   [Non-subject-initial]

        Yesterday had   Kari misunderstood assignment.DEF

        'Yesterday Kari had misunderstood the assignment'

This word order is best illustrates with non-subject-intitial declarative main clauses, such as (20c) and (21c) for two reasons: first, the surface structure of subject-initial declarative main clauses is not different from the surface structure of sentences with an SVO word order (*subject<verb<object*). Secondly, there is a lot of dialectal variation in Norwegian with respect to where the adverb occurs in sentences like (20b) and (21b) (e.g. Rognes 2011; Bentzen 2009), and in general, the placement of adverbs is a widely discussed topic. For these reasons, non-subject-initial declarative main clauses are used to illustrate the V2 word order in Norwegian in this thesis.

The V2 word order is caused by V-to-C (or V-to-I-to-C) movement. Movement must always be triggered, i.e. it cannot happen for no reason (see Pesetsky and Torrego 2001: 3), and it is argued that the trigger in this case is an Extended Projection Principle (EPP) in the C-domain, which requires C to be lexicalized (Westergaard and Vangsnes 2010: 130). In order to fill this requirement, the verb moves. The syntactic tree in figure 3 illustrates the fact that the finite verb has moved to the C-position of the sentence structure in order to fill the V2 requirement:

Figure 3
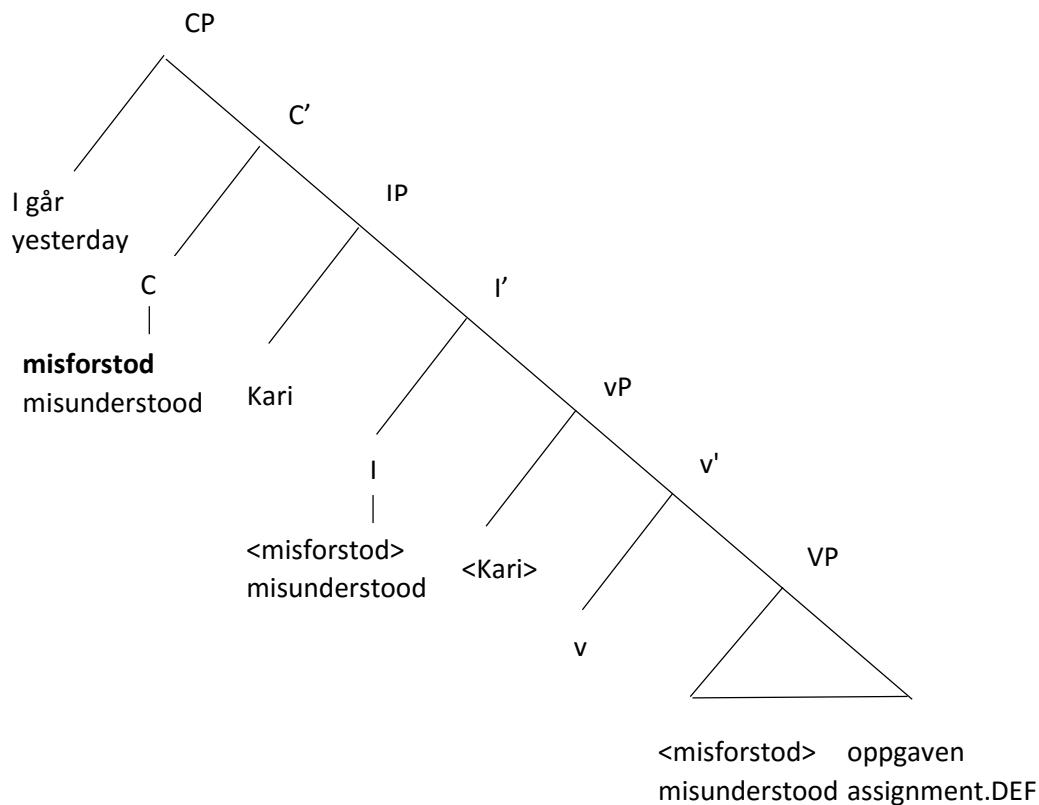
This further suggests that there should be no V2 pattern in embedded clauses in Norwegian, as the movement to C is blocked by a complementizer in the C-position. This is illustrated in example (22) and in figure 4, in which the verb *misforstod* ('misunderstood') does not move, and thus, occurs below the subject in the embedded clause.

(22)   I går   spurte Per om Kari **misforstod**   oppgaven
         Yesterday   asked Per if Kari misunderstood assignment.DEF
         'Yesterday Peter asked if Kari misunderstood the assignment'

Figure 4
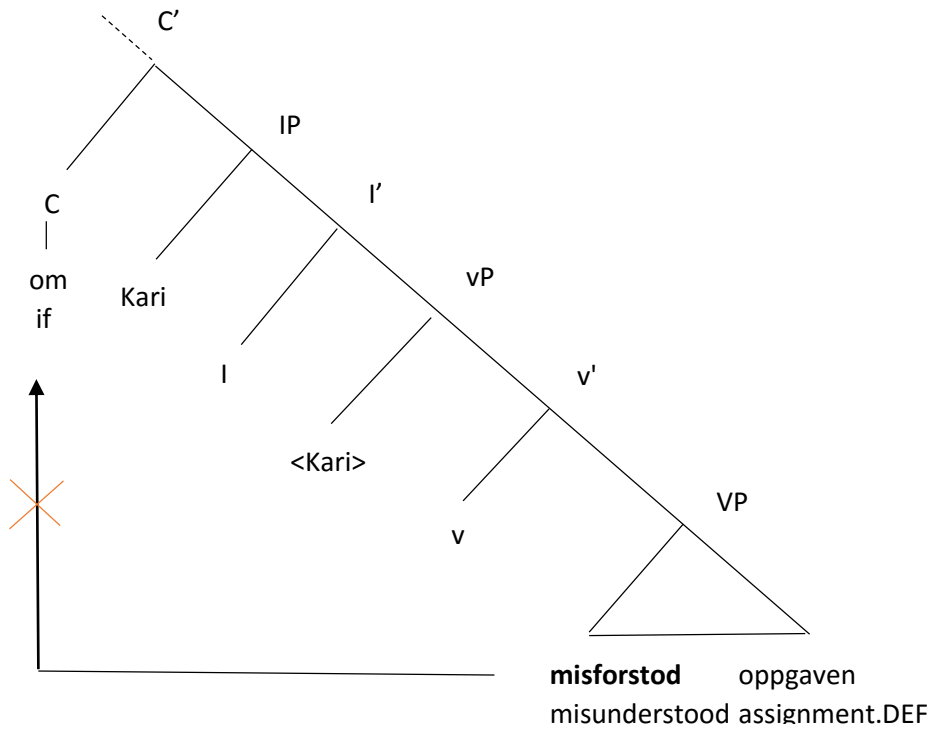
English, on the other hand, is not a V2-language, but rather, an SVO-language. This means that the verb always follows the subject in main declarative clauses. According to Kayne (1994), this is the universally underlying word order, which suggests that nothing moves. In other words, unlike Norwegian, there is no V2 requirement in English, i.e. there is not a strong EPP-feature in the C-domain that needs to be lexicalized. Instead, the lexical verb stays in the VP. This is illustrated in the declarative clauses in (23). As the difference between the SVO and the V2 word orders is best illustrated in non-subject-initial declarative clauses like (23c), the English syntactic structure of this sentence type is illustrated in figure 5.

(23)   *Lexical verbs*

    a.   Peter **often** went to the shop   [Main clause with adverbs]

    b.   Mary asked if Peter often **went** to the shop [Embedded clause with adverbs]

    c.   Yesterday Peter **went** to the shop   [Non-subject-initial main clause]

Figure 5
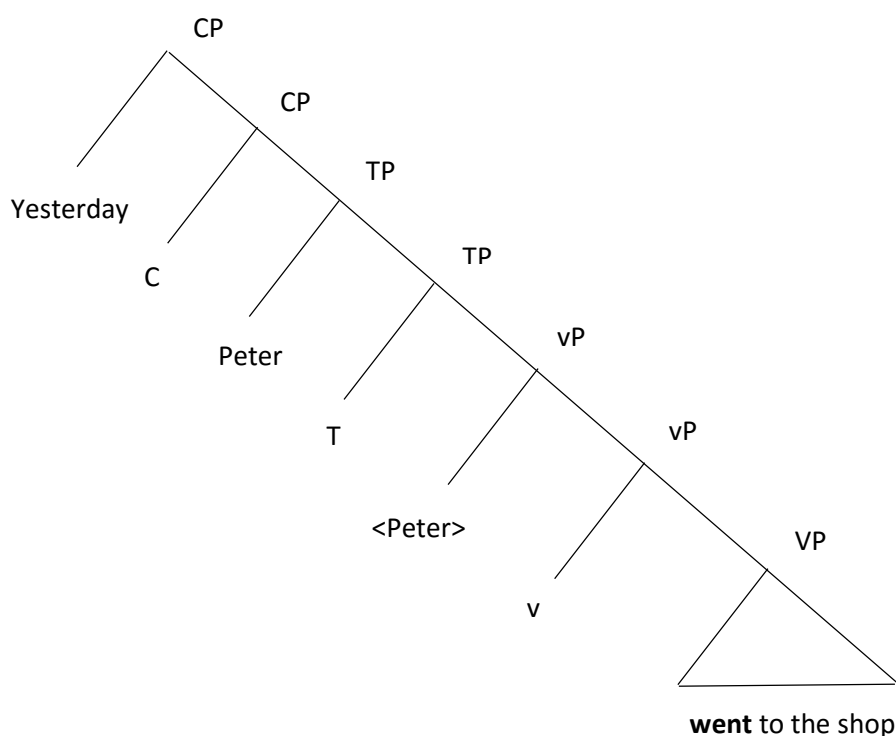
Moreover, although the lexical verb never moves out of the VP, English allows movement to the CP-domain in questions (I-to-C movement). In these cases, the auxiliary verb moves above the subject (Westergaard 2003: 78). This is illustrated in example (24). However, this does not affect the mismatch between Norwegian and English in non-subject-initial declarative clauses, as it only happens in interrogatives. This is exemplified in (25).

(24)   Where **did** John go yesterday?

(25)   I morgen   **skal** Jon   gå til butikken
             Tomorrow shall John go to shop.DEF
             'Tomorrow John **will** go to the shop'

Based on Kayne's (1994) proposal of the underlying SVO word order, one can say that the learning task involved for Norwegian learners of English L2 is to unlearn the V2 rule, rather than to learn the SVO rule. That is, during L1 acquisition of Norwegian, the learners should initially assume that the word order is SVO, until they are exposed to input cues which tell them that Norwegian is a V2 language. In other words, L1 speakers of Norwegian first have to acquire the V2 rule during their L1 acquisition, and then, they must unlearn it during L2 acquisition of

English. The mismatch between English and Norwegian is repeated here in examples (26) and (27), for convenience.

(26)   Norwegian: I går   **gikk**   Jon   til   butikken

                                   Yesterday   went Jon to   shop.DEF

               English: Yesterday John **went** to the shop

(27)   Norwegian: I morgen   **skal**   Jon   gå   på   kino

                                   Tomorrow   shall Jon go   on   cinema

               English: Tomorrow John is going to the cinema

2.3.2 Subject-verb agreement

With respect to morphology, there is a mismatch between overt subject-verb agreement in Norwegian and English. That is, Norwegian has no overt agreement morphology, whereas English marks the verb when the subjects are 3rd person singular. In the following text, I illustrate this difference by using subject-initial main declarative clauses.

First, in Norwegian, the present tense is marked with the suffix –*r* on the verb (Enger and Kristoffersen 2000: 83). This is exemplified in (28), where the suffix is marked in bold.

(28)   Per   snakke**r**   norsk

            Per   speaks   Norwegian

            'Per speaks Norwegian'

Furthermore, as mentioned, there is no overt morphology agreement in Norwegian, which means that the verb *snakke,* ('speak'), in (28) does not change its form, regardless of the subject's number and person. This is illustrated in (29).

(29)

        a. Per   og   Mari   snakke**r**   norsk

             Per   and   Mari   speak   Norwegian

             'Per and Mari speak Norwegian'

b. Jeg   snakke**r**   norsk

I   speak   Norwegian

'I speak Norwegian'

c. Per   snakke**r**   norsk

Per   speaks   Norwegian

'Per speaks Norwegian'

In English, the verb agrees with the number and person features of the subject, as present tense verbs receive the suffix *–(e)s* when the subject is 3rd person singular[6]. This is exemplified in (30) and (31). As the latter example shows, the verb is bare when the subject is not 3rd person singular.

(30)

    a. Peter speak**s** English

    b. Peter go**es** to the cinema every Friday

(31)

    a. Peter and Mary speak English

    b. I speak English

In other words, English overtly marks the agreement between the subject and present tense verbs, which shows that there is a mismatch between English and Norwegian. Consequently,

---

[6] In English there are two exceptions with respect to subject-verb agreement when the subject is 3rd person singular. First, modal auxiliaries show no agreement, as illustrated in (i). Secondly, the verb *be* has several different forms. That is, in the present tense, there are three forms when the subject is singular (*am, are* and *is*), and one form when the subject is plural (*are*). In the past tense, *be* has two forms when the subject is singular (*was* and *were)* and one form when it is plural (*were*). This is exemplified in (ii). (Dypedahl et. Al 2002: 103)

  i.   a. John should work harder

       b. John and Mary should work harder

  ii.   a. John is/was here

       b. John and Mary are/were here

the learning task for Norwegian speakers of English L2 with respect to subject-verb agreement is to lexically learn the target agreement morphology in English.

2.4 Previous research on L2 acquisition of agreement and verb movement

In this section, I discuss previous research on L2 acquisition with focus on the two constructions discussed in section 2.3, i.e. agreement and verb movement.

2.4.1   Dröschel (2011)

Dröschel (2011) addresses the grammar of non-native English spoken as a lingua franca in Switzerland. The informants' native languages are French (34 % of the informants), German (51 %) and Italian (11 %), and some of the speakers have two native languages (2 %). The study is based on a Swiss English database, created by a number of researchers for a project called *Language Contact and Focusing: The Linguistics of English in Switzerland* (SNSF) (Dröschel 2011: 151). The database includes both spoken and written material, but the emphasis is on spoken data, which constitutes 71 % of the database (Dröschel 2011: 161). Examples of the spoken data are interviews and recordings from meetings, lectures and conferences, and examples of the written data are e-mails sent between medical students and applications for a Swiss business school (Dröschel 2011: 156-157). There are 167,086 words in the database, and 94 informants in total. The informants' age ranges from 20 to 59 (Dröschel 2011: 160).

One of the linguistic phenomena Dröschel (2011) discusses is agreement errors with simple present tense verbs. These errors are split up into two groups: Utterances with verbs that have not received the 3rd person singular mark –s although the subject is singular, and verbs that have received an –s when the subject is plural. Dröschel (2011: 213) refers to the former type as missing 3rd persong sg –s, and the latter error type as hypercorrection. These are exemplified in (32) and (33), which are taken from the Swiss English database mentioned above (Dröschel 2011: 214). The suggested explanation of the occurrence of both error types is a regularisation process of English, i.e. simplification of the verb morphology in English (Dröschel 2011: 218).

(32)   *3rd person singular –s is missing*

          But I-I try it and er after that Simon er and I go through er and er he give me
          some I don't know how to say – Verbesserungen?

(33)   *Hypercorrection of 3rd person singular –s*

So you should drive down the level to a er to a level that all people, all involved
per-people understands

Furthermore, the Swiss English database shows that there are slightly more errors with the
missing 3rd person singular –s, rather than hypercorrection, as there are 44 instances of the
former, and 30 instances of the latter (Dröschel 2011: 214). Both types of agreement errors have
also been reported in other studies of overt agreement morphology in English a non-native
language (see e.g. Breiteneder 2005; Hülmbauer 2010; Vettorel 2014). In Vettorel's (2014)
study, there were also more errors with the missing 3rd person singular –s.

2.4.2   Håkansson and Collberg (1994)

In this study, Håkansson and Collberg (1994) look at L1 and L2 acquisition of Swedish, with
focus on the word order in sentences that contain modal auxiliary verbs and negation. The target
word order in Swedish is *modal< negation* in main declarative clauses like (34a), and
negation<modal in embedded clauses, as exemplified in (34b)

(34)

    a. Barn **kan inte** tala på samma sätt som vuxna

       children cannot speak in the same way as adults

       'Children cannot speak the same way as adults'

    b. Vi vet [att barn **inte kan** tala på samma sätt som vuxna]

       we know that children not can speak in same way as adults

       'We know that children cannot speak in the same way as adults'

(Håkansson and Collberg 1994: 96)

In L2 acquisition of Swedish word order, studies have shown that the learners go through four
stages with respect to their placement of negation in relation to the finite verb (see e.g.
Hyltenstam 1977; 1978; Bolander 1988; Colliander 1993). It is not until the last stage that the
non-target word order *modal<negation* is abandoned (Håkansson and Collberg 1994: 95). The
same was found for children who acquire Swedish L1, i.e. that they also initially prefer the non-
target like word order *modal<negation,* and do not acquire the target order until the final stage.
These results come from the Child Language Syntax Project (Söderberg 1974) in which the

speakers' age range from 2-3 years old, and their preference for *modal<negation* lasts for approximately one month (Håkansson and Collberg 1994: 96). In addition, the same result was found in Håkansson's (1989) experimental study of L1 speakers of Swedish. In other words, both L1 and L2 learners initially assume the word order *modal<negation* in embedded clauses. Figure 6 illustrates the non-target like word order, and figure 7 illustrates the target word order in Swedish embedded clauses.

Figure 6:

```
                  CP
                 /  \
                /    C′
               /    /  \
           Spec    /    IP
                  /    /  \
                 C    /    I′
                 |   Spec / \
                som      /   VP
                that    I   /  \
                        |  /    VP
                      ville NEG /  \
                      would  | Spec V′
                            inte     |
                            not      V
                                    gå
                                    go
```

Figure 7:

CP
　Spec
　C'
　　C
　　　som
　　　that
　　IP
　　　Spec
　　　I'
　　　　I
　　　　NEG
　　　　　inte
　　　　　not
　　　　VP
　　　　　Spec
　　　　　VP
　　　　　　V
　　　　　　　ville
　　　　　　　would
　　　　　　V'
　　　　　　　Spec
　　　　　　　VP
　　　　　　　　V
　　　　　　　　　gå
　　　　　　　　　go
　　　　　　　　V'
　　　　　　　　　(XP)
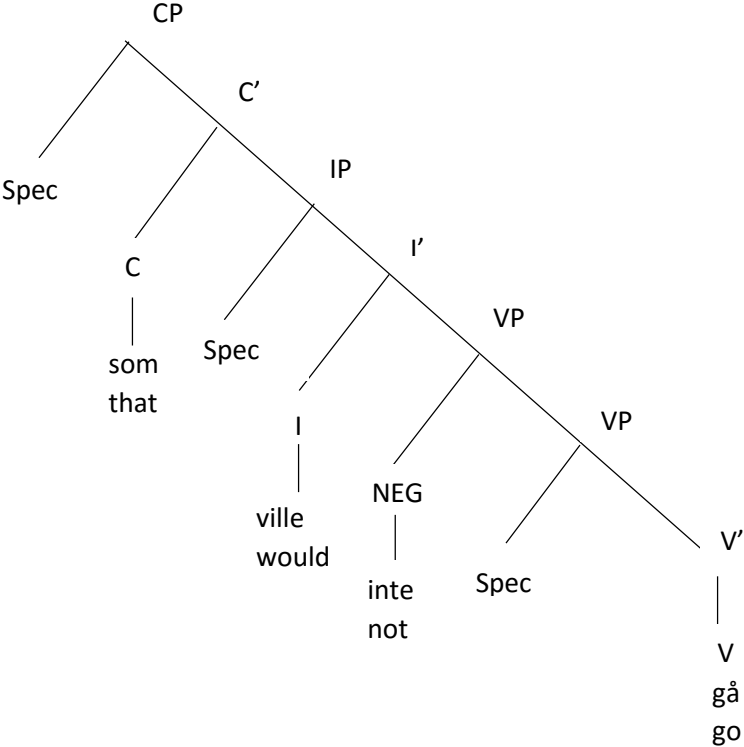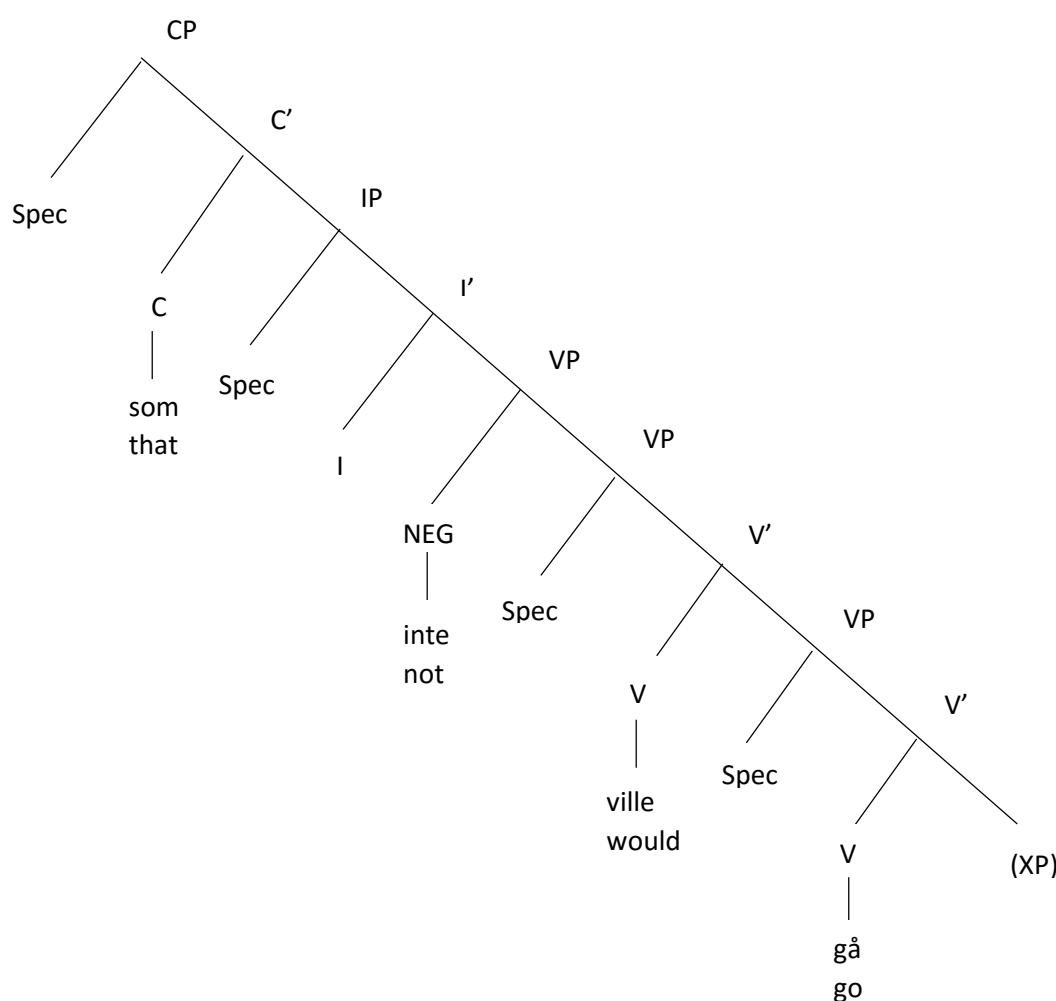
Håkansson and Collberg's (1994) explanation of why *negation<modal* is delayed in Swedish builds on the claim that the unmarked placement of auxiliary verbs in UG is in the inflectional phrase (IP), whereas Swedish, Danish and Norwegian auxiliaries are generated in the VP-domain, i.e. Swedish speakers place the auxiliaries in a marked position (Håkansson and Collberg 1994: 96 and 97). In contrast, it is argued that languages like English generate auxiliaries in IP (Håkansson and Collberg 1994: 103). Their argumentation for this claim is that Norwegian, Swedish and Danish modal auxiliaries do not behave in the same manner as the English modals do. That is, in English, only the auxiliaries are affected by inversion, as illustrated in (35), they are the only verb type that can be used with sentential negation (example (36)), and they are the only ones that can contract with negation (example (37)). In addition, modal auxiliaries do not inflect in English, as illustrated in (38), and they do not have regular infinitive forms (example (39)) (Håkansson and Collberg 1994: 104).

(35)

      a. **Can** you understand this?

      b. ***Understand** you this?

(36)

      a. You **cannot** understand this

      b. *You **understand** not this

(37)

      a. You **can't** understand this

      b. *You **understandn't** this

(38)

      a. The tourist **can** speak English

      b. The tourist **speaks** English

(39)

      a. *To can

      b. To understand

In contrast, Norwegian, Danish and Swedish modal auxiliaries behave differently. Håkansson and Collberg (1994) use Swedish modal auxiliary to illustrate this, and I extend this explanation to Norwegian modals. First, there are no subject-auxiliary inversion, as all verbs are all placed above the subjects in *yes/no* interrogatives (example (40)), and secondly, the auxiliaries have regular infinitive and participle forms, as exemplified in (41).

(40)

      a. Kan   du   **forstå**   dette?
         Can   you   understand   this?
         'Can you understand this?'

b. Forstår   du   dette?

Understand   you   this?

'Do you understand this?'

(41)   Å   kunne

To can

'can'

Furthermore, as the unmarked representations are preferred by default, the learners assume that the auxiliaries are placed in unmarked position (IP) until the input shows them that they need to restructure their grammar (Håkansson and Collberg 1994: 118). This leads to the initial non-target word order *modal<negation* in acquisition of Swedish embedded clauses. Lexical verbs, on the other hand, are generated in the VP, which is the underlying position in UG (see Kayne 1994), i.e. this causes no problems for the word order *negation<lexical verbs.*

As Håkansson and Collberg's (1994) study of Swedish L1 learners' embedded clauses shows, word order is acquired in a sequence of four stages, in which the target word orders *negation<modal* occurs in the fourth stage and *negation<lexical verb* occurs in end of the second stage (Håkansson and Collberg 1994: 108). Based on this, they put forward the following hypothesis: "…the difficulty lies in a series of parametrically related characteristics which make Swedish (and Danish and Norwegian) a marked type of language among the other recognized verb second languages" (Håkansson and Collberg 1994: 96).

2.4.3 Ocampo (2013)

Ocampo's (2013) dissertation aims to investigate how structural distance, as well as the number of the subject, affects subject-verb agreement. In order to do this, she uses three groups of participants: the first group consists of 20 Spanish leaners of English L2, the second group is a control group with 28 native speakers of English, and the third group consists of 28 native speakers of English who participate in a stressed test (Ocampo 2013: 20). The latter group is henceforth referred to as the stressed English L1 group. The research questions are the following:

- "Will L2 learners of English be less sensitive to subject-verb agreement errors when agreement is established across a more structurally complex intervening phrase?" (Ocampo 2013: 17)

- "Will the effects of plural markedness facilitate successful subject-verb agreement resolution for L2 learners as structural distance between the two agreeing elements is increased?" (Ocampo 2013: 18)

- "Will native speakers placed under a processing burden show variability similar to the L2 learners in their processing of agreement?" (Ocampo 2013: 19)

To elaborate, the following is tested in the study: first, whether or not more structural distance between the subject and the agreeing verb causes more problems for subject-verb agreement, or as Ocampo (2013: iii) states, whether or not this affects the speakers "…sensitivity to agreement violations". Secondly, the study tests if the number of the subject affects the speakers' performance in subject-verb agreement, and finally, if there are learner-like patterns in the stressed English L1 group.

In research question 1 and 2, only the English L2 learners and the control group were tested. In order to address the first question (whether or not structural distance between the subject and the verb affects sensitivity to subject-verb agreement errors), prepositional phrases and more complex relative clauses are used (Ocampo 2013: 41). These are exemplified in (42) and (43), in which (a) and (b) illustrate sentences with singular subjects, and (c) and (d) illustrate sentences with plural subjects. The examples are taken from Ocampo (2013: 22). In order to test the second research question, sensitivity to agreement errors with plural subjects was compared to sensitivity to agreement errors with singular subjects (Ocampo 2013: 41).

(42)   *Prepositional phrase intervener*

    a. The tourist *in warm southern Mexico* often fishes in the ocean

    b. *The tourist *in warm southern Mexico* often fish in the ocean

    c. The tourists *in warm southern Mexico* often fish in the ocean

    d. *The tourists *in warm southern Mexico* often fishes in the ocean

(43)   *Relative clause intervener*

    a. The tourist *who hunted in Mexico* often fishes in the ocean

    b. * The tourist *who hunted in Mexico* often fish in the ocean

    c. The tourists *who hunted in Mexico* often fish in the ocean

    d. *The tourists *who hunted in Mexico* often fishes in the ocean

For the third research question, the stressed English L1 group completed the same test as the L2 speakers and the control group, but with an additional digit load, i.e. an additional cognitive burden, when they processed long-distance subject-verb agreement (Ocampo 2013: 41). The purpose for this was to investigate whether there are qualitative differences between L1 speakers and L2 speakers with respect to storage and processing of agreement, which may contribute to the identification of the sources of variability (Ocampo 2013: 2). To specify, Ocampo (2013: 3) states that the purpose is to "…further contribute to the discussion of morphological variability as a problem either in grammatical deficiency or processing limitations". This topic, i.e. the nature of the variability in morphology has been widely debated in L2 acquisition research. The fact that there is variability in L2 morphology, on the other hand, is well documented, both in production studies (Lardiere 1998a; White 2003c) and in comprehension studies, such as acceptability judgement tests or reading times (Johnson and Newport; 1989; Keating 2009; Clahsen et al. 2010; Jiang 2004; Slabakova and Gajdos 2008). In both production and comprehension studies, errors in agreement has been observed even at advanced stages (Ocampo 2013: 2).

Moreover, previous research has shown that in both L1 acquisition (see e.g. Bock and Miller 1991; Bock and Cutting 1992) and L2 acquisition (see e.g. Franck et al. 2002; McCarthy 2008; Lopez-Prego 2012), accurate long-distance agreement is affected by the number of the intervening noun, i.e. the noun that is closest to the verb. This phenomenon is called attraction errors, and refers to the fact that speakers often make the verb agree with the noun that is immediately preceding it, rather than with the subject (Franck et al. 2002: 371). This is seen in sentences in which the subject and the intervening verb disagrees in number, such as those exemplified in (42) and (43).

Most studies have found that attraction errors usually happens when the intervening noun is plural, i.e. when the subject is singular (see e.g. Bock and Miller 1991; Franck et al. 2002). To exemplify, this means that speakers more often produce incorrect agreement in sentences like (42a) and (43a) than in sentences such as (42c) and (43c) (Ocampo 2013: 10-11). These findings have led to the argument that the feature [plural] is marked in English. That is, when an intervening noun is plural, it disrupts the establishment of agreement between the subject and the verb, whereas when the intervening noun is singular, i.e. unmarked, there is no disruption. This is referred to as the plural markedness effect (Ocampo 2013: 11 and 12).

With respect to Ocampo's (2013) experiment, she used a moving window self-paced reading task which all participants completed individually. The task is the same for both the L2 group and the L1 groups, except for the fact that additional stress is added to the task for the stressed English L1 group, as mentioned previously (see Ocampo 2013: 14). To elaborate, the L2 learners (and the L1 control group) were first exposed to a series of dashes on a blank computer screen which concealed a sentence. The participants then clicked the mouse button to read the sentence word for word. Only one word was visible at all times. A control question was presented after the last word of each sentence, and the participants had to answer "yes" or "no" (Ocampo 2013: 24). The additions in the stressed English L1 group's task was the following: the participants first saw a fixation cross at the centre of the screen, which lasted for 1500 ms, before they were presented with a string of six digits for 3000 ms. Following this, they progressed to the same reading task as the L2 group (and the L1 control group). When the participants in the stressed English L1 group had answered the control questions, a six-digit number was presented again, and the participants' task was to determine if this string of numbers was the same as the one they saw at the beginning of the test (Ocampo 2013: 25).

The results showed that subject-verb agreement is affected by the sentences' structural distance. That is, the participants in the L2 group were sensitive to agreement errors in sentences with prepositional phrases, but not in sentences with the more complex relative clauses. The L1 control group, on the other hand, were sensitive to errors in both sentence structures (Ocampo 2013: 43). The participants who completed the stressed test, i.e. the stressed English L1 group, showed the same behaviour as the L2 group: they were less sensitive to agreement errors in sentences with relative clauses, in comparison to sentences with prepositional phrases (Ocampo 2013: 50).

Another result was that, with sentences that contained relative clauses, the participants only showed sensitivity to agreement errors when the subject was plural (Ocampo 2013: 50). For this reason, Ocampo (2013: 46) suggests that the plural markedness effect seems to be present. However, she argues that this effect is weak, as it for instance was only found in the prepositional phrases in the L2 group, and not in the relative clauses. In addition, no such behaviour was found in the results of the L1 control group.

Based on this, Ocampo (2013) argues that her findings illustrate that subject-verb agreement in English L2 acquisition is affected by the structural distance in sentences with long-distance agreement. In addition, she argues that the fact that similar results were found for native speakers of English indicates that processing limitations are the source for variability in long-distance subject-verb agreement, i.e. that the establishment of long-distance agreement is based on non-linguistic factors (Ocampo 2013: 43 and 50). Furthermore, she holds that it is hard to maintain that there is a strong plural markedness effect, as there were few differences between the sensitivity of agreement errors on the basis of the subjects' number (Ocampo 2013: 56).

2.4.4 Slabakova and Gajdos (2008)

Slabakova and Gajdos' (2008) experimental study of German L2 speakers shows that functional morphology is difficult in L2 acquisition. In this study, beginners and intermediate learners of L2 German, with English as their L1, were tested in their knowledge of the German copula *sein* ('be'). The participants are University students of German, and their proficiency was determined on the basis of how many hours of classroom teaching they had. The beginners had 40 hours of classroom instruction, and the intermediate learners had 140 hours.

The test consisted of 40 sentences that lacked subjects, of which 10 were fillers. In the remaining 30 sentences, there were six sentences for each form of *sein: bin, bist, ist, sind* and *seid.* The participants' task was to choose the correct subject out of four alternatives, of which more than one alternative may be correct. There were both full DP subjects and pronominal subjects. An example from the test is illustrated in (44), in which the only correct answer is *du* ('you').

(44)   ____ bist ein guter Freund

          are a good friend

     ☐ Moritz

     ☐ Du
        'you'

     ☐ Die Shüler
        'the students'

     ☐ Er
        'he'   (Slabakova and Gajdos 2008: 39)

The results of the experiment are illustrated in figure 8, in which the error rate is higher for both proficiency groups when the subjects are full DPs, in comparison to pronominal subjects. Furthermore, the error rate is higher for the intermediate group than for the beginners when there are DP subjects, which indicates that knowledge of subject-verb agreement with full DPs does not improve much, despite more exposure to German. Based on these results, Slabakova and Gajdos's (2008) study indicates that acquisition of functional morphology is difficult, and in particular, with sentences that have full DP subjects.

Figure 8: Percentage errors in all forms of sein *depending on type of subject*

Type of errors	Beginners	Intermediate learners
Errors in choosing correct pronoun subjects	7.50	4.50
Errors in choosing correct DP subjects	20.18	29.80

(Slabakova and Gajdos 2008:40)

2.4.5   Westergaard (2003)

Westergaard (2003) investigates transfer of word order from Norwegian L1 to English L2 in the grammar of primary school students between the age of 7 and 12. As the following text illustrates, this experimental study found that unlearning the V2 rule is difficult, and requires exposure to certain syntactic structures in English which shows the learner that English is an SVO language, rather than a V2 language.

To briefly repeat the Norwegian and English word order (see section 2.3.1), the verb moves to the C-domain in declarative main clauses in Norwegian, whereas it stays in the VP in English. In other words, Norwegian has V-to-I-to-C movement (Westergaard 2003: 78). This leads to mismatches between Norwegian and English word order in some types of main clauses, such as, for instance, non-subject-initial declaratives and sentences with adverbs. This is illustrated in examples (45) and (46), in which the verb occurs in the second position in Norwegian, and in the third position in English.

(45)   I går   **gikk** Peter på kino   [non-subject-initial declarative]
           Yesterday went   Peter to   cinema.DEF
           'Yesterday Peter **went** to the cinema'

(46)   Peter **går**   ofte på kino   [Subject-initial declarative with adverb]
           Peter goes often to cinema.DEF
           'Peter often **goes** to the cinema'

Furthermore, although the lexical verb never moves out from the VP domain in English, the auxiliary can move in certain sentence types, as illustrated in (47) and (48). This is because English allows I-to-C movement in questions, as explained in section 2.3.1, but also V-to-I movement in declarative clauses (Westergaard 2003: 78). Consequently, the auxiliary verb is in the second position of these sentences in English, whereas the lexical verb stays in the VP. In Norwegian, the auxiliary verb is in the second position of both the declarative clause and the question in (47 and (48)

(47)   Peter **har**   alltid   **elsket** kino
           Peter has   always loved   cinema.DEF
           'Peter **has** always **loved** the cinema'

(48)   Hvor **gikk** Peter i   går?

     Where went Peter in   yesterday?

    'Where **did** Peter **go** yesterday?'

Regarding Westergaard's (2003) study, there are 100 participants in the experiment, and these are all Norwegian school children who attend the 2nd to 7th grade of primary school (7 to 12 years old). In order to test their knowledge of English word order, the older children were given a written test (5th to 7th grade) that consisted of an acceptability judgment test, an assessment of sentence pairs, and an elicited production task (Westergaard 2003: 80). The younger children (2nd to 4th grade) were given an oral test that mirrored the written one, but only the 4th graders were given the elicited production task. An example of a sentence pair from the test is seen in (49). The word order in (49a) is grammatical in English, as the verb stays in the VP, and the word order in (49b) is ungrammatical in English, as the verb has moved to the second position of the sentence. As suggested above, the latter is the grammatical word order in Norwegian. In other words, if the participants use sentences like (49b), it suggests that they transfer their V2 word order to English.

(49)

   a. Every day John plays football

   b. Every day plays John football   (Westergaard 2003: 79)

The results from the experiment showed that there is considerable transfer from Norwegian L1 to English L2 in all age groups (Westergaard 2003: 86). In non-subject-initial declarative main clauses, 5th graders had 14 % correct performance on English word order, the 6th graders had 38 % correct performance, and the 7th graders had 61 % correct performance (Westergaard 2003: 86). This illustrates that, first, transfer happens not only among the youngest speakers, but also, after many years of classroom teaching. Secondly, the results show that a developmental jump happens in the 7th grade.

The suggested reason for this frequent transfer is that the learners are not exposed to the necessary input cues. This argument follows the cue-based approach to L2 acquisition, which assumes full, or partial, transfer from the L1 to the L2. That is, the learners initially transfer their L1 system to the L2, and then change their L2 system on the basis of input cues that show them that the new language is different from their native one. It follows that if the learners are

not exposed to these input cues, they continue to assume that the target system is the same as their L1. Westergaard (2003) identifies two input cues that are necessary for the learners to realize that English is not a V2 language: Non-subject-initial declarative main clauses and sentences with *do*-support. The former is exemplified in (45), and the latter is exemplified in (48). However, non-subject-initial main clauses are infrequent in English, and sentences with *do*-support are not introduced in the instruction material until the 7th grade (Westergaard 2003: 91). This suggests that it is difficult for the learners to discover that English is an SVO language. The fact that the developmental jump happens in the 7th grade, i.e. at the time in which *do*-support is introduced, supports this argument.

3   Research questions and methodology

In the following text, I first describe the research questions in section 3.1, as well as the hypothesis and predictions in section 3.2. In section 3.3, I discuss the acceptability judgement test, which is the method used in this thesis. In section 3.4, I briefly describe the pilot experiment, and in section 3.5, I discuss the main experiment, including the procedure, the sentences, the acceptability judgement test, and the participants. The project has been approved by the Norwegian Centre of Research Data (NSD).

3.1 Research questions

The following three research questions are addressed in the current study:

RQ1: Is functional morphology more difficult than narrow syntax in L2 acquisition?

RQ2: Is functional morphology a more persistent problem than narrow syntax?

RQ3: Which of the syntactic and morphological conditions are more difficult?

Research question 1 and 2 are raised in order to test the Bottleneck Hypothesis' predictions about functional morphology and narrow syntax (see section 2.2). Question 1 is addressed by comparing the participants' acceptability judgements of agreement and verb movement (the sentences are discussed in section 3.5.2), and the participants are then split up into proficiency groups and compared in order to address research question 2. The reason why research question 3 is raised is that different sentence structures are used to test agreement and verb movement: long distance-agreement and local agreement with singular and plural subjects, and non-subject initial declarative main clauses with lexical verbs and auxiliary verbs, respectively (see section 3.5.2). Previous research has shown that the difficulty of acquiring both agreement and verb movement varies on the basis of the sentence structure (see section 2.4), and for that reason, it is necessary to identify the more challenging morphological and syntactic conditions. The way in which this research question is approached is by comparing the participants' judgements of the different morphological constructions and syntactic constructions in turn.

3.2 Hypothesis and predictions

The research hypothesis, or alternative hypothesis, is that it is more difficult for Norwegian learners of English L2 to acquire agreement than it is for them to acquire verb movement. The null hypothesis is that there is no difference between how difficult acquisition of agreement and

acquisition of verb movement is. This is a directional hypothesis, which means that there is an expectation about the participants' behaviour (Levshina 2015: 9). The following predictions are based on this hypothesis:

Prediction 1: Subject-verb agreement is more difficult than verb movement

Prediction 2: Subject-verb agreement is a more persistent problem than verb movement

Prediction 3: Syntactic conditions

> Word order in sentences with auxiliary verbs is more difficult than in sentences with lexical verbs.

Prediction 4: Morphological conditions

> A. Long-distance agreement is more difficult than local agreement
>
> B. Long- distance agreement is more difficult when the subject is 3rd person singular, in comparison to when it is 3rd person plural
>
> C. Local agreement is more difficult when the subject is 3rd person singular, than when it is 3rd person plural

Prediction 1 and 2 are based on the Bottleneck Hypothesis' predictions, i.e. that functional morphology is the most challenging part of L2 acquisition, and that narrow syntax comes before accurate knowledge of functional morphology. As explained in section 2.2, these predictions are based on observations in a number of studies on English L2 acquisition, summarized by White (2003a). The studies show that speakers are considerably more accurate in their use of syntactic phenomena in obligatory contexts, such as the verb staying in the VP, than they are with functional morphology, such as subject-verb agreement.

Predicition 3 is based on the fact that auxiliary verbs and lexical verbs behave differently, as disussed in section 2.3 and 2.4.2. Håkansson and Collberg's (1994) claim is that Swedish, Danish and Norwegian marked languages, as they generate the auxiliary verbs in the VP. In contrast, English auxiliary verbs are generated in the unmarked IP position, which is preferred by default in language acquisition (see section 2.4.2). This means that when Norwegian L1 speakers acquire English L2, they must not only unlearn the V2 rule, but also the unmarked position of the auxiliary verb. However, this should not affect the difficulty of acquiring the target English word order *XP<subject<modal* in comparison to *XP<subject<lexical verb,* as the surface structure in these sentence types are the same, regardless of whether or not the auxiliary is placed in VP or IP. That is, regardless of whether or not the participants have

unlearned the marked position of the auxiliary verb, the surface structure is correct in English, as long as they have unlearned the V2 rule. In other words, the question investigated here is whether or not the V2 rule is unlearned, i.e. whether or not the participants incorrectly keep moving the auxiliary verbs and lexical verbs to C in English.

However, as pointed out in section 2.3, the auxiliary verb and the lexical verb behave differently from each other in English with respect to verb movement, as auxiliaries are allowed to move above the subject in *yes/no* interrogatives (I-to-C movement). Pollard (1996) argues that the reason why auxiliaries behave differently than lexical verbs is that finite auxiliary verbs are the only lexical entries that do not have the feature [-inverted]. Instead, auxiliaries are unspecified for this feature, which means that they are able either to invert, or not to invert (Pollard 1996: 291). Lexical verbs, on the other hand, never invert. Consequently, it might be possible that the Norwegian learners allow more verb movement with auxiliary verbs than with lexical verbs.

Prediction 4A is based on the argument that agreement is usually more challenging when the structural distance between the subject and the agreeing verb is increased. This is based on Ocampo's (2013) study of more and less complex sentences with long-distance agreement, which found that learners are less sensitive to subject-verb agreement errors when the structure of the sentence is complex (see section 2.4.3). Applying this to the current experiment, it suggests that sentences with long-distance agreement should be more problematic than sentences with local agreement because there are no intervening elements in the latter sentence structure.

Prediction 4B builds on findings from several studies, which indicate that long-distance agreement is more challenging when the subject is singular, and the noun that is closest to the verb is plural. For instance, Ocampo (2013) found that learners are less sensitive to agreement errors in sentences with prepositional clauses when the subject is singular and the intervening noun is plural (see section 2.4.3). As Ocampo's (2013) experiment was also a comprehension study, and the sentences with long-distance agreement contain prepositional clauses in the current experiment, the same behaviour is expected here too.

Prediction 4C is based on studies which have found that the most frequent error type with local agreement is to drop the 3rd person plural *–s,* in comparison to inserting a superfluous 3rd person singular *–s* (hypercorrection)*,* i.e. that there are more errors with sentences that have singular

subjects. However, both error types have been reported in the studies mentioned in section 2.4.1, and for that reason, it is expected that both error types are represented in the current study.

3.3 Method

In this study, data was gathered by the means of an acceptability judgement test. This is a quantitative research method, which makes it possible to systematically investigate the participants' judgements of sentences. More specifically, a quantitative method collects numeric data, and with statistical analyses, it is possible to look for common aspects, or trends, in the observations, in addition to causal or descriptive patterns (Johnson 2008: 4). This method is widely used in linguistic research, and especially in the generative tradition, because it is considered a window into the leaners' language system, i.e. it makes it possible to get information about the L2 learners mental grammar (Leow 1996: 126; Sprouse and Almeida 2011:2).

In the test, the participants rank sentences on a Likert scale from 1-4, in addition to the option "I don't know". 1 means unacceptable, and 4 means acceptable. In other words, the participants' task is to decide if the sentences are "good" or "bad" in English. This is a widely used method, as it is an easy task, and a natural way of assessing sentences (Dabrowska 2010:8). However, this method poses some issues. First, there is an issue with respect to what type of variable the Likert scale should be treated as. The scale is numeric, however, it is not quantitative, as the numbers represent levels or categories rather than measurable numbers. In other words, it is either nominal or ordinal. The latter type is most commonly used, as the numbers represent an ordered scale, i.e. 1 is lower than 2, 3 and 4, and 2 is lower than 3 and 4, and so forth (Levshina 2015:17; Johnson 2008:4). The scale is non-measurable, as mentioned above, as there is not necessarily the same interval between each level. In other words, it is impossible to know if the difference between 1 and 2 is the same as the difference between 2 and 3, or 3 and 4. This may pose problems to the analysis of the data, as parametric tests assume an interval between the levels (Dabrowska 2010:9). However, Dabrowska (2009:9) argues that this assumption has little effect on the results of these types of tests.

Nevertheless, in this experiment, the Likert scale may also be considered a nominal variable. The reason for this is that the numbers on the Likert scale may be treated as categories, rather than levels. More specifically, the scale consists of the levels 1-4, and the participants were informed that 1-2 should be given if they think the sentence is unacceptable, and 3-4 should be

given if they think the sentence is acceptable. For that reason, the Likert scale can be seen as a binary variable, i.e. a nominal variable, with the two categories *acceptable* and *unacceptable*, rather than an ordered scale. To some degree, this limits the issue of the possibly varying intervals between the levels, although it is still impossible to know why the participants have ranked a sentence as 2, rather than 3, for instance. For that reason, as Dabrowska (2010:9) points out, a common problem for both ordinal and nominal variables is that fine contrasts between the acceptability of the sentences may be missed, as the scale consists of fixed levels, or categories.

Furthermore, acceptability judgement tests are also referred to as grammaticality judgement tests, but only the former term is used in this text. The reason for this is that grammaticality and acceptability is not the same: a grammatical sentence refers to a sentence that follows the rules of grammar in a particular language, whereas an acceptable sentence refers to a sentence that a native speaker deems permissible, i.e. grammaticality is only one of many possible factors that makes a sentence acceptable (Chomsky 1965: 11; Dabrowska 2010: 4). Chomsky (1965) distinguishes these two terms by arguing that the notion of acceptability is "…a concept that belongs to performance, whereas grammaticalness belongs to the study of competence" (Chomsky 1965:11; Dabrowska 2010: 4). This statement suggests that sentences that do not follow the grammar rules of a particular language may be considered acceptable, and vice versa, due to performance factors. To exemplify, a sentence such as (50a), which follows the grammar rules of Standard English, may be considered unacceptable because of its semantics. Similarly, a sentence like (50b), which do not follow Standard English grammar rules, may be considered an acceptable answer to the question "What did you do last night?". Example (50a) is taken from Chomsky (1965:11) and (50b) is taken from Dabrowska (2010:4).

(50)

    a. The man who the boy who the students recognized pointed out is a friend of mine

    b. Watched some TV, then went to bed

Furthermore, as White (2003a: 17-18) argues, the data obtained from acceptability judgement tests can be described as intuitional data, although it cannot directly reflect the speakers' linguistic competence. That is, it cannot directly tell us whether the L2 learners' systematic

behaviour is the same as the grammatical system of English, as performance factors may affect their judgements (Paradis 2004: 8; Dabrowska 2010: 4). In other words, the underlying competence of L2 learners may be hidden by performance factors (White 2003b: 37). As Leow (1996: 126) argues, this is one of the main criticisms against acceptability judgement tests, as it has been argued to reflect performance rather than competence.

Similarly, another criticism is that acceptability judgement tests often lead researchers to draw false conclusions, as it is easy to make type I or type II errors (Sprouse and Almeida 2011:1). This is related to the notion of performance and competence, as the judgements in an acceptability judgement test may be affected by non-linguistic factors (Sprouse et al. 2012: 8). Type I errors, or false positives, refer to observations that show a difference between two conditions that does not exist, i.e. the results suggest that the null hypothesis should be rejected, although there is no true difference between the conditions. Instead, the reason why a difference is found may be due to a sampling error (Levshina 2015: 13). Making a type II error, or false negatives, means that, although there is a real difference, the null hypothesis is accepted, i.e. the results suggest that there are no differences between the experimental conditions (Levshina 2015: 13). The fact that false negatives and false positives may occur has been argued to be one of the weaknesses of acceptability judgement tests.

Nevertheless, Sprouse et al. (2012:22) argue that acceptability judgement tests are in fact reliable, as they have found no evidence for a reliability problem. More specifically, they found a minimum replication rate of 98 % when they tested 365 phenomena from Adger's (2003) textbook in syntax, and a minimum replication rate of 95 % when they tested 146 phenomena from a native language. This is supported by Leow (1996), who compared acceptability judgements of Spanish L2 to oral and written production tasks. To specify, Leow (1996:130) tested 30 college students who were in their first year of an undergraduate Spanish language study. They were given the tasks at two stages of their course: first, after three weeks, and secondly, after fourteen weeks. At these points, the participants had approximately six and 35 hours of formal language exposure, respectively. The results showed that there was a significant correlation between the scores of the production tasks and the acceptability judgement test (Leow 1996: 133). This correspondence indicates that acceptability judgement tests are reliable sources when it comes to reflecting learners' competence in the L2 (Leow 1996: 134: Mandell 1999: 76).

3.4 The pilot study

There are eight participants in the pilot study, who all have Norwegian as their only native language. Their age rages from 11 to 19, and their proficiency scores range between 21 and 35 out of 40 possible points (see appendix 1). The proficiency test is discussed in more detail section 3.5.3. Several age groups were tested in order to determine the age groups for the main experiment. That is, if some of the age groups had judged all of sentences correctly, or incorrectly, it would not have given any indications of the most difficult parts of L2 acquisition.

The results showed that the sentences were not too difficult or too easy for any of the participants, although some of them made very few incorrect judgements (see appendix 1). However, this was expected, as they were usually also the most advanced speakers. The participants were also asked if anything in the test was unclear, confusing or too difficult or easy. On the basis of this feedback, minor changes were made to the main experiment. Examples of this is that some of the sentences were adjusted with respect to vocabulary, and more information was added in the instructions. However, no major changes were done to the main test.

During the time in which the experiment was planned, one issue concerned whether the acceptability judgement test should be speeded or not, i.e. whether or not the participants should have a limited amount of time to judge the sentences. The advantage with a speeded test is that it gives information about more automatic and implicit knowledge, as it would force the participants to make a judgement based on first-pass parsing. That is, they would not have time to assess the sentence based on explicit and metalinguistic knowledge, which means that a speeded test would minimize post-processing effects (see Hopp 2007). Consequently, this would have given information about the speakers' on-line processing, as opposed to off-line processing, which is measured in acceptability judgement tests that are not speeded.

The initial idea was to give the participants one test that was speeded, and one test that was untimed. The purpose of this was to test both off-line and on-line processing. With respect to the speeded test, it is necessary to do millisecond timing accuracy, which is usually not possible in web-based survey software such as SurveyGizmo, SurveyMonkey or LimeSurvey, for instance. Instead, programs such as E-prime or OpenSesame are necessary, and, in order to ensure timing accuracy, the experiment must be administered locally. This means that the participants have to do the experiment one by one. This caused problems, as there is a large

number of participants in the main experiment (see section 3.5.3), as well as a limited amount of time to finish the project. For that reason, it was decided that an on-line speeded acceptability judgement test was not the best option for this experiment. Consequently, the pilot study is a web-based, untimed acceptability judgement test, created with the online survey software SurveyGizmo. The advantage with a web-based test is that all participants can complete the test simultaneously, which makes it possible to gather data from several participants at the same time.

3.5 The main experiment

In the following sections, I discuss the main experiment, which includes an untimed and web-based acceptability judgement test, a proficiency test, and a background questionnaire. In addition, I account for the test sentences in section 3.5.2, and the participants in section 3.5.3.

3.5.1   Procedure

The experiment took place during school hours, and the participants spent approximately 25-35 minutes to complete the whole test. Before the participants began the experiment, they were given instructions about what they were going to do in the test, as well as what the different numbers on the Likert scale mean. This information was given in Norwegian, both orally and in writing, as illustrated in appendix 2.

In the first part of the experiment, the participants completed the acceptability judgment test. Here, they saw five sentences at a time, and they were asked to rank them on a Likert scale from 1-4 (see section 3.3). Each sentence had to be ranked, and the participants were not able to return to any of the previous pages once they had moved to the next page. The reason for this is to avoid that the participants change their judgements, i.e. we want their judgements to be as spontaneous as possible in an off-line test. Example (51) illustrates how the sentences are presented to the participants (the whole test is provided in appendix 3).

(51)

16. Very soon will the band play their favourite song. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

17. The girl drinks a lot of water every day. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

18. Yesterday went the teacher to the shop. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

19. The house with yellow and white doors looks nice. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

20. Father my went to the shop every Monday. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

19%

Furthermore, the test is pseudo-randomized, which means that the five sentences that occur on each page have a permanent order, i.e. all participants receive the same sequence of sentences on each page. The pages, on the other hand, are randomized so that each participant gets a unique sequence of pages. The purpose of pseudo-randomizing is to make sure that a sentence pair never appears at the same page or that the sentences in a pair immediately follow one another from one page to the next. In addition, pseudo-randomizing makes it possible to control the types of constructions that occur on each page, i.e. that there is always one sentence that tests syntax, and one sentence that tests functional morphology, as well as no more than one filler on each page. This would not be possible to control for if all of the sentences were randomized. In addition, in order to avoid a recognizable pattern, there is not a set number of construction types on each page.

The second part on the experiment is a subset of a Standardized Oxford Proficiency test with 40 questions (see appendix 4). This test has been used in several other studies on language acquisition, such as Slabakova and Garcia Mayo (2015). As examples (52) and (53) illustrate, the test is a multiple-choice task, i.e. it consists of sentences with a blank spot, and three options below each of them. The participants' task is to choose the option that makes the sentence acceptable, and they get one point for each correct answer. It is only possible to pick one of the

options, and it is obligatory to fill out all of the blank spots. In the last 20 questions, the sentences form a continuous story.

(52)   *Multiple-choice task with individual sentences*

86. Water _____ at a temperature of 100° C. _*
    ○ is boiling
    ○ is to boil
    ○ boils

87. In some countries _____ very hot all the time. *
    ○ there is
    ○ is
    ○ it is

(53)   *Multiple-choice task with a continuous story*

106. The history of _____ is *
    ○ airplane
    ○ the airplane
    ○ an airplane

107. _____ short one. For many centuries men *
    ○ quite a
    ○ a quite
    ○ quite

The third, and final, part is the background questionnaire. Here, the students are asked to answer questions about their age and linguistic background (see appendix 5). In order to avoid any confusion about what a native language is, the participants were also asked which language they use with their mother, father, siblings and friends. These questions are all in Norwegian, in order to avoid any possible misunderstandings.

3.5.2   Sentences

There are 85 sentences in total in the study, of which 13 are fillers (see appendix 6). The actual test sentences are sentence pairs with one ungrammatical and one grammatical version of the same sentence. There are six different types of sentences: non-subject-initial declaratives with lexical verbs and auxiliary verbs, and subject-initial declaratives with 3rd person plural and singular subjects, as well as long and short distance agreement. The non-subject-initial declaratives test verb movement, i.e. syntax, and the subject-initial declaratives test subject-verb agreement, i.e. morphology. There are six sentence pairs of each type, which gives a total of 36 sentence pairs, and a total of 72 test sentences. These are discussed in turn shortly.

Three important notions apply for all sentences: first, they all consist of 10-12 syllables, so that all sentences have approximately the same length. Secondly, all sentences consist of words that are taken from a word frequency list of English, in order to ensure that most participants are familiar with the vocabulary (Word Frequency 2014). The reason for this is based on the argument described in section 3.3, namely, that the acceptability of a sentence may be determined on the basis of other factors than its grammaticality, which suggests that it is important to make the sentences as similar as possible in terms of length and vocabulary. This is in agreement with Dabrowska (2010: 5), who states that,

> … [R]esearchers must take care to either neutralize [extra-grammatical factors] (by balancing stimuli for length, lexical content, processing difficulty, plausibility, etc., whenever possible) or to control for them (by setting up control conditions which will allow them to assess the extent to which the confounding factors affect speakers' judgments.

Finally, all subjects are full DPs. The reason for this is based on the results of Slabakova and Gajdos's (2008) experimental study, discussed in section 2.4.4, as they show that learners struggle more with subject-verb agreement when there are DP subjects rather than pronominal subjects.

Regarding the non-subject-initial declarative main clauses, the sentences with lexical verbs are all in the past form in order to avoid judgments based on subject-verb agreement, and for the same reason, the auxiliary verbs are all modals. These are exemplified in (54) and (55), respectively (all of the sentences are listed in appendix 6).

(54)   *Lexical verbs:*

    a. *Yesterday went the teacher to the shop

    b. Yesterday the teacher went to the shop

(55)   *Auxiliary verbs:*

    a. *Every day should the students bring their books to school

    b. Every day the students should bring their books to school

The reason why subject-initial declarative main clauses are used to test subject-verb agreement in English is based on the mismatch between English and Norwegian with respect to overt agreement morphology (see chapter 2.3).

With respect to subject-verb agreement, there are, as mentioned previously in this section, four different sentence structures: long-distance agreement and local agreement, and both plural and singular subjects. These are exemplified in (56)-(58) (all sentences are illustrated in appendix 6).

(56)   *Singular subjects and local distance agreement*

    a. *The boys in the black car looks very scary

    b. The boys in the black car look very scary

(57)   *Plural subjects and local distance agreement*

    a. *The teachers gives their students a lot homework

    b. The teachers give their students a lot of homework

(58)   *Singular subjects and long distance agreement*

    a. *The teacher with black shoes walk to work every day

    b. The teacher with black shoes walks to work every day

(59)   *Plural subjects and long distance agreement*

    a. *The brown dog play with the yellow football

    b. The brown dog plays with the yellow football

These sentences all include prepositional phrases. In addition, there are no irregular verbs, i.e. they all receive the regular suffix –s when the subject is 3rd person singular, and a verb is only

used once in each of the four sentence types. Similarly, the subjects are all regular nouns, i.e. they receive the regular plural suffix –s. Furthermore, as the examples show, the noun that is in the prepositional phrase has the opposite number of the subject. If the two nouns had the same number, it would not give any information about whether or not the participants have judged the sentence on the basis of the subject noun or the intervening noun.

Finally, there are 13 fillers in the test. All of the fillers are ungrammatical, and thus, there are no sentence pairs. These sentences are non-target like possessives and definitive noun phrases, as exemplified in (60a) and (62a), respectively. In both of these structures, there is a mismatch between Norwegian and English, which I will briefly account for in the following. First, with possessives, the specifiers are pre-nominal in English, as exemplified in (60a).

(60)

      a.   *Father my went to the shop every Monday

      b.   My father went to the shop every Monday

In Norwegian, on the other hand, both pre- and post-nominal possessives are acceptable, as exemplified in (61). However, the word order in (61a) is most widely used by adult speakers of Norwegian L1 (see Andersen and Westergaard 2010).

(61)

      a. Faren   min   dro   til   butikken   hver   mandag
           Father.DEF my   went to   shop.DEF   every Monday
           'My father went to the shop every Monday'

      b. Min   far   dro   til   butikken   hver mandag
           My   father went to   shop   every Monday
           'My father went to the shop every Monday'

Similarly, definite noun phrases are marked with the pre-nominal article *the* in English, as exemplified in (62b), whereas it is marked with a suffixal article in Norwegian, as seen in (63) (see Andersen 2006).

(62)

    a.   *Thief the stole many paintings yesterday

    b.   The thief stole many paintings yesterday

(63)   **Tyven**   stjal   mange malerier   i går
        Thief.DEF   stole   many   paintings   yesterday
        'The thief stole many paintings yesterday'

3.5.3   Participants

Three school classes with a total of 69 students participated in the experiment. The age of the participants was decided on the basis of the pilot study, as described in section 3.4. The purpose was to get different groups of proficiency, and for that reason, 7th graders (11-12 years old) and upper secondary school students (15-18 years old) were chosen. The reason for this is that the difference between the lower secondary school students (14 years old) and the upper secondary school students was very little (see appendix 1). Consequently, the participants include 25 primary school students who attend the 7th grade (11-12 years old), 27 upper secondary school students who attend their first year (15-17 years old), and 17 upper secondary school students who attend their second year (18-19 years old). The schools were contacted by e-mail and telephone, and they were given an information letter about the project (see appendix 7). The schools themselves chose the classes that participated in the experiment, and the test was carried out during school hours.

In addition, the participants who were under the age of 15 were given a form of consent (see appendix 8), which was signed by the students' parents and handed back to the school. Students who are older than 15 years old are allowed to agree to participate themselves, as no sensitive information was collected in this experiment (NSD: Norsk Senter for Forskningsdata). This information was given orally to all students before the test started, as well as in written in the information section of the test. By clicking the "start" button, the students agreed to participate, and allowed their results to be used in research (see appendix 2).

As this study focus solely on speakers who have Norwegian as their L1 and English as their L2, multilinguals or speakers who did not have Norwegian as their native language were not included in the dataset. Consequently, 9 participants were excluded, which means that there is a total of 60 speakers in the dataset, of which 22 are primary school students, and 40 are upper secondary school students.

4   Results

The results were retrieved from SurveyGizmo and analysed in R. The main focus in this section is on whether there are statistically significant differences between the way in which agreement and verb movement are judged in the acceptability judgement test, and also, if there are statistically significant differences between the proficiency groups. This provides information about the difficulty of agreement and verb movement, as well as the way in which the participants' performance develops. As stated in section 3.2, the null hypothesis is that acquisition of verb movement and agreement is equally difficult, and the alternative hypothesis is that one construction is more difficult than the other. The p-value is set to 0.05, which means that any value smaller than this number is statistically significant. Furthermore, the dependent variable is the mean scores of the judgements in the acceptability judgement test, and the independent variables are the constructions discussed in section 3.5.2 and the proficiency scores (see section 4.1). The mean score of each sentence in the acceptability judgement test is illustrated in appendix 18.

In the following, I first discuss the participants' proficiency scores in comparison to their age and number of incorrect acceptability judgements in order to check if the proficiency scores reflect the participants general performance and exposure to English. Secondly, I discuss the results in the acceptability judgement test. The fillers are excluded from the analyses.
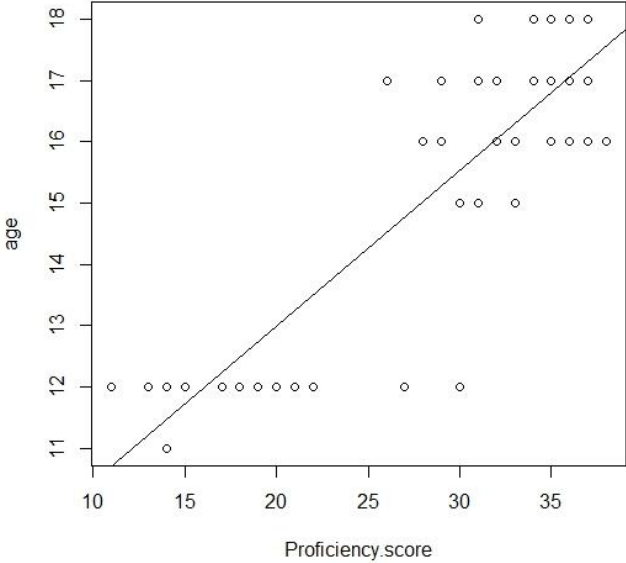
4.1 The proficiency test

As explained in section 3.5.1, the speakers' proficiency is measured in a multiple-choice task, which is a subset of the Standardized Oxford Proficiency test. There are 40 questions in this test, and the participants get one point for each correct answer. In other words, the highest possible proficiency score is 40. If the participants have a score that is lower than 10, they are considered beginners, if they have a score between 10 and 32, they are intermediate speakers, and if their score is higher than 32, they are considered advanced speakers. In addition, the intermediate level is split up into low intermediate, intermediate and high intermediate.

In this experiment, the participants' scores rage between 11 and 38 points (see appendix 9). This results in the following groups: low intermediate learners (between 11 and 17 points), intermediate learners (between 18 and 22 points), high intermediate learners (between 26 and 32 points) and advanced learners (between 33 and 38 points). The low intermediate group

consists of nine speakers (11- and 12- year-olds), the high intermediate group consists of 11 speakers (12- year-olds), the intermediate group has 20 speakers (12- to 18-year-olds), and finally, there are also 20 speakers in the advanced group (15- to 18-year-olds). The number of participants in each group was not decided beforehand. The following figure illustrates the correlation between age and proficiency:

*Figure 9: Correlation between proficiency scores and age*

Note: The x-axis shows the age of the participants (11-18) and the y-axis shows the proficiency scores (11-38).

The adjusted r^2-value for the correlation between age and proficiency scores is 0.7753 (see appendix 11A), which suggests that 78 % of the proficiency scores can be explained by the participants' age. In other words, there is a strong correlation between proficiency scores and age, and this correlation is positive, as the proficiency scores increase when the participants' age increases. Based on this, it is possible to argue that the proficiency test is reliable, as it reflects the participants' amount of exposure to English.

Furthermore, figure 10 shows the relationship between the participants' proficiency scores and the number of correct judgements in the acceptability judgement test. In order to illustrate this relationship, the number of correct judgements for each participant was summarized (see appendix 10). It is considered a correct judgement when a grammatical sentence is judged as acceptable, i.e. given a high score on the Likert scale (3-4), and similarly, it is considered correct when an ungrammatical sentence is deemed unacceptable, i.e. given a low score on the Likert scale (1-2) (see section 3.3).

Figure 10: Correlation between proficiency scores and correct judgements of agreement and verb movement, all participants
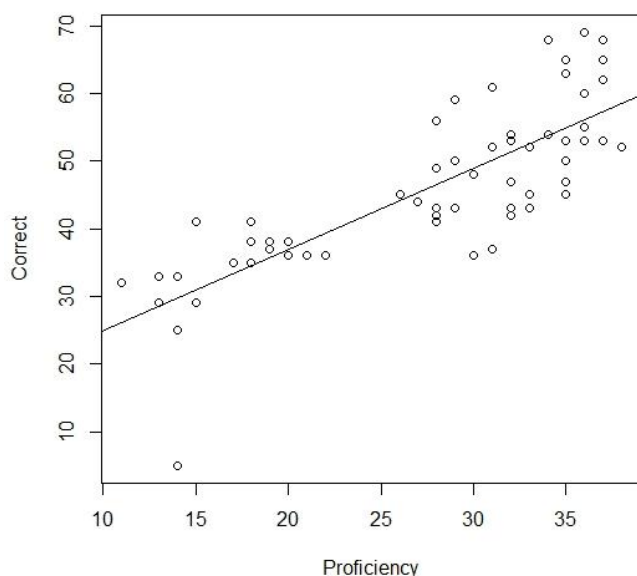
Note: the y-axis shows the number of correct judgements (5-68 out of 72 possible correct answers), and the x-axis shows the proficiency score (11-38 out of 40 possible points).

As the figure illustrates, the number of correct judgements increases as the proficiency scores increase, which suggests that there is a positive correlation. The fact that the adjusted r^2-value is 0.6428 (see appendix 11B) shows that the correlation is strong, as around 64 % of the number of correct judgements may be explained by the participants' proficiency scores. In other words, as the participants become more advanced speakers of English, they improve their performance in the acceptability judgement test.

4.2 The acceptability judgement test

To repeat some of the information in chapter 3, there are 36 sentence pairs in the experiment, and each pair consists of an ungrammatical and a grammatical version of the same sentence. The participants are asked to rank the sentences on a Likert scale from 1-4, where 1 means completely unacceptable, and 4 means completely acceptable. The scores on the Likert scale are treated as binary variable with the categories *unacceptable* (levels 1 and 2 on the Likert scale) and *acceptable* (levels 3 and 4 on the Likert scale). For that reason, the mean scores of the judgements range on a scale between 0 and 1 in the following text, where 0 represents the value 1-2, or *unacceptable,* and 1 represents the value 3-4, or *acceptable*. It follows that a mean score close to 0 for ungrammatical sentences indicates few errors, and similarly, that a mean score close to 1 for grammatical sentences indicates few errors with grammatical sentences.

Furthermore, t-tests and regression are used to analyse the data. These tests require normal distribution. Figure 11 shows that this requirement is fulfilled with respect to the judgements of agreement and verb movement in the acceptability judgement test, as the p-value is higher than 0.05 (see appendix 12A). That is, the null hypothesis here is that the data is normally distributed, and consequently, as the p-value is higher than 0.05, this hypothesis cannot be rejected (Levshina 2015: 12).

Figure 11: Density plot of the participants' number of correct judgements
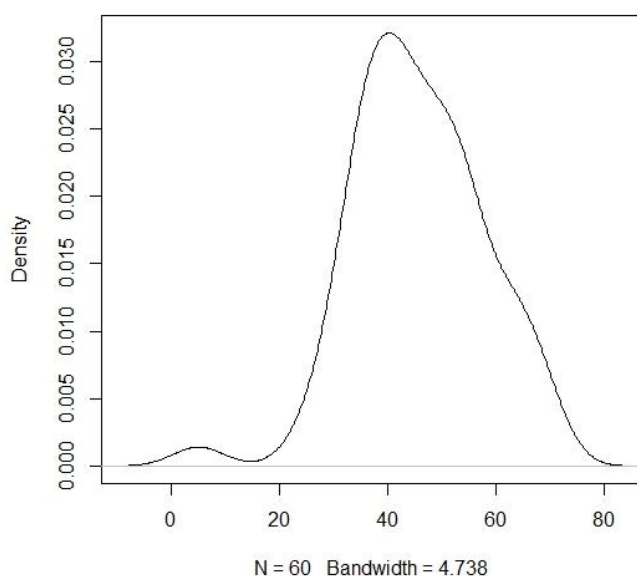
Note: The x-axis shows the number of correct judgements, which varies from 5 to 68 out of 72 possible correct answers.

In the following sections, I first discuss the participants' judgements of agreement and verb movement (sections 4.2.1 and 4.2.2), and then I compare the judgements of these two constructions in section 4.2.3. In section 4.2.4, the difference between the proficiency groups' judgements are discussed, including a closer look at the correlation between proficiency scores and judgements of verb movement and agreement. Sections 4.2.5 and 4.2.6 presents the judgements of the morphological and syntactic conditions, respectively, and finally, in section 4.2.7 I look at the results when long-distance agreement is excluded from the dataset.

4.2.1 Agreement

Figure 12 shows the mean scores for judgements of ungrammatical and grammatical agreement when all 60 participants are included. The mean score for ungrammatical sentences is significantly lower than the mean score for grammatical sentences, i.e. ungrammatical sentences are more often ranked as unacceptable (see appendix 16B). However, the mean scores are both closer to 1, than they are to 0, which means that both ungrammatical and grammatical sentences are often accepted.

Figure 12: Mean score for the judgements of grammatical and ungrammatical agreement, all participants

	All participants
Grammatical	0.8680556
Ungrammatical	0.6777778
P-value	<0.05

The same result is found in the high intermediate and advanced groups, but not in the low intermediate and intermediate groups. In other words, in the two highest proficiency groups, the mean score for ungrammatical sentences with agreement is significantly lower than the mean score for grammatical sentences with agreement, and in the two lowest proficiency groups, there is no statistically significant difference between the way in which ungrammatical and grammatical sentences with agreement are judged (see appendix 16G-16J). This is illustrated in figure 13.

Figure 13: Mean scores for judgements of grammatical and ungrammatical agreement, proficiency groups

	Low intermediate	Intermediate	High intermediate	Advanced
Grammatical	0.7546296	0.8923611	0.8837719	0.8895833
Ungrammatical	0.8055556	0.8888889	0.7061404	0.4666667
P-value	>0.05	>0.05	<0.05	<0.05

4.2.2   Verb movement

Similarly, with verb movement, figure 14 shows that ungrammatical sentences have a significantly lower mean score in comparison to grammatical sentences when the judgements of all participants are included (see appendix 16A). In this case, the mean score for grammatical sentences is close to 1, whereas the mean score for ungrammatical sentences is close to 0, i.e. grammatical sentences are most often accepted, and ungrammatical sentences are usually rejected.

Figure 14: Mean scores for judgements of grammatical and ungrammatical verb movement, all participants

	All participants
Grammatical	0.8541667
Ungrammatical	0.3305556
P-value	<0.05

This result is found in the four highest proficiency groups as well, i.e. in all of the proficiency groups except for the low intermediate one, the mean score for ungrammatical sentences is significantly lower than the mean score for grammatical sentences (see appendix 16C-16F). Nevertheless, the mean score is still high for both ungrammatical and grammatical sentences until the high intermediate stage, i.e. both ungrammatical and grammatical sentences are most often accepted until this level of proficiency. This is illustrated in figure 15.
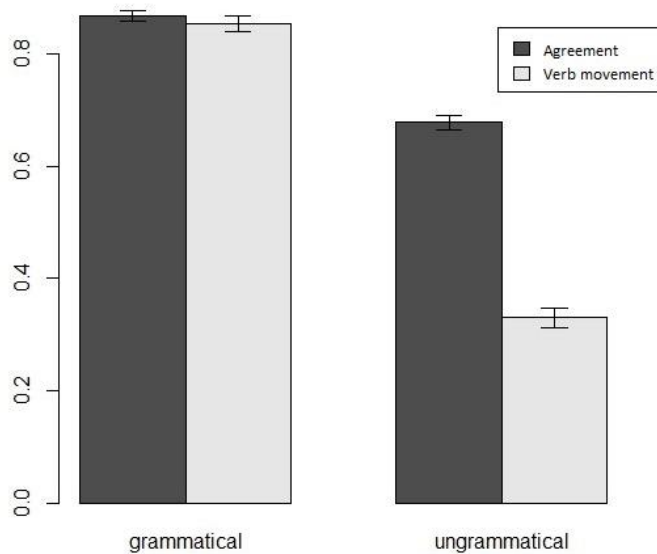
Figure 15: Mean scores for judgements of grammatical and ungrammatical verb movement, proficiency groups

	Low intermediate	Intermediate	High intermediate	Advanced
Grammatical	0.7222222	0.8680556	0.8640351	0.8958333
Ungrammatical	0.5833333	0.6875000	0.2280702	0.1000000
P-value	>0.05	<0.05	<0.05	<0.05

4.2.3 Comparison of agreement and verb movement

Comparing the judgements of verb movement and agreement in section 4.2.1 and 4.2.2, respectively, there is no statistically significant difference between the mean scores of grammatical sentences, and these are both high when the judgements of all participants are considered. This is illustrated in figures 16 and 17. In other words, there are few incorrect judgements of both grammatical verb movement and grammatical agreement. The judgements of ungrammatical sentences, on the other hand, differ more, as there is a statistically significant difference between them. To specify, the mean score for ungrammatical verb movement is significantly lower than the mean score for ungrammatical agreement (see appendix 13A and 14A). In other words, there are more incorrect judgements of ungrammatical agreement in comparison to ungrammatical verb movement.

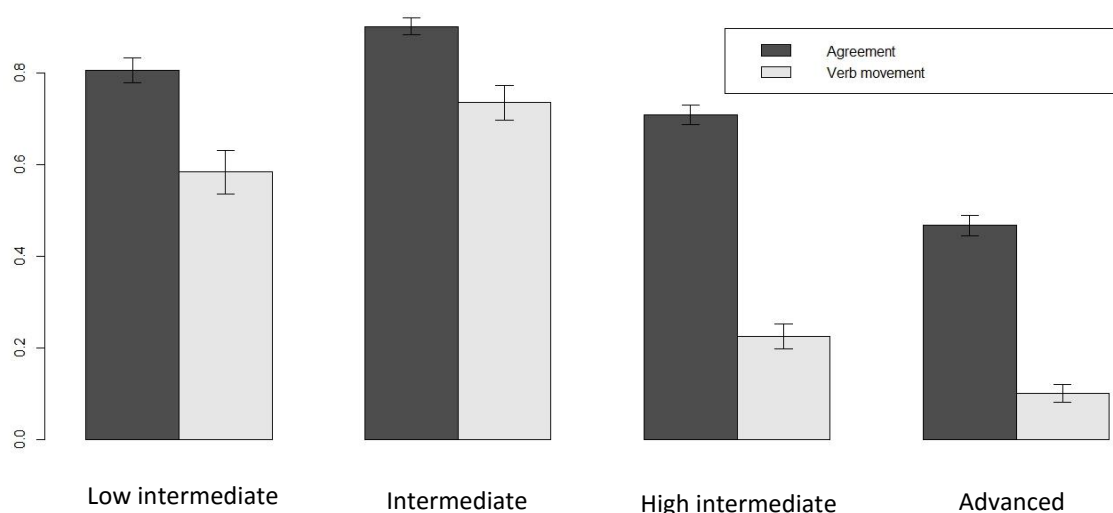*Figure 16: Mean scores for judgements of agreement and verb movement, all participants*

Note: The y-axis illustrates the mean score (0-1) of the grammatical and ungrammatical sentences.

Figure 17: Mean scores for judgements of agreement and verb movement, all participants

	Grammatical	Ungrammatical
Agreement	0.8680556	0.6777778
Verb movement	0.8541667	0.3305556
P-value	>0.05	<0.05

When the participants are divided in proficiency groups, the same result is found in each group, i.e. first, that the mean scores for verb movement and agreement differ significantly from one another when the sentences are ungrammatical, and more specifically, the mean score for verb movement is significantly lower than the mean score for subject-verb agreement (see appendix 14N-14P). This is illustrated in figure 18 and 19. Secondly, as figure 20 and 21 show, there are no statistically significant differences between the mean scores for verb movement and agreement when the sentences are grammatical.

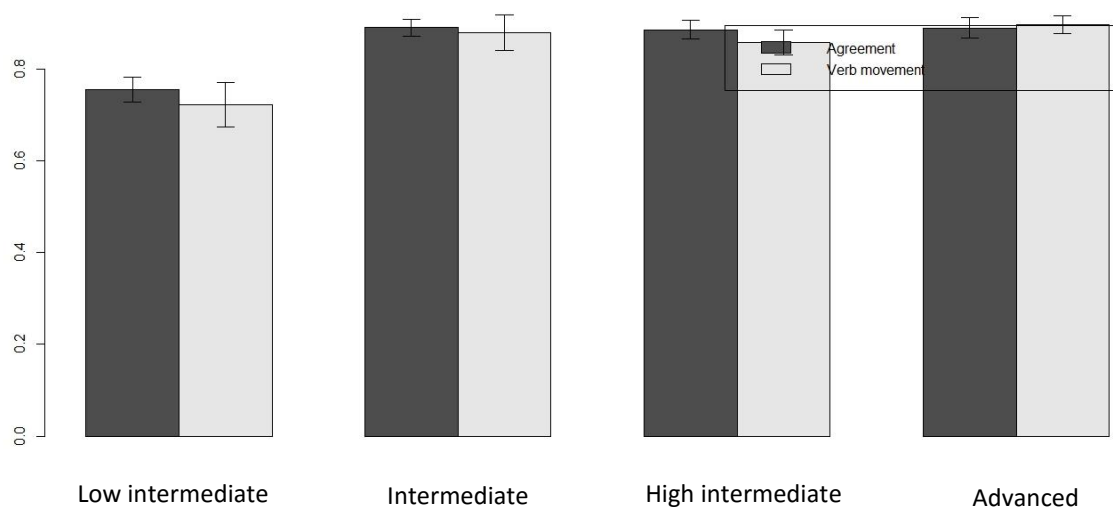*Figure 18: Mean score for judgements of ungrammatical sentences, proficiency groups*

Note: The y-axis illustrates the mean score (0-1) of the grammatical and ungrammatical sentences.

Figure 19: Mean scores for judgements of ungrammatical sentences, proficiency groups

	Low intermediate	Intermediate	High intermediate	Advanced
Agreement	0.8055556	0.8888889	0.7061404	0.4666667
Verb movement	0.5833333	0.6875000	0.2280702	0.1000000
P-value	<0.05	<0.05	<0.05	<0.05

Figure 20: Mean scores for judgements of grammatical sentences, proficiency groups

Note: The y-axis illustrates the mean score (0-1) of the grammatical and ungrammatical sentences.

Figure 21: Mean scores for judgements of grammatical sentences, proficiency groups

	Low intermediate	Intermediate	High intermediate	Advanced
Agreement	0.7546296	0.8923611	0.8837719	0.8895833
Verb movement	0.7222222	0.8680556	0.8640351	0.8958333
P-value	>0.05	>0.05	>0.05	<0.05

4.2.4 Differences between proficiency groups

In the previous section, I discussed the difference between verb movement and agreement within each of the proficiency groups. In this section, I focus on the differences between the proficiency groups with respect to how they judge these two constructions in order to illustrate the way in which performance on agreement and verb movement develops. The mean scores that are addressed are those illustrated in figures 18-21 above.

Regarding grammatical agreement, there is a statistically significant difference between the mean score in low intermediate group in comparison to the mean scores in each of the other proficiency groups, as illustrated in figure 21, and to specify, the mean score is significantly

lower in the low intermediate group than in the other proficiency groups (see appendix 13K). The mean scores in the intermediate, high intermediate and advanced group do not differ significantly from each other. Similarly, with grammatical verb movement, the mean score in the low intermediate group is significantly lower than the mean scores in the other groups, and there are no statistically significant differences between the latter three groups (see appendix 13L).

With respect to ungrammatical agreement, the mean score in the low intermediate group is significantly higher than the mean score in the advanced group, but does not differ significantly from the mean scores in the high intermediate and intermediate groups (see appendix 13M). With ungrammatical verb movement, the low intermediates differ significantly from the high intermediate and advanced group, but not from the intermediate group (see appendix 13N). The high intermediate and advanced group do not differ significantly from one another neither with agreement, nor with verb movement. The mean scores are illustrated in figure 19.

In other words, the results show that there are more significant differences between the proficiency groups' mean scores with ungrammatical verb movement. As I demonstrate in the following paragraphs, this behaviour is also found when the correlation between proficiency scores and the participants' judgements of agreement and verb movement are compared. Figure 22 and figure 23 illustrate the relationship between proficiency scores and judgements of grammatical sentences, and figure 24 and 25 show the same correlation for ungrammatical sentences. I first discuss the grammatical sentences, and then the ungrammatical sentences.

In figures 22 and 23, most mean scores are closer to 0 than to 1, which illustrates that there are generally few errors with grammatical sentences. The adjusted r^2-value supports this, as it is 0.1159 for the correlation in figure 22 and 0.06865 for the correlation in figure 23 (see appendix 11C and 11D). This means that the proficiency scores may explain around 11 % of the judgements of grammatical verb movement and around 6 % of the judgements of grammatical verb movement, i.e. the correlation is weak.

Figure 22: Mean scores of judgements (0-1) and proficiency scores, grammatical verb movement
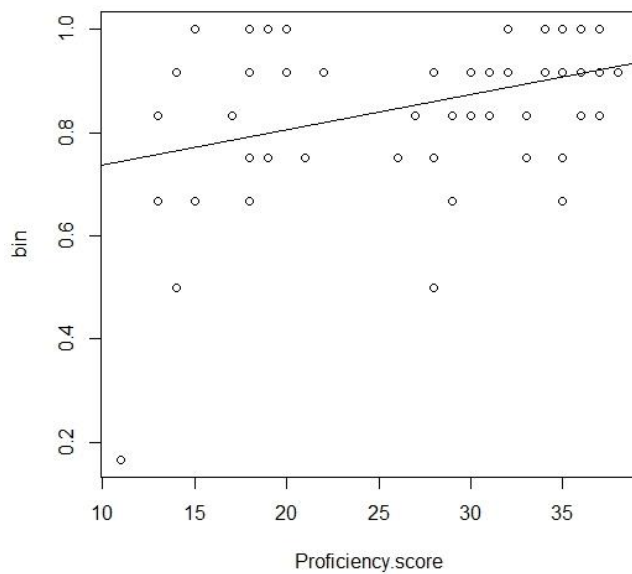
Note: The y-axis shows the mean scores of judgements (0-1), and the x-axis show the proficiency scores (11-38).

Figure 23: Mean scores of judgements and proficiency scores, grammatical agreement
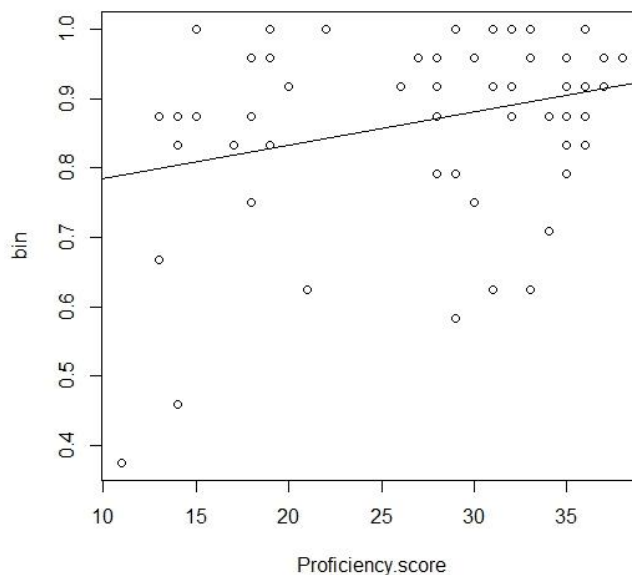
Note: The y-axis shows the mean scores of judgements (0-1), and the x-axis show the proficiency scores (11-38).

Figure 24 and figure 25 illustrate the relationship between proficiency scores and judgements of ungrammatical verb movement and agreement, respectively. With respect to the judgements of verb movement, figure 24 shows that the participants with lower proficiency scores differ from the participants with higher proficiency scores. The adjusted r^2-value is 0.4688 (see

71

appendix 11E), which suggests that the negative correlation is strong. In other words, the proficiency score can explain around 47 % of the correct judgements of ungrammatical verb movement. With ungrammatical agreement, figure 25 illustrates that the participants with higher proficiency scores make more errors than they do with ungrammatical verb movement. The participants with lower proficiency scores also make more errors with agreement, but the difference from verb movement is not as dramatic as for the participants with higher proficiency scores. In other words, the negative correlation between proficiency scores and judgements is stronger in figure 25 than in figure 24. This is supported by the adjusted r^2-value in the latter figure, which is 0.2227 (see appendix 11F), i.e. around 22% of the correct judgements can be explained by the proficiency scores.

Figure 24: Mean scores of judgements on the Likert scale and proficiency scores, ungrammatical verb movement

Note: The y-axis shows the mean scores of judgements (0-1), and the x-axis show the proficiency scores (11-38).

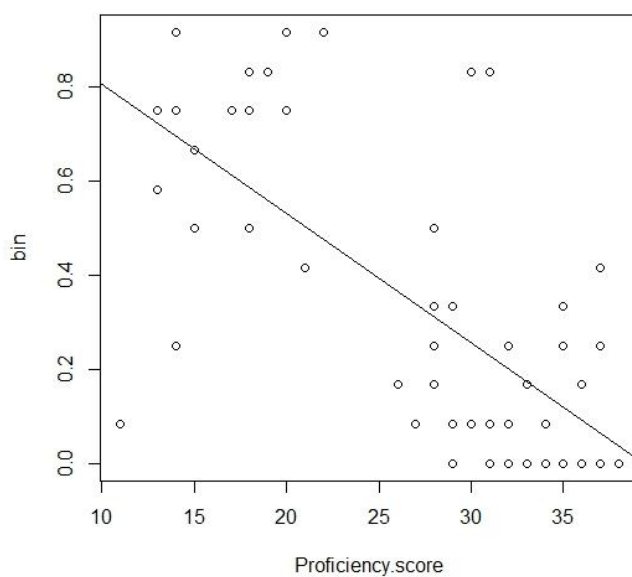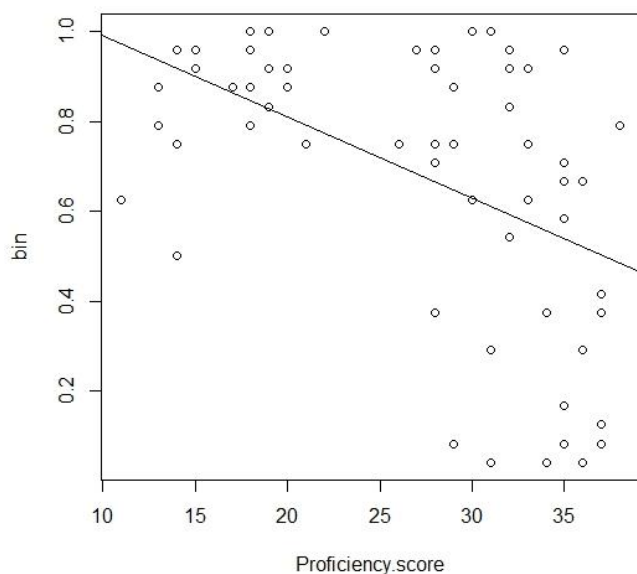*Figure 25: Mean scores of judgements and proficiency scores, ungrammatical agreement*

Note: The y-axis shows the mean scores of judgements (0-1), and the x-axis show the proficiency scores (11-38).

In other words, all participants make few errors with grammatical agreement and verb movement, which results in a weak correlation between proficiency scores and acceptability judgements for both agreement and verb movement. With ungrammatical sentences, on the other hand, the correlation is stronger for both conditions. However, the correlation is still relatively weak for ungrammatical agreement, whereas it is strong for ungrammatical verb movement.

4.2.5   Morphological conditions

As previously explained, six different sentence structures are used to test subject-verb agreement: local and long-distance agreement with plural and singular subjects. These are referred to as the morphological conditions. In this section, I discuss and compare the judgements of these structures.

Figure 26 illustrates the mean score of the judgements of each morphological condition i.e. local and long-distance agreement with plural and singular subjects when the judgements of all participants are considered:

Figure 26: Mean scores of judgements (0-1) for morphological conditions, all participants

	Grammatical	Ungrammatical
Pl-long	0.7305556	0.7027778
Pl-local	0.9000000	0.6861111
Sg-long	0.9111111	0.6750000
Sg-local	0.9305556	0.6472222

Note: The abbreviation *pl* and *sg* refers to plural and singular, respectively, *local* refers to local agreement, and *long* stands for long-distance agreement

With respect to the grammatical sentences, there is a statistically significant difference between the mean scores of long-distance agreement and plural subjects and the rest of the conditions (see appendix 13B). In other words, the mean score for judgements of sentences like (64a) are significantly lower than the mean scores for judgements of sentences such as (64b) to (64d). There are no significant differences between the mean scores of the rest of the conditions, as they all range around 0.90.

(64)

    a. The cats with long white fur drink milk every day

    b. The house that is on top of the hill looks nice

    c. The cats play with the yellow and green ball

    d. The teacher eats fish for dinner every Friday

With ungrammatical sentences, there are no statistically significant differences between the mean scores of the morphological conditions. The sentences are exemplified in (65):

(65)

    a.   *The cats with long white fur drinks milk every day

    b.   *The house that is on top of the hill look nice

    c.   *The cats plays with the yellow and green ball

    d.   *The teacher eat fish for dinner every Friday

With respect to the proficiency groups, the following addresses each group's judgements of the morphological conditions in turn. First, the low intermediates' scores are illustrated in figure 27. Here, there are no statistically significant differences between the way in which the structures are judged, neither with grammatical sentences, nor with ungrammatical sentences.

Figure 27: Mean scores for judgements of morphological conditions, low intermediates

	Grammatical	Ungrammatical
Pl-long	0.7037037	0.8148148
Pl-local	0.7777778	0.7407407
Sg-long	0.7407407	0.8333333
Sg-local	0.7962963	0.8333333

The intermediate group's mean scores are illustrated in figure 28. The same result as for the low intermediate group is found here, i.e. that there are no statistically significant differences between the means scores for judgements of morphological conditions.

Figure 28: Mean scores for judgements of morphological conditions, intermediates

	Grammatical	Ungrammatical
Pl-long	0.8333333	0.8787879
Pl-local	0.8787879	0.8939394
Sg-long	0.9090909	0.9090909
Sg-local	0.9305556	0.9305556

The mean scores in the high intermediate group, on the other hand, differs more from one another, as shown in figure 29. However, this only applies for the grammatical sentences.

Figure 29: Mean scores for judgements of morphological conditions, high intermediates

	Grammatical	Ungrammatical
Pl-long	0.7583333	0.7000000
Pl-local	0.9000000	0.7583333
Sg-long	0.9666667	0.6500000
Sg-local	0.9166667	0.7250000

More specifically, the high intermediates' mean score of long-distance agreement with plural subjects is significantly lower than the mean scores for the rest of the conditions (see appendix 14C and 14D). The other three conditions do not differ significantly from one another. To exemplify, this suggests that sentences like (66a) are rejected significantly more often than sentences such as (66b)-(66d).

(66)

    a. The cats with long white fur drink milk every day

    b. The house that is on top of the hill looks nice

    c. The cats play with the yellow and green ball

    d. The teacher eats fish for dinner every Friday

Furthermore, figure 30 illustrates the advanced learners' mean scores of judgements. In this case, there are statistically significant differences with both the grammatical and ungrammatical sentences.

Figure 30: Mean scores for judgements of morphological conditions, advanced

	Grammatical	Ungrammatical
Pl-long	0.6583333	0.5583333
Pl-local	0.9666667	0.4750000
Sg-long	0.9333333	0.5000000
Sg-local	1.0000000	0.3333333

With respect to grammatical sentences, the differences between the mean scores are similar to the ones in the high intermediate group (see figure 29 above). That is, the mean score for long-distance agreement are significantly lower than the morphological conditions' mean scores (see appendix 14F and 14I). In other words, sentences like (67a) have a significantly lower mean score in comparison to the rest of the morphological conditions in the advanced group too.

(67)

    a.   The cats with long white fur drink milk every day

    b.   The house that is on top of the hill looks nice

    c.   The cats play with the yellow and green ball

    d.   The teacher eats fish for dinner every Friday

In addition to this, there are more significant differences between the judgements in this group. First, the mean score for long-distance agreement with singular subjects is significantly different from the mean score for local agreement with singular subjects, i.e. in this group, the mean score for long-distance agreement, in comparison to local agreement, is not only significantly lower when the subject is plural, but also, when it is singular (see appendix 14H). Local and long-distance agreement with singular subjects are illustrated above in (67d) and (67b), respectively.

Secondly, there is a statistically significant difference between the mean scores for local agreement with plural subjects and local agreement with singular subjects (see appendix 14G). That is, the mean score for the former is significantly lower than the mean score for the latter. In other words, the participants more often reject sentences like (67c), in comparison to (67d).

With respect to ungrammatical sentences, there is a statistically significant difference between the mean scores for local agreement with singular subjects, and the mean scores for local agreement with plural subjects (see appendix 14J). That is, the mean score for sentences like (68d) is significantly lower than the mean score for sentences like (68c), i.e. the participants more often correctly reject sentences such as the former.

(68)

    a.   *The cats with long white fur drinks milk every day

    b.   *The house that is on top of the hill look nice

    c.   *The cats plays with the yellow and green ball

    d.   *The teacher eat fish for dinner every Friday

In addition, there is a significant difference between the mean scores for ungrammatical long-distance agreement with singular subjects, and ungrammatical local agreement with singular subjects, i.e. sentences such as (68a) and (68b), respectively (see appendix 14K). In this case, the mean score for long distance agreement is significantly higher than the mean score for local agreement.

4.2.6 Syntactic conditions

Two structures are used to test verb movement: non-subject-initial declarative clauses with lexical verbs and auxiliary verbs (see section 3.5.2). These are referred to as the syntactic conditions, and in this section, I discuss the way in which these two structures are judged. Figure 31 illustrates the judgements of all of the participants, and figures 32 to 35 show the proficiency groups' judgements.

Figure 31: Mean scores of judgements of syntactic condition, all participants

	Grammatical	Ungrammatical
Auxiliary verbs	0.8611111	0.3944444
Lexical verbs	0.8472222	0.2666667

As seen in figure 31, the grammatical sentences have a high mean score for both auxiliary verbs and lexical verbs, and there is thus no statistically significant difference between the judgements of the two conditions, i.e. sentences such as (69a) and (69b).

(69)

      a.   Yesterday the teacher went to the shop

      b.   In December the kids will celebrate Christmas

With ungrammatical sentences, on the other hand, there is a statistically significant difference, and more specifically, the mean score for sentences with lexical verbs is significantly lower than the mean score for sentences with auxiliary verbs (see 13E and 14B). This suggests that sentences like (70a) are more often incorrectly accepted than sentences like (70b).

(70)

      a.   *In December will the kids celebrate Christmas

      b.   *Yesterday went the teacher to the shop

As mentioned above, figure 32 to figure 35 show the proficiency groups' judgements of the syntactic conditions, and I now turn to discuss their judgements in turn. First, in the low intermediate group, there are no statistically significant differences, as both ungrammatical and grammatical sentences with auxiliary and lexical verbs are ranked frequently accepted.

Figure 32: Mean scores for judgements of syntactic conditions, low intermediates

	Grammatical	Ungrammatical
Auxiliary verbs	0.7037037	0.5925926
Lexical verbs	0.7407407	0.5740741

In the intermediate group, on the other hand, there is a statistically significant difference between the mean scores for lexical verbs and auxiliary verbs when the sentences are ungrammatical (see appendix 14M). As figure 33 illustrates, the mean score for ungrammatical lexical verbs is lower than the mean score for ungrammatical auxiliary verbs, i.e. there are

significantly more errors with ungrammatical auxiliaries. The sentences are exemplified in (72) above.

Figure 33: Mean scores for judgements of syntactic conditions, intermediates

	Grammatical	Ungrammatical
Auxiliary verbs	0.8939394	0.8030303
Lexical verbs	0.8636364	0.6666667

Similarly, there a statistically significant difference between the two syntactic conditions when the sentences are ungrammatical in the high intermediate group (see appendix 14E)., but not when the sentences are grammatical. This is shown in figure 34, in which the mean score for sentences with lexical verbs is significantly lower than the mean score of sentences with auxiliary verbs (see the sentences in example (70)).

Figure 34: Mean scores for judgements of syntactic conditions, high intermediates

	Grammatical	Ungrammatical
Auxiliary verbs	0.8500000	0.3250000
Lexical verbs	0.8666667	0.1250000

The advanced group's mean scores are illustrated in figure 35. Here too, there is a statistically significant difference between the mean scores when the sentences are ungrammatical, and again, the mean score for lexical verbs is significantly lower than the mean score for auxiliary verbs (see appendix 14L). As in the intermediate and high intermediate group, this is not found when the sentences are grammatical (sentences with lexical verbs and auxiliaries are exemplified in (70)).

Figure 35: Mean scores for judgements of syntactic conditions, advanced

	Grammatical	Ungrammatical
Auxiliary verbs	0.9250000	0.1500000
Lexical verbs	0.8666667	0.0500000

4.2.7   Results without long-distance agreement

The following sections address the results discussed in the above sections when long-distance agreement is excluded from the dataset. The reason why this is done is related to the argument that variability in long-distance agreement is caused by processing limitations, i.e. non-linguistics factors, rather than limited knowledge of subject-verb agreement (see section 2.4.3). For that reason, it is necessary to check if there is still a significant difference between the judgements of verb movement and agreement. As the following results show, there are few differences between the results described in the above sections and those that are found when judgements of long-distance agreement is excluded from the dataset. As this only affects the mean scores for agreement, the judgements of verb movement are only repeated when the two constructions are compared.

First, with respect to the judgements of agreement and verb movement, figure 38 shows that the mean scores are slightly higher for grammatical sentences and slightly lower for ungrammatical sentences when long-distance agreement is not included in the dataset. In other words, the difference between judgements of ungrammatical and grammatical agreement is still significant, as section 4.2.1 illustrates (see appendix 17A).

Furthermore, the t-test shows that there is a significant difference between agreement and verb movement in both grammatical and ungrammatical sentences (see appendix 15A and 15B, respectively). This is also illustrated in figure 36. The mean score for agreement is significantly higher than the mean score for verb movement, both when sentences are grammatical and ungrammatical. In other words, when long-distance agreement is excluded from the dataset, there are significantly more correct judgements with agreement in grammatical sentences, and significantly more correct judgements with verb movement in ungrammatical sentences. This difference in grammatical sentences was not found in section 4.2.1.

Figure 36: Judgements of grammatical and ungrammatical sentences, all participants

	Grammatical	Ungrammatical	P-value
Agreement (all conditions)	0.8680556	0.6777778	<0.05
Agreement (without long-distance agreement)	0.9152778	0.6666667	<0.05
P-value	<0.05	<0.05	

Regarding the mean scores of agreement in each proficiency group, figure 37 illustrates the participants' judgements when long-distance agreement is excluded, and figure 38 repeats the mean scores presented in section 4.2.1, i.e. when long-distance agreement is included in the dataset. A comparison of these two tables shows that removing long-distance agreement from the dataset does not affect the fact that the judgements of ungrammatical and grammatical agreement only differ significantly from each other in the high intermediate and advanced groups, and that the participants usually accept all sentences until the advanced stage (see appendix 17B-17E).

Figure 37: Judgements of grammatical and ungrammatical agreement, proficiency groups

Agreement (without long-distance agreement)	Low intermediate	Intermediate	High intermediate	Advanced
Ungrammatical	0.787037	0.9097222	0.7324561	0.4041667
Grammatical	0.787037	0.9097222	0.9078947	0.9833333
P-value*	>0.05	>0.05	<0.05	<0.05

Figure 38: Judgements of grammatical and ungrammatical agreement, proficiency groups

Agreement (all conditions)	Low intermediate	Intermediate	High intermediate	Advanced
Ungrammatical	0.8055556	0.8888889	0.7061404	0.4666667
Grammatical	0.7546296	0.8923611	0.8837719	0.8895833
P-value	>0.05	>0.05	<0.05	<0.05

When judgements of agreement and verb movement are compared, figure 39 shows that the two constructions do not differ significantly from each other when the sentences are grammatical. This result was also found in section 4.2.1. The first row repeats the mean scores presented in section 4.2.1, i.e. when long-distance agreement is included in the dataset. Similarly, the same result as in 4.2.1 is found with ungrammatical sentences, as illustrated in figure 40, i.e. the mean scores for verb movement and agreement differ significantly from each other in all of the proficiency groups (see appendix 15C-15F).

Figure 39: Judgements of grammatical sentences, proficiency groups

	Low intermediate	Intermediate	High intermediate	Advanced
Agreement (all morphological conditions	0.7546296	0.8923611	0.8837719	0.8895833
Agreement (without long distance agreement)	0.7870370	0.9097222	0.9078947	0.9833333
Verb movement	0.7222222	0.8680556	0.8640351	0.8958333
p-value*	>0.05	>0.05	>0.05	<0.05

*The p-value shows the difference between agreement without long distance agreement and verb movement

Figure 40: Judgements of ungrammatical sentences

	Low intermediate	Intermediate	High intermediate	Advanced
Agreement (all morphological conditions	0.8055556	0.8888889	0.7061404	0.4666667
Agreement (without long distance agreement)	0.7870370	0.9097222	0.7324561	0.4041667
Verb movement	0.5833333	0.6875000	0.2280702	0.1000000
P-value*	<0.05	<0.05	<0.05	<0.05

*The p-value shows the difference between agreement without long distance agreement and verb movement

With respect to the correlation between agreement and proficiency scores, as discussed in section 4.2.4, the correlation is still weak with both grammatical and ungrammatical agreement.

To elaborate, the adjusted r^2-value is 0.2062 with ungrammatical agreement, as illustrated in figure 41 (see appendix 11G), which is similar to the value found in section 4.2.4, i.e. for the correlation between ungrammatical agreement and proficiency scores when long-distance agreement is included in the dataset (adjusted r^2 = 0.2227). With grammatical agreement, as seen in figure 42, the adjusted r^2-value is 0.2524 when the scores for long-distance agreement excluded (see appendix 11H). This is also a relatively weak correlation, although it is considerably higher than when the mean scores for long-distance agreement are considered, which is 0.06865 (see section 4.2.4). In other words, the correlation between proficiency scores and judgements of agreement is still weak with both grammatical and ungrammatical sentences.

Figure 41: Correlation between proficiency scores and ungrammatical agreement (without long-distance agreement)
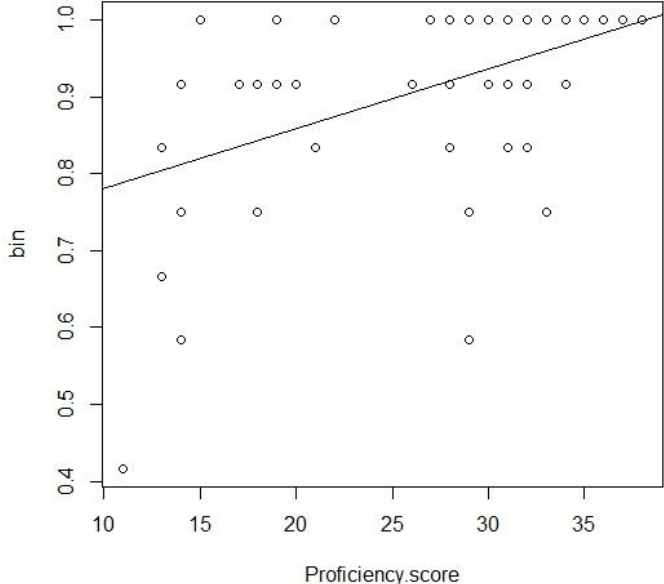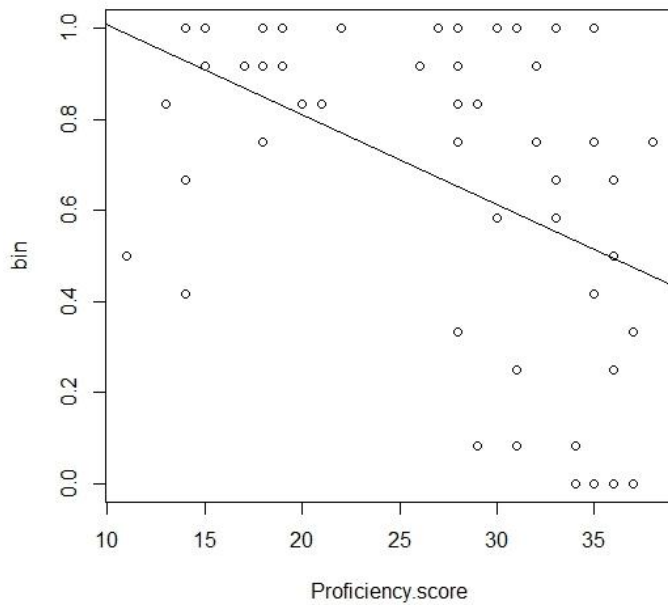
Figure 42: Correlation between proficiency scores and grammatical agreement (without long-distance agreement)

Finally, when considering the differences between the proficiency groups with respect to how they judge verb movement and agreement, there are no differences between the results found in section 4.2.3. That is, with respect to ungrammatical agreement, there is still a significant difference between the low intermediate speakers and advanced speakers, but no significant difference between the low intermediate, the intermediate and the high intermediate groups (see appendix 13Q). With grammatical agreement, there is a statistically significant difference between the low intermediate group and the rest of the groups (see appendix 13R).

5   Discussion

In this chapter, I discuss the results in chapter 4, with focus on the research questions posed in chapter 3. The questions, as well as the predictions, are repeated here for convenience, and I discuss them in turn.

> RQ1: Is functional morphology more difficult than narrow syntax in L2 acquisition?
>
> RQ2: Is functional morphology a more persistent problem than narrow syntax?
>
> RQ3: Which of the syntactic and morphological conditions are more difficult?

Prediction 1: Subject-verb agreement is more difficult than verb movement

Prediction 2: Subject-verb agreement is a more persistent problem than verb movement

Prediction 3: Syntactic conditions

> Word order in sentences with auxiliary verbs is more difficult than in sentences with lexical verbs.

Prediction 4: Morphological conditions

> A. Long-distance agreement is more difficult than local agreement
>
> B. Long-distance agreement is more difficult when the subject is 3rd person singular, in comparison to when it is 3rd person plural
>
> C. Local agreement is more difficult when the subject is 3rd person singular, than when it is 3rd person plural

As explained in sections 3.1 and 3.2, the predictions are based on previous findings in research on L2 acquisition of functional morphology and narrow syntax and the predictions that the Bottleneck Hypothesis puts forward. To briefly repeat some of the main points in chapters 2 and 3, the claim is that narrow syntax is easier to acquire than functional morphology because it is processed by means of universal operations, and is thus transferable. Functional morphology, on the other hand, is not transferable, which means that it must be lexically learned. In the current experiment, this suggests that learning English subject-verb agreement is more challenging than unlearning the V2 rule, as stated in prediction 1 and 2 above.

Furthermore, non-subject-initial declarative clauses with auxiliary verbs and lexical verbs are used to test knowledge about verb movement, and declarative clauses with local and long-distance agreement and singular and plural subjects are used to test knowledge about subject-verb agreement (see section 3.5.2). Prediction 3 (acquiring target-like word order in sentences

with auxiliary verbs is more difficult than in sentences with lexical verbs), is based on Pollard's (1996) argument that auxiliaries, unlike lexical verbs, are unspecified for the feature [inverted] in English. Lexical verbs, on the other hand, always stay in the VP, as they have the feature [-inverted] (see section 3.2). In other words, it is possible to move the auxiliary to C in some structures of English, which may cause the participants to be more inclined to move the auxiliary, in comparison to the lexical verb, in non-subject-initial declarative main clauses too.

Prediction 4A is based on studies which show that the structural distance between the subject and the verb increases the difficulty level of subject-verb agreement, i.e. it is expected that long-distance agreement is more challenging than local agreement because the former involves an additional element between the subject and the agreeing verb. In other words, there is an increased cognitive load in these types of sentences (see sections 3.2 and 2.4.3). With respect to prediction 4B (long-distance agreement is more difficult when the subject is 3rd person singular, in comparison to when it is 3rd person plural) is based on studies that have found that long-distance agreement is more challenging when the noun that is closest to the verb is plural, and the subject noun is singular. Ocampo (2013) points out that this was only found in prepositional phrases, and not in relative clauses (see sections 3.2 and 2.4.3). As the sentences in the current experiment contains prepositional phrases, it is expected that the same result is found here. Predicition 4C (local agreement is more difficult when the subject is 3rd person singular, than when it is 3rd person plural) builds on the fact that most studies on local agreement errors have found that dropping the 3rd person singular –*s* is more common than inserting a superfluous –*s* (hypercorrection) (see sections 3.2 and 2.4.1). In other words, agreement errors more frequently occur with singular subjects, as these require the suffix –*s* on the verb. Nevertheless, Dröschel (2011) points that both error types are frequently found, and for that reason, it is expected that the difference between the frequency of error types with local agreement is slight.

To repeat what the mean scores refer to, the Likert scale is treated as a binary variable (see sections 3.3 and 4.2) with the categories *unacceptable* and *acceptable*. These are represented by 0 (the number 1-2 on the Likert scale) and 1 (the number 3-4 on the Likert scale), respectively. If the mean score is low, it suggests that most participants have judged the sentence as unacceptable, and if it is high, it suggests that most participants have judged the sentence as acceptable. Consequently, a low mean score for an ungrammatical sentence and a high mean score for a grammatical sentence are considered correct judgements.

Furthermore, it is important to note that there is always a possibility that the participants judgements are based on performance factors, such as the naturalness of the sentence or processing difficulties, rather than their knowledge of grammar rules in English (see section 3.3). For instance, Ocampo (2013) argues that the variability in long-distance subject-verb agreement is caused by processing limitations, rather than the speakers' knowledge of agreement (see section 2.4.1). In other words, the participants might in fact know the rules of agreement in English, but are not able to apply them (see section 2.4.1). To specify, the rules of English subject-verb agreement are explicitly taught in Norwegian schools, whereas verb movement is implicitly acquired, i.e. acquired without awareness of the structures (Dekeyser 2003: 314) (see section 2.2). However, knowing a rule does not automatically mean that the performance is good, as the knowledge must be internalized and automated. In other words, judgements of long-distance agreement tell something about the speakers' attention to syntactic cues and the challenges that are involved in acquiring accurate functional morphology in English, but they do not necessarily say anything about the participants' knowledge of subject-verb agreement. Similarly, non-subject initial declarative clauses are infrequent in English, which means that the participants may not be exposed to these types of sentences very often, and thus find them unnatural (Westergaard 2003: 91). For that reason, it is not possible to be certain that the acceptability judgements show the participants' knowledge of the two conditions per se, and thus, I use the term *performance* rather than *knowledge* in the following sections.

5.1 Is functional morphology more difficult than narrow syntax in L2 acquisition?

The prediction in research question 1 is that subject-verb agreement is more difficult than verb movement. As illustrated in section 4.2.3 the mean score for judgements of ungrammatical agreement is significantly higher than the mean score for judgements of ungrammatical verb movement. This means that there are significantly more incorrect judgements with agreement when the sentences are ungrammatical. In contrast, the mean scores for both grammatical verb movement and grammatical agreement are high, i.e. the participants make few errors with both constructions when the sentences are grammatical. These results are also found in each of the proficiency groups, i.e. there are always more errors with agreement when the sentences are ungrammatical, and always few errors with both constructions when the sentences are grammatical (see section 4.2.3).

Nevertheless, the fact that there are few errors with grammatical sentences does not necessarily mean that the participants' performance is good. In fact, the low number of errors with subject-verb agreement indicates weak performance skills on this condition, as the participants accept both ungrammatical and grammatical sentences (see sections 4.2.1). In contrast, the participants usually reject ungrammatical verb movement and accept grammatical verb movement, which suggests good performance in verb movement.

Furthermore, the difference between judgements of agreement and verb movement was also tested when the judgements of long-distance agreement is excluded from the dataset (see section 4.2.7). The reason for this is this is that previous research has argued that processing limitations, rather than a lack of knowledge about subject-verb agreement, cause variability with long-distance subject-verb agreement (see section 2.4.3). In order to get a more nuanced picture of the participants' knowledge about subject-verb agreement, it is thus necessary to assess the participants' behaviour without this condition.

The results in section 4.2.7 show that excluding the judgements of long-distance agreement had few consequences for the mean scores of agreement, which means that the performance is still significantly weaker than with verb movement. To elaborate, the mean score for ungrammatical agreement changed minimally (0.67 when long-distance agreement is excluded from the dataset, in comparison to 0.68 when it is included in the dataset). The mean score for grammatical agreement changed slightly more (0.92 when long-distance agreement is excluded from the dataset, and 0.87 when it is included in the dataset). Consequently, there is still significantly more incorrect judgements with ungrammatical subject-verb agreement in comparison to ungrammatical verb movement. However, there is now also a significant difference between the judgements of grammatical agreement and grammatical verb movement, as the mean score for the former construction is now significantly higher than the mean score for the latter. In other words, there seems to be significantly more errors with ungrammatical verb movement when the sentences are grammatical. Nevertheless, as mentioned previously, few incorrect judgements does not mean that the participants' performance is good if it is a consequence of a general tendency towards ranking the sentences high, i.e. to accept both ungrammatical and grammatical sentences. As mentioned above, this is the case even when the judgements of long-distance agreement are not considered, which illustrates that the participants still accept both grammatical and ungrammatical agreement. It is thus possible to

argue that the results show that performance on subject-verb agreement is weak, whereas performance in verb movement is strong.

These observations indicate that unlearning the V2 rule is less difficult than learning the rules of subject-verb agreement in L2 acquisition of English by Norwegian speakers. This lends support to prediction 1: subject-verb agreement is more difficult than verb movement. Furthermore, these results support findings in previous studies, such as Haznedar (2001), Ionin and Wexler (2001) and Lardiere (1998a,b), who all found that L2 speakers of English are more accurate with syntactic phenomena in obligatory contexts, in comparison to their accuracy in morpho-syntactic phenomena related to the same functional category, such as subject-verb agreement and verb movement. Slabakova (2013) argues that these results indicate that L2 learners are able to engaged knowledge of syntactic operations, although they have not acquired the target functional morphology. This argues for the *syntax-before-morphology* view, and against the *morphology-before-syntax* view which states that acquisition of functional morphology drives acquisition of functional categories (see section 2.2).

5.2 Is functional morphology a more persistent problem than narrow syntax?

In order to test the Bottleneck Hypothesis, it is necessary not only to investigate the difficulty level of agreement and verb movement, but also the way in which acquisition of these two constructions develop. For that reason, the participants are split up into proficiency groups (see section 4.1), and the performance in each group is compared. In the following, I argue that although there is a significant difference between agreement and verb movement at each proficiency level (see section 4.2.1 and 4.2.2), the way in which performance in the two constructions develops shows that agreement is a more persistent problem than narrow syntax, as stated in prediction 2. I first discuss the correlation between proficiency scores and the way in which agreement and verb movement are judged, and secondly, I discuss and compare the proficiency groups' judgements of these two constructions.

As shown in section 4.2.4, there is a negative correlation between proficiency scores and the way in which ungrammatical sentences are judged, and a positive correlation between proficiency scores and judgements of grammatical sentences. This applies for both agreement and verb movement. However, the correlations differ with respect to how strong they are. A weak correlation suggests that all participants judge the sentences similarly, regardless of their proficiency score, and a strong correlation indicates that the sentences are judged differently

depending on the participants' proficiency level. Consequently, the fact that there is a strong negative correlation between proficiency scores and judgements of ungrammatical verb movement suggests that the participants' performance clearly increases as they become more advanced speakers of English (adjusted r^2-value = 0.4688). In comparison, the negative correlation between proficiency scores and judgements of ungrammatical agreement is weak, which indicates that a higher proficiency level and more exposure to English does not affect the judgements of agreement to the same degree as it affects the judgements of ungrammatical verb movement (adjusted r^2-value = 0.2227). This result is also found when long-distance agreement is excluded from the dataset (see section 4.2.7). In other words, this indicates that performance in ungrammatical verb movement develops more than performance on ungrammatical subject-verb agreement.

With grammatical sentences, the positive correlation between proficiency scores and judgements is weak for both agreement and verb movement (adjusted r2-value = 0.06865 and 0.1159, respectively). With respect to agreement, the correlation becomes stronger when the judgements of long-distance agreement are not considered, but it is still relatively weak (adjusted r^2-value = 0.2524) (see section 4.2.7). Furthermore, as the mean scores are high for both constructions, as illustrated in section 5.1, the weak correlations indicate that that all proficiency levels make few incorrect judgements. However, as also noted in the previous section, this does not mean that the performance is equally good in agreement and verb movement, as the participants with lower proficiency scores tend to accept both ungrammatical and ungrammatical sentences.

With respect to the proficiency groups, the general tendency is that the participants accept both grammatical and ungrammatical agreement until the advanced stage, whereas only the low intermediate and intermediate speakers accept ungrammatical and grammatical verb movement. This strengthens the above argument that performance in verb movement develops more than performance on subject-verb agreement. More specifically, the results in section 4.2.3 show that ungrammatical and grammatical agreement is often accepted by the low intermediate, intermediate and high intermediate groups, and that it is not until the advanced stage that there is a clear difference between the way in which sentences with ungrammatical and grammatical agreement are judged. In addition, the low intermediate and intermediate groups also often accept ungrammatical and grammatical verb movement. In other words, these two groups illustrate low performance skills in both agreement and verb movement. At the high intermediate stage, there is a clear difference between the way in which grammatical and

ungrammatical sentences with verb movement are judged, i.e. ungrammatical sentences are usually rejected, and grammatical sentences are most often accepted. The same result is found for the advanced learners. This indicates that in the high intermediate group, the performance is weak in agreement and strong in verb movement, and for the advanced group, the results suggest that their performance is relatively strong in both agreement and verb movement, although it is stronger in verb movement. In other words, ungrammatical verb movement improves considerably from the intermediate stage to the high intermediate stage, whereas ungrammatical agreement seems to stagnate. This supports prediction 2: subject-verb agreement is a more persistent problem than verb movement in L2 acquisition of English, and consequently, also the Bottleneck Hypothesis.

5.3 Which of the morphological and syntactic conditions are more difficult?

In this section, I discuss the judgements of the syntactic and morphological conditions in turn with the purpose of identifying the structures that are more challenging for the L2 learners.

5.3.1 Syntactic conditions

The prediction for research question 3 with respect to syntactic conditions is that acquiring target-like English word order with auxiliary verbs is more difficult than with lexical verbs (see section 5.1). In the following, I argue that the results found in the current study supports this prediction.

When the judgements of all participants are considered, the performance is good with both auxiliaries and lexical verbs, as the mean scores are low for ungrammatical sentences and high with grammatical sentences (see section 4.2.6). Examples of these types of sentences are illustrated in (71) and (72). However, the mean score for ungrammatical sentences with auxiliary verbs is significantly higher than the means score for ungrammatical sentences with lexical verbs, which means that the participants make more incorrect judgements with the former.

(71) *Grammatical non-subject-initial declarative main clauses*

    a. In December the kids will celebrate Christmas

    b. Yesterday the teacher went to the shop

(72)   *Ungrammatical non-subject-initial declarative main clauses*

    a.   *In December will the kids celebrate Christmas

    b.   *Yesterday went the teacher to the shop

Regarding the proficiency groups' judgements, section 5.2 argues that the participants' performance in verb movement is weak at the low intermediate and intermediate stages, and strong at the high intermediate and advanced stages. That is, at the lowest proficiency levels, the participants generally accept both ungrammatical and grammatical sentences, whereas they reject ungrammatical sentences and accept grammatical sentences at the two highest proficiency levels. In the following, I discuss which one of the syntactic conditions the participants struggle more with. As the grammatical sentences are ranked high in all proficiency groups, and there are no statistically significant differences between the auxiliary verbs and the lexical verbs, I focus on the judgements of ungrammatical sentences.

In the low intermediate group, ungrammatical sentences with lexical verbs and auxiliaries are both ranked high, i.e. they are usually accepted. There is no statistically significant difference between the mean scores for each structure. In other words, the mean scores cannot tell us with which syntactic structure there are more problems. Instead, it is possible to conclude that this supports the argument in 5.2, i.e. that low intermediates struggle with verb movement in general. In the intermediate group, there are also a lot of errors with both syntactic conditions. However, in this group, there are significantly more incorrect judgements of sentences with auxiliary verbs. Similarly, in the high intermediate and advanced group, in which the performance is considerably better, there are significantly more errors with auxiliary verbs (see section 4.2.6). In other words, this suggests that once the participants have acquired the V2 rule, they make few errors with both auxiliaries and lexical verbs, but they make significantly more errors with auxiliary verbs.

The above observations give support to prediction 3, which is that verb movement in non-subject initial declarative clauses with auxiliary verbs is the most difficult syntactic condition. As suggested in the introduction of this section and in section 3.2, a possible explanation for this is that learners are more inclined to move auxiliaries in L2 English, as these verbs move in some structures in English, such as in (73), i.e. they are unspecified for the feature [inverted] (see section 2.3).

(73)   Hvor **gikk** Peter i   går?

         Where went Peter in   yesterday?

         'Where **did** Peter **go** yesterday?'

As non-subject-initial declarative main clauses are infrequent in the input (see section 2.4.5), the participants may not be aware of the target structure. When learners acquire a language, they make hypothesis about the structure on the basis of their input, i.e. the more frequent main clauses. Consequently, as the English input shows the learners that there are more options for the placement of the auxiliary, in comparison to the lexical verb, they may be more inclined to move auxiliaries. This may explain why the participants in the current study seem to spend more time on unlearning the Norwegian word order *XP<modal<subject* in comparison to *X<lexical verb<subject.*

5.3.2   Morphological conditions

When the judgements of all 60 participants are considered, all of the morphological conditions, with both ungrammatical and grammatical sentences, are ranked relatively high (see section 4.2.5). This indicates that the performance level is weak for all conditions. However, long-distance agreement with plural subjects seems to be the most problematic condition, as it is ranked significantly lower than the other morphological conditions when the sentences are grammatical. In other words, whereas there are tendencies towards rejecting ungrammatical sentences and accepting grammatical sentences with the three other morphological conditions (long-distance agreement with singular subjects, local agreement with singular subjects, and local agreement with plural subjects), long-distance agreement with plural subjects is frequently accepted regardless of its grammaticality.

Regarding the proficiency groups, there are no significant differences between the judgements of the morphological conditions in the low intermediate and intermediate groups. The mean scores in these groups are all high, which suggests that all structures are difficult (see section 4.2.5). This changes in the high intermediate group, in which the judgements are similar to those seen in the overall mean scores. In other words, the performance is relatively low on all morphological conditions, but the performance on long-distance agreement with plural subjects seems to be the weakest. That is, the high intermediate participants reject grammatical sentences with long-distance agreement and plural subjects significantly more often than they reject the other grammatical sentence structures and they often accept the ungrammatical sentences too.

The same result is also found in the advanced group, but with some additions: Participants now differentiate between grammatical and ungrammatical sentences more clearly, as the mean scores for ungrammatical sentences is considerably lower than the mean scores for grammatical sentences. In other words, their performance on subject-verb agreement has generally improved, as suggested in section 5.2. As most differences between the morphological conditions are found in the advanced group, I look more closely at these participants' behaviour in the following paragraphs.

The advanced learners' judgements of grammatical long-distance agreement with plural subjects is around 0.67, which suggests that most participants accept sentences like (74a), whereas the mean score for ungrammatical sentences is around 0.56, which indicates that there are approximately the same amount of acceptance and rejection of sentences like (74b), although there are slightly more acceptance. It is possible to argue that these types of structures are more problematic than the other structures, in which the mean scores differ more.

(74)

      a.   The girls in the red car drink coffee

      b.   *The girls in the red car drinks coffee

In contrast, the performance on local agreement with singular subjects, as illustrated in (75) is clearly good, as the advanced participants accept grammatical sentences and reject ungrammatical sentences.

(75)

      a.   The girl drinks coffee

      b.   *The girl drink coffee

Furthermore, the mean scores for long-distance agreement with singular subjects and local agreement with plural subjects, as illustrated in (76) and (77), respectively, are similar. That is, the grammatical sentences are frequently accepted, whereas around half of the participants reject the ungrammatical sentences (around 0.50 and 0.48, respectively). It is thus hard to maintain that one of the sentence types in (76) and (77) is more difficult than the other, which

is unexpected, as long-distance agreement is predicted to be more difficult than local agreement (see prediction 4A).

(76)

    a. The girl with the heavy suitcases drinks coffee

    b. *The girl with the heavy suitcases drink coffee

(77)

    a. The girls drink coffee

    b. *The girls drinks coffee

I argue that the reason why local agreement and long-distance agreement seem to be equally problematic is caused by the fact that plural subjects generate more errors with agreement than singular subjects, whereas long-distance agreement cause more problems than local agreement. The result is a similar difficulty level in long-distance agreement with singular subjects and local agreement with plural subjects. This is based on the two following arguments: the most frequent agreement error is hypercorrection and long-distance agreement is more difficult than local agreement. This pattern is seen in the three most challenging conditions: first, with long-distance agreement and plural subjects, half of the participants accept a superfluous –s and just under half of them reject the target-like lack of the suffix –s. In long-distance agreement and singular subjects, most learners accept the correct suffix –s on the verb and as much as half of the participants accept the incorrect occurrence of the –s. In local agreement with plural subjects, the participants tend to correctly accept the absence of the –s on the verb, whereas as half of the participants accept the incorrect occurrence of the –s. In other words, the participants tend to prefer the occurrence of the suffix –s on the verb, which cause more errors with plural subjects in both long-distance and local agreement.

The fact that the participants seem to accept the absence of the –s in grammatical local agreement with plural subjects suggests that local agreement simplifies the establishment of agreement. This is also seen in the least problematic morphological construction, i.e. local agreement with singular subjects, as the participants seem to have no problems with this condition.

The above discussion shows that the results in the current experiment support the prediction in 4A, i.e. that long-distance agreement is more difficult than local agreement. Predictions 4B and 4C, on the other hand, are not supported, as 4B states that long-distance agreement with singular subjects is more difficult than long-distance agreement with plural subjects, and similarly, 4C states that local agreement with singular subjects is more difficult than local agreement with plural subjects (see section 3.2). This is not the case here, as hypercorrection is the most frequent error type with both local and long-distance agreement, which is in disagreement with the findings in previous studies that gave rise to prediction 4B (see section 2.4.1). Nevertheless, as mentioned in the introduction to this section and in section 3.2, both hypercorrection and missing 3rd person singular –s were expected to be found, i.e. errors with both singular and plural subjects.

Furthermore, prediction 4B was based on results found in, amongst others, Ocampo (2013) (see section 3.2). However, the results described above suggests the exact opposite, as Ocampo's (2013) reading task found that learners are less sensitive to agreement errors when the subjects are singular and the noun that is closest to the verb is plural in prepositional phrases (see section 2.4.3). Based on this, Ocampo (2013) argues that her results illustrate a weak plural markedness effect, i.e. that plural nouns are marked in English, and when these are placed between the subject and the verb, the markedness disrupts the establishment of agreement. When the noun that is closest to the verb is singular, no disruption occurs, as it is unmarked. This effect was not found in the current study, as there are significantly more errors with long-distance agreement when the subject is plural and the intervening noun is singular. I thus argue against the plural markedness effect as a cause of subject-verb agreement errors, and rather, claim that hypercorrection is the reason for the behaviour seen in the current experiment. In other words, prediction 4B, that long-distance agreement with singular subjects is more difficult than long-distance agreement with plural subjects, is not supported.

This further suggests that participants are in fact good at identifying the subject in long-distance agreement, although the difference in difficulty between local and long-distance agreement is clearly seen. I argue that this difference is caused by an increased cognitive load and lack of attention to syntactic cues.

6 Conclusion

In this thesis, the Bottleneck Hypothesis (Slabakova 2008; 2013) has been tested in L2 acquisition of English by Norwegian learners. According to this hypothesis, functional morphology is the bottleneck of L2 acquisition. Consequently, the prediction is that functional morphology is more difficult for the Norwegian speakers to acquire in comparison to syntactic operations. As the previous chapters have shown, the experiment presented in this thesis supports the Bottleneck Hypothesis. More specifically, the results show that the speakers' performance with functional morphology is weaker than their performance with syntactic operations, and also, that performance with functional morphology seems to become stagnant, although the speakers' proficiency in English increases.

In order to investigate this, we tested two constructions that do not match in English and Norwegian: Subject-verb agreement and verb movement. The former represents functional morphology and the latter represents syntax. To briefly repeat from chapter 2, the differences between Norwegian and English with respect to verb movement is that Norwegian is a V2 language, i.e. the finite verb moves to the second position in non-subject-initial declarative clauses, whereas English is an SVO language, i.e. the verb stays in the VP in these types of clauses. The learning task for Norwegian speakers of English is thus to unlearn the V2 rule. Westergaard (2003) has shown that this is challenging for Norwegian learners at an early stage. Regarding functional morphology, Norwegian has no overt agreement morphology, whereas English marks simple present tense verbs when the subject is 3rd person singular. Several studies have shown that functional morphology is not only challenging to acquire in a second language (see e.g. Slabakova and Gajdos 2008), but also that it comes later than syntax (see e.g. Ionin and Wexler 2002; Haznedar 2002; Lardiere 1998a,b).

In this thesis, syntax was tested by non-subject-initial declarative clauses with lexical verbs and auxiliary verbs, and functional morphology was tested with subject initial main clauses with plural and singular subjects, as well as local and long-distance agreement. In other words, two experimental conditions tested syntax, and four conditions tested functional morphology. The participants were asked to rank these sentences on a Likert scale from 1 to 4 in an untimed acceptability judgement test. In my analysis of the data, the Likert scale was treated as a binary variable, where 1-2 meant *unacceptable* and 3-4 stood for *acceptable*. The web-based test was carried out with SurveyGizmo.

Sixty native speakers of Norwegian in two age groups (11–12 and 15–18) participated in the experiment. They all had Norwegian as their only L1 and English as their L2. They were divided into four levels of proficiency on the basis of a Standardized Oxford Proficiency test with 40 multiple-choice questions. The proficiency scores ranged from low intermediate to advanced.

According to Sprouse et al. (2012:22), acceptability judgement tests are reliable sources of information about L2 speakers' interlanguage. However, it is not possible to be completely certain that the speakers' judgements illustrate their linguistic competence, as performance factors like processing limitations or naturalness of the sentences may affect their evaluation. As variability in long-distance agreement has been argued to be caused by processing limitations rather than linguistic knowledge, the acceptability judgements of subject-verb agreement and verb movement were also compared when the judgements of long-distance agreement were excluded from the data. However, this had a minimal effect on the overall mean scores of the sentences.

Furthermore, as several different structures were tested, we investigated whether some of them were significantly more difficult than others. With respect to the morphological conditions, the experiment shows that long-distance agreement is more difficult than local agreement, i.e., adding an additional element between the subject and the verb causes problems for the learners. This supports findings in previous experiments, such as Ocampo's (2013) comprehension study, which shows that more structural complexity in a sentence makes subject-verb agreement more challenging. In addition, subject-verb agreement is more difficult when the subjects are plural, in comparison when they are singular, with both local and long-distance agreement. In the sentences with long-distance agreement, the subject and the intervening noun differed in number, i.e. there was a singular intervening noun when the subject was plural. The results indicate that the participants prefer sentences that contain the 3rd person singular suffix –*s,* regardless of whether or not the subject is singular or plural. One possible explanation for this is that the participants overgeneralize the 3rd person singular –*s*. Based on previous studies on the effect of the subjects' number on agreement errors, this result was not expected, as research has shown that most errors in both comprehension studies, such as the current experiment, and production studies happen when the subject is singular in local agreement and when the subject is singular and the intervening noun is plural in long-distance agreement. The results thus argue against the plural markedness effect, which holds that the marked feature [plural] on the noun that immediately precedes the verb disrupts the establishment of agreement.

Regarding the syntactic conditions, the sentences with auxiliary verbs were more problematic than the sentences with lexical verbs, as expected. I argue that the reason for this is related to the fact that auxiliary verbs move to the CP-domain in some structures of English, whereas lexical verbs never move out of the VP. The fact that non-subject-initial declarative clauses are infrequent in English means that the learners do not receive much exposure to the target word order in these types of sentences. Consequently, they are more inclined to move the auxiliary verb, in comparison to the lexical verb, as the input from the more frequent main clauses has shown them that this is in fact possible in some situations. The consequence of this is more incorrect verb movement with auxiliary verbs than with lexical verbs.

In sum, although there are remaining questions, the experiment presented in this thesis supports the Bottleneck Hypothesis, which may contribute to current knowledge about the cognitive process of L2 acquisition. Furthermore, the results give implications for language teaching, as they illustrate that even at advanced proficiency levels, learners struggle with subject-verb agreement, although this is explicitly taught in Norwegian schools. In order to further investigate the Bottleneck Hypothesis, it is necessary to also look at functional morphology in comparison to other domains than narrow syntax, such as semantics and the interfaces, as well as different language combinations. It would also be interesting to test the Bottleneck Hypothesis in an on-line experiment, such as eye-tracking or a speeded acceptability judgement test, in order to investigate the L2 learners' automatic and implicit knowledge.

7 References

Anderssen, Merete. 2006. The Acquisition of Compositional Definiteness in Norwegian. PhD dissertation, University of Tromsø.

Adger, David. 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Adger, David. 2006. Combinatorial Variability. *Journal of Linguistics* 42.3, 503-530.

Anderssen, Merete and Marit Westergaard. 2010. Frequency and economy in the acquisition of variable word order. *Lingua* 120.11, 2569-2588.

Bentzen, Kristine. 2009. Subject positions and their interaction with verb movement. *Studia Linguistica* 63.3, 261-291.

Bentzen, Kristine. 2014. Verb Placement in Main and Embedded Clauses. *Nordic Atlas of Language Structures (NALS) Journal* 1, 207-210.

Bini, Milena F. 1993. Laadquisicíon del italiano: mas allá de las propiedades sintácticas del parámetro pro-drop [The acquisition of Italian: beyond the syntactic properties of the pro-drop parameter]. *La linguistica y el analisis de los sistemas no natives.* Ottawa: Doverhouse.

Birdsong, David. 1992. Ultimate Attainment in Second Language Acquisition. *Language* 68.4, 706-755

Bley-Vroman, Robert. 2009. The Evolving Context of the Fundamental Difference Hypothesis. *SSLA* 31, 175-198.

Bock, Kathryn and J. Cooper Cutting. 1992. Regulating mental energy: Performance units in language production. *Journal of Memory and Language* 31, 99-127.

Bock, Kathryn and Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23, 45-93.

Bolander, Maria. 1988. Is there any order? On word order in Swedish learner language. *Journal of Multilingual and Multicultural Development* 9, 97-113.

Breiteneder, Angelika. 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.

Cattell, N.R. 2006. *An Introduction to Mind, Consciousness and Language.* London: Continuum.

Chen, Lang et. al. 2007. ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition* 10.2, 161-174.

Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton.

Chomsky, 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. *Step by Step: essays on Minimalist syntax in honour of Howard Lasnik*, 89–115. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. *Ken Hale: a Life in Language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. *Structures and Beyond: The Cartography of Syntactic Structures, Vol. 3,* 104-131. Oxford: Oxford University Press.

Chomsky, Noam. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36.1, 1-22.

Clahsen, Harald, Claudia Felser, Kathleen Neubauer, Mikako Stato and Renita Silva. 2010. Morphological structure in native and nonnative language processing. *Language Learning* 60, 21-43.

Clahsen, Harald, Martina Penke and Teresa Parodi. 1993/1994. Functional Categories in Early Child German. *Language Acquisition* 3.4, 395-429.

Colliander, Gunilla. 1993. Profiling Swedish second language acquisition: a method for assessing L2 proficiency. *Problem, process, product in language learning. Papers from the Stockholm-Åbo conference, 21-22 October 1992*. Department of Linguistics, Stockholm University.

Cook, V.J. and Mark Newson. 2007. *Chomsky's Universal Grammar: An Introduction.*
Malden, MA: Blackwell Publishing.

Coppieters, René. 1987. Competence Differences between Native and Near-Native Speakers.
Language 63.3, 544-573.

Corder, Stephen P. 1967. The Significance of Learner's errors. *IRAL: International Review of Applied Linguistics in Language Teaching* 5.1-4, 161-170.

Crain, Stephen and Mineharu Nakayama. 1987. Structure Dependence in Grammar Formation. *Language* 63.3, 522-543.

Dabrowska, Ewa. 2010. Naïve v. expert intuitions: An empirical study of acceptability judgments. *The Linguist Review* 27, 1-23.

Dekeyser, Robert. 2003. Implicit and Explicit Learning. *The Handbook of Second Language Acquisition.* Malden, MA: Blackwell.

Duskova, Libuse. 1969. On sources of error in foreign language learning. *International Review of Applied Linguistics* 4, 11-36.

Dröschel, Yvonne. 2011. *Lingua Franca English: The Role of Simplification and Transfer.*
Bern: Peter Lang.

Dypedahl, Magne, Hilde Hasselgård and Berit Løken. 2012. *Introducing English Grammar.*
Bergen: Fagbokforlaget.

Eide, Kristin Melum. 2015. *Norsk som Andrespråk.* Oslo: Novus Forlag.

Enger, Hans-Olav and Kristian Emil Kristoffersen. 2000. *Innføring i Norsk Grammatikk: Morfologi og Syntaks.* Oslo: Cappelen akademisk forlag.

Falk, Yehuda N. 2012. Superiority Effects. The Hebrew University of Jerusalem: Department of Linguistics

Felix, Sascha W. Felix. 1988. UG-generated knowledge in adult second language acquisition.
Linguistic Theory in Second Language Acquisition. Dordrecht: Kluwer Academic Publishers.

Franck, Julie, Gabriella Vigliocco and Janet Nicol. 2002. Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes* 17.4, 371-404.

Gass, Susan M. 2013. *Second Language Acquisition: An Introductory Course.* Florence, KY: Taylor and Francis.

Guasti, Maria T. 2004. *Language Acquisition: The Growth of Grammar.* Cambride, MA: MIT Press.

Guido Mendes, Celia and I.C Iribarren. 2007. Fixação do parâmetro do sujeito nulo na acquisição do português europeu por hispanofalantes. *XXII Encontro Nacional da Associação Portuguesa de Linguística: Textos seleccionados.* Lisbon: Associação Portuguesa de Linguística.

Haznedar, Belma. 2001. The Acquisition of the IP system in child L2 English. *Studies in second Language Acquisition* 23, 1-39.

Hermas, Abdelkader. 2014. Multilingual transfer: L1 morphosyntax in L3 English. *International Journal of Language Studies* 8.2, 1-24.

Holmberg, Anders. 2010. Verb Second. *Syntax: an International Handbook of Contemporary Syntactic Research. 2nd Edition.* Berlin: de Gruyter.

Hopp, Holger. 2007. Ultimate attainment at the interfaces in second language acquisition: Grammar and processing. PhD dissertation, University of Groningen.

Hülmbauer, Cornelia. 2010. *English as a lingua franca between correctness and effectiveness. Shifting constellations.* Saarbrüchen: VDM Verlag Dr.Müller.

Hyltenstam, Kenneth. 1977. Implicational patterns in interlanguage syntax variation. *Language Learning* 27, 383-411.

Håkansson, Gisela. 1989. The acquisition of negative placement in Swedish. *Studia Linguistica* 43, 47-58.

Håkansson, Gisela and Sheila Dooley Collberg. 1994. The preference for Modal + Neg: an L2 perspective applied to Swedish L1 children. *Second Language Research* 10.2, 95-124.

Ionin, Tania and Kenneth Wexler. 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.

Jackendoff, Ray. 2002. *Foundationsof Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jiang, Nan. 2004. Morphological insensitivity in second language processing. *Applied Psycholinguistics* 25, 603-634.

Johnson, Keith. 2008. *Quantitative methods in Linguistics.* Malden, MA: Blackwell Publishing.

Johnson, Jaqueline S and Elissa L. Newport. 1989. Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology* 21, 60-99.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax.* Cambridge, MA: MIT Press.

Keating, Gregory D. 2009. Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59, 503-535.

Kellerman, Eric and Michael Sharwood Smith. 1986. *Crosslinguistic influence in second language acquisition.* Oxford: Pergamon Press.

Krashen, Stephen D. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

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

Lardiere, Donna. 1998b. Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research* 14, 359-375.

Larsson, Ida and Janne Bondi Johannessen. 2015. Incomplete Acquisition and Verb Placement in Heritage Scandinavian. *Moribound Germanic Heritage Languages in North America: Theoretical Perspectives and Empirical Findings.* Leiden: BRILL.

Leow, Ronald P.1996. Grammaticality judgement tasks and second-language development. *Georgetown University Round Table on Languages and Linguistics 1996: Linguistics, language acquisition and language variation: Current trends and future prospects.* Washington, DC: Georgetown University Press.

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis.*Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lopez Prego, Beatriz. 2012. Examining the nature of variability in gender and number agreement in native and non-native Spanish. Master's Thesis, University of Kansas.

Lozano, Cristobal. 2006. The development of the syntax-discourse interface: Greek learners of Spanish. *The acquisition of syntax in Romance languages.* Amsterdam: John Benjamins.

Margaza, Panagiota and Aurora Bel. 2006. Null subjects at the syntax-pragmatics interface: Evidence from Spanish interlanguage of Greek speakers. *Proceedings of GASLA 2006.* Sommerville, MA: Cascadilla Press.

McCarthy, Corrine. 2008. Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research* 24, 469-486.

NSD: Norsk Senter for Forskningsdata. *NSD: Personvernombudet for forskning.* [ONLINE] Available at: http://www.nsd.uib.no/personvern/meldeplikt/vanlige_sporsmal.html?id=6. Accessed 30 April 2016]

Ocampo, Sally. 2013. Morphological variability in long-distance subject-verb agreement: A study of native and nonnative processing. PhD dissertation, University of Kansas.

Odlin, Terence. 2003. Cross-Linguistic Influence. *The Handbook of Second Language Acquisition.* Malden, MA: Blackwell.

Paradis, Michel. 2004. *A Neurolinguistic Theory of Bilingualism.* Amsterdam: John Benjamins.

Pesetsky, David and Esther Torrego. 2001. T-to-C Movement: Causes and Consequences. *Current Studies in Linguistics* 36*,* 355-426.

Pollard, Carl. J. 1996. On head non-movement. *Discontinuous Constituency*. Berlin: Muoton de Gruyter.

Rognes, Stig. 2011. V2, V3, V4 (and maybe even more): The syntax of questions in the | Rogaland dialects of Norwegian. Master's Thesis, University of Oslo.

Rothman, Jason, Jennifer Cabrelli Amaro and Kees de Boot. 2013. Third Language Acquisition. *Cambridge Handbook of Second Language Acquisition.* West Nyack, NY: Cambridge University Press.

Santorini, Beatrice and Anthony Kroch. 2007. *The syntax of natural language: An online introduction using the Trees program.* [ONLINE] Available at: http://www.ling.upenn.edu/~beatrice/syntax-textbook [Accessed 20 April 2016]

Schwatz, Bonnie D. and Rex A. Sprouse. 1994. Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. *Language Acquisition Studies in Generative Grammar.* Amsterdam: John Benjamins.

Schwartz , Bonnie D. and Rex A. Sprouse. 1996. L2 Cognitive States and the Full Transfer/Full Access Model. *Second Language Research* 12, 40-72.

Slabakova, Roumyana. 2006. Is There a Critical Period for Semantics? *Second Language Research* 22.3, 302-338.

Slabakova, Roumyana. 2008. *Meaning in the Second Language*. Berlin: Mouton de Gruyter

Slabakova, Roumyana and Johnathan Gajdos. 2008. The Combinatorial Variability Hypothesis in the Second Language. *Selected Proceedings of the 2007 Second Language Research Forum.* Somerville, MA: Cascadilla Proceedings Project.

Slabakova, Roumyana. 2013. What is easy and what is hard to acquire in a second language: A generative perspective. *Contemporary approaches to second language acquisition* 9.5, 528.

Slabakova, Roumyana. 2015. What is easy and what is hard to acquire in a second language? [Power Point Slides]. Retrieved from <http://www.researchgate.net/publication/267917650_What_is_easy_and_what_is_hard_to_a cquire_in_a_second_language!>

Slabakova, Roumyana and Maria del Pilar Garcia Mayo. 2015. The L3 syntax-discourse interface. *Bilingualism: Language and Cognition* 18.2, 208-226.

Sorace, Antonella and Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22.3, 339–368.

Sorace, Antonella. 2011. Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism* 1.1, 1-33.

Söderbergh, R. 1974. *Projektet Barnespråkssyntax. Forskningsplan.* PM 2. Stockholm Universitet: Institutionen för nordiska språk.

Sprouse, Jon and Diogo Almeida. 2011. Power in acceptability judgment experiments and the reliability of data in syntax. Ms., University of California, Irvine & Michigan State University.

Sprouse, Jon and Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *J. Linguistics* 48.03, 609-652.

Sprouse, Jon, Carson T. Schütze and Diogo Almeida. 2012. Assessing the reliability of journal data in syntax: Linguistic Inquiry 2001-2010. Ms., University of California and Michigan State University.

Ullmann, Michael T. 2001. A Neurocognitive Perspective on Language: the Declarative/ Procedural Model. *Nature reviews neuroscience* 2.10, 717-726.

Vainikka, Anne and Martha Young-Scholten. 1994. Direct Access to X'-theory: evidence from Korean and Turkish adults learning German. *Language Acquisition Studies in Generative Grammar.* Amsterdam: John Benjamins.

Vainikka, Anne and Martha Young-Scholten. 1996. Gradual development of L2 phrase structure. *Second Language Research* 12.1, 7-39.

Valenzuela, Elena. 2006. L2 end state grammars and incomplete acquisition of Spanish CLLD constructions. *Inquiries in linguistic development: In honor of Lydia White*. Amsterdam: John Benjamins.

Vangsnes, Øystein A. 1996. Om-spørsmål i rogalandsk. *Nordica Bergensia* 9, 166-188.

Vettorel, Paola. 2014. *English as a Lingua Franca in Wider Networking: Blogging Practices*. Berlin: De Gruyter Mouton.

Westergaard, Marit. 2003. Unlearning V2. Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA Yearbook* 3, 77-101.

Westergaard, Marit and Øystein Vangsnes. 2005. *Wh*-questions, V2, and the Left Periphery of Three Norwegian Dialect Types. *Journal of Comparative Germanic Linguistics* 8, 117-158.

Westergaard, Marit. 2009. Microvariation as Diachrony: A View from Acquisition. *J Comp German Linguistics* 12, 49–79.

White, Lydia. 2000. Second Language Acquisition: From Initial to Final State. *Second Language Acquisition and Linguistic Theory.* Malden, MA: Blackwell.

White, Lydia. 2003a. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

White, Lydia. 2003b. On the Nature of Interlanguage Representation. *The Handbook of Second Language Acquisition.* Malden, MA: Blackwell.

White, Lydia. 2003c. Fossilization in steady state grammars: Persistent problems with inflectional morphology. *Bilingualism and Cognition* 6, 128-141.

White, Lydia and Fred Genesee. 1996. How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research* 12.3, 233-265.

Word Frequency. 2014. *Word Frequency Data: Corpus of Contemporary American English.* [ONLINE] Available at: http://www.wordfrequency.info/free.asp?s=y. [Accessed 1 December 2015]

8 Appendix

Appendix 1: The pilot tests

A: The pilot participants' mean score of judgements for ungrammatical sentences when the Likert scale is treated as an ordinal variable

Prof. score	Age	Agreement, plural singular	Agreement, long singular	Agreement, local plural	Agreement, local singular	Verb movement, lexical verbs	Verb movement, auxiliaries
21	12	3.20	3.00	3.20	3.00	1.80	2.20
25	17	2.50	2.50	2.83	3.17	1.00	1.50
27	16	3.33	2.17	3.00	2.50	1.83	2.33
28	14	3.50	3.33	3.83	3.50	2.83	3.33
33	18	3.00	3.17	3.00	3.00	1.83	2.50
34	17	1.83	2.00	1.00	1.00	1.00	1.00
35	12	3.08	3.25	3.50	3.08	1.17	2.00

B: The pilot participants' mean score of judgements for ungrammatical sentences when the Likert scale is treated as a binary variabel

Prof. score	Age	Agreement, long plural	Agreement, long singular	Agreement, local plural	Agreement, local singular	Verb movement, lexical verb	Verb movement, auxiliaries
21	12	0.83	0.83	0.83	0.67	0.17	0.17
25	17	0.50	0.67	0.83	1.00	0.00	0.17
27	16	0.83	0.33	0.83	0.50	0.17	0.50
28	14	1.00	0.83	1.00	1.00	0.67	0.83
33	18	0.67	1.00	0.83	1.00	0.00	0.50
34	17	0.33	0.33	0.00	0.00	0.00	0.00
35	12	0.67	0.75	0.83	0.75	0.08	0.33

Appendix 2: The instructions

Velkommen!

På de neste sidene kommer du til å se noen engelske setninger, og vi vil vite om du syns de er bra eller dårlige på engelsk. Noen setninger inneholder feil. Ingen av feilene er stave-, punktum- eller kommafeil.

Du kan velge mellom fem alternativer når du bedømmer setningene:

1	2	3	4	Vet ikke
☹	☹	☺	☺	○
1 = Dårlig	2 = Litt dårlig	3 = Litt god	4 = God	

Her er to eksempler. Hvis du syns setningen er dårlig på engelsk, trykker du på det sinte fjeset. Hvis du syns setningen er god på engelsk, trykker du på det smilende fjeset. Hvis du er usikker, trykker du på "Vet ikke".

1. Mother my house was in

1	2	3	4	Vet ikke
😠	☹	☺	☺	○

2. I speak English

1	2	3	4	Vet ikke
☹	☹	☺	😀	○

Takk for at du deltar! Ved å trykke på "next" godtar du at resultatene kan brukes i forskning.

Next

0%

Appendix 3: The acceptability judgement test

1. The boys in the black car look very scary. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

2. Every year the students must pass their exams. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

3. The girl drive to work every Wednesday morning. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

4. Mother my read a book about flowers. •
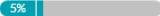
1	2	3	4	Vet ikke
☹	☹	☺	☺	○

5. The parents with the nice car talks to their kids. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

5%

6. The kids likes to play in the park every weekend. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

7. Boy the with the blue jacket ate an apple. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

8. The parents with the nice car talk to their kids. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

9. Last weekend the students drank lots of coffee. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

10. The girl drink a lot of water every day. •

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

10%

11. Thief the stole many paintings yesterday. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

12. The cats with long white fur drinks milk every day. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

13. The student loves to read books about football. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

14. Sometimes the dog ate the family's shoes. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

15. Every day the students should bring their books to school. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

14%

16. The students sit in the park after school. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

17. Last weekend drank the students lots of coffee. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

18. The man enjoyed to drive car his to work. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

19. Tomorrow the football player might score a goal. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

20. The girls with blonde hair likes to sit in the park. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

19%

21. The teacher eats fish for dinner every Friday. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

22. The kids with the red bike play in the garden. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

23. Every year must the students pass their exams. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

24. The boy with blue eyes seems very happy. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

25. The black shoes are too small for woman the. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

24%

---

26. Many people like to wear shorts in summer the. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

27. Yesterday celebrated the girl her birthday. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

28. The teacher with black shoes walks to work every day. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

29. The sisters love to run in the forest. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

30. The boy with blue eyes seem very happy. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

29%

31. Yesterday the teacher went to the shop. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

32. Monkey the loved to eat bananas for dinner. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

33. Those tourists with the heavy suitcase seems tired. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

34. The brothers attend football practice every day. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

35. Hopefully the student passed her exam. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

33%

---

36. The girl with golden earrings take the bus to school. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

37. Tomorrow the sisters must clean their room. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

38. Garden my had many different types of flowers. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

39. The teachers gives their students a lot homework. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

40. Yesterday the girl celebrated her birthday. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

38%

41. Dogs the barked at the postman with the red jacket. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

42. The boy likes to go swimming in the ocean. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

43. Every Monday the girls played football in the park. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

44. The students sits in the park after school. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

45. The boy with broken arms try to read a book. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

43%

46. The kids with the red bike plays in the garden. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

47. Tomorrow must the sisters clean their room. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

48. The teacher eat fish for dinner every Friday. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

49. In December the kids will celebrate Christmas. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

50. The girls with blonde hair like to sit in the park. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

48%

51. Every Monday played the girls football in the park. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

52. The brown dog plays with the yellow football. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

53. The house with yellow and white doors look nice. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

54. The cats with long white fur drink milk every day. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

55. The boy with broken arms tries to read a book. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

Next

52%

---

56. In December will the kids celebrate Christmas. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

57. The girl with golden earrings takes the bus to school. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

58. The brown dog play with the yellow football. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

59. Hopefully passed the student her exam. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

60. The boy like to go swimming in the ocean. *

| 1 ☹ | 2 ☹ | 3 ☺ | 4 ☺ | Vet ikke ○ |

Next

57%

118

61. The teachers give their students a lot of homework. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

62. Student the often drank coffee in the morning. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

63. The brothers attends football practice every day. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

64. The book about fast cars makes the girl happy. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

65. Sometimes ate the dog the family's shoes. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

62%

66. The teacher with black shoes walk to work every day. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

67. The sisters loves to run in the forest. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

68. The student love to read books about football. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

69. The cats plays with the yellow and green ball. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

70. Every day should the students bring their books to school. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

67%

119

71. Tomorrow might the football player score a goal. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

72. The girl drives to work every Wednesday morning. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

73. The boys in the black car looks very scary. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

74. The cats play with the yellow and green ball. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

75. The man wrote a book about mother his. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

71%

120

76. Those tourists with the heavy suitcase seem tired. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

77. Children the played the guitar very well. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

78. The kids like to play in the park every weekend. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

79. Very soon the band will play their favourite song. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

80. The book about fast cars make the girl happy. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

76%

---

81. The boys in the black car look very scary. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

82. Every year the students must pass their exams. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

83. The girl drive to work every Wednesday morning. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

84. Mother my read a book about flowers. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

85. The parents with the nice car talks to their kids. *

1	2	3	4	Vet ikke
☹	☹	☺	☺	○

Next

81%

Appendix 4: The Standardized Oxford Proficiency test

Part 1:
Instructions: Please complete the sentences by selecting the best answer from the available answers below.

1) Water _____ at a temperature of 100° C.

☐   is to boil

☐   is boiling

☐   boils

2) In some countries _____ very hot all the time.

☐   there is

☐   is

☐   it is

3) In cold countries people wear thick clothes _____ warm.

☐   for keeping

☐   to keep

☐   for to keep

4) In England people are always talking about _____.

☐   a weather

☐   the weather

☐   weather

5) In some places _____ almost every day.

☐   it rains

☐   there rains

☐   it raining

6) In deserts there isn't _____ grass.

- ☐ the

- ☐ some

- ☐ any

7) Places near the Equator have _____ weather even in the cold season.

- ☐ a warm

- ☐ the warm

- ☐ warm

8) In England _____ time of year is usually from December to February.

- ☐ coldest

- ☐ the coldest

- ☐ colder

9) _____ people don't know what it's like in other countries.

- ☐ The most

- ☐ Most of

- ☐ Most

10) Very _____ people can travel abroad.

- ☐ less

- ☐ little

- ☐ few

11) Mohammed Ali _____ his first world title fight in 1960.

- ☐ has won

- ☐ won

- ☐ is winning

12) After he _____ an Olympic gold medal, he became a professional boxer.

- ☐ had won

- ☐ have won

- ☐ was winning

13) His religious beliefs _____ change his name when he became a champion.

- ☐ have made him

- ☐ made him to

- ☐ made him

14) If he _____ lost his first fight with Sonny Liston, no one would have been surprised.

- ☐ has

- ☐ would have

- ☐ had

15) He has traveled a lot _____ as a boxer and as a world-famous personality.

- ☐ both

- ☐ and

- ☐ or

16) He is very well known _____ the world.

- ☐ all in

- ☐ all over

- ☐ in all

17) Many people _____ he was the greatest boxer of all time.

- ☐ is believing

- ☐ are believing

- ☐ believe

18) To be the best _____ the world is not easy.

- ☐ from

- ☐ in

- ☐ of

19) Like any top sportsman, Ali _____ train very hard.

- ☐ had to

- ☐ must

- ☐ should

20) Even though he has now lost his title, people _____ always remember him as a champion.

- ☐ would

- ☐ will

- ☐ did

Part 2:

21) The history of _____ is

- ☐ airplane

- ☐ the airplane

- ☐ an airplane

22) _____ short one. For many centuries men

- ☐ quite a

- ☐ a quite

- ☐ quite

23) _____ to fly, but with

- ☐ are trying

- ☐ try

- ☐ had tried

24) _____ success. In the 19th century a few people

- ☐ little

- ☐ few

- ☐ a little

25) succeeded _____ in balloons. But it wasn't until

- ☐ to fly

- ☐ in flying

- ☐ into flying

26) the beginning of _____ **century that anybody**

- ☐ last

- ☐ next

- ☐ that

27) _____ **able to fly in a machine**

- ☐ were

- ☐ is

- ☐ was

28) _____ **was heavier than air, in other words, in**

- ☐ who

- ☐ which

- ☐ what

29) _____ **we now call a 'plane'. The first people to achieve**

- ☐ who

- ☐ which

- ☐ what

30) 'powered flight' were the Wright brothers. _____ was the machine

☐ His

☐ Their

☐ Theirs

31) which was the forerunner of the Jumbo jets and supersonic airliners that are _____ common

☐ such

☐ such a

☐ some

32) sight today. They _____ hardly have imagined that in 1969,

☐ could

☐ should

☐ couldn't

33) _____ more than half a century later,

☐ not much

☐ not many

☐ no much

34) a man _____ landed on the moon.

☐ will be

☐ had been

☐ would have

35) Already _____ is taking the first steps towards the stars.

☐ a man

☐ man

☐ the man

36) Although space satellites have existed _____ less

- ☐ since

- ☐ during

- ☐ for

37) than forty years, we are now dependent _____ them for all

- ☐ from

- ☐ of

- ☐ on

38) kinds of _____. Not only

- ☐ informations

- ☐ information

- ☐ an information

39) _____ being used for scientific research in

- ☐ are they

- ☐ they are

- ☐ there are

40) space, but also to see what kind of weather _____.

- ☐ is coming

- ☐ comes

- ☐ coming

Appendix 5: Background questionnaire

1. Hvor mange år er du?
2. Hva er morsmålet ditt?
3. Hvilket språk snakker du med…
   - Moren din?
   - Faren din?
   - Søsknene dine?
   - Vennene dine?
4. Hvor gammel var du da du begynte å lære engelsk?

Appendix 6: The sentences

	Grammatical	Ungrammatical
Non-subject-initial declarative main clauses with lexical verbs	Yesterday the teacher went to the shop	*Yesterday went the teacher to the shop
	Every Monday the girls played football in the park	* Every Monday played the girls football in the park
	Hopefully the student passed her exam	*Hopefully passed the student her exam
	Sometimes the dog ate the family's shoes	*Sometimes ate the dog the family's shoes
	Last weekend the students drank lots of coffee	*Last weekend drank the students lots of coffee
	Yesterday the girl celebrated her birthday	* Yesterday celebrated the girl her birthday
Non-subject-initial declarative main clauses with auxiliary verbs	Very soon will the band play their favourite song	*Very soon the band will play their favourite song
	Every day the students should bring their books to school	*Every day should the students bring their books to school
	Tomorrow the sisters must clean their room	*Tomorrow must the sisters clean their room
	Every year the students must pass their exams	*Every year must the students pass their exams
	Tomorrow the football player might score a goal	*Tomorrow might the football player score a goal
	In December will the kids celebrate Christmas	*In December the kids will celebrate Christmas
Main clauses with local agreement, singular subjects	The girl drinks a lot of water every day	*The girl drink a lot of water every day
	The boy likes to go swimming in the ocean	*The boy like to go swimming in the ocean
	The girl drives to work every Wednesday morning	*The girl drive to work every Wednesday morning
	The student loves to read books about football	*The student love to read books about football

	The teacher eats fish for dinner every Friday	*The teacher eat fish for dinner every Friday
	The brown dog plays with the yellow football	*The brown dog play with the yellow football
Main clauses with local agreement, plural subjects	The kids like to play in the park every weekend	*The kids likes to play in the park every weekend
	The teachers give their students a lot of homework	*The teachers gives their students a lot homework
	The cats play with the yellow and green ball	*The cats plays with the yellow and green ball
	The students sit in the park after school	*The students sits in the park after school
	The sisters love to run in the forest	* The sisters loves to run in the forest
	The brothers attend football practice every day	*The brothers attends football practice every day
Main clauses with long distance agreement, singular subjects	The house with yellow and white doors looks nice	*The house with yellow and white doors look nice
	The teacher with black shoes walks to work every day	*The teacher with black shoes walk to work every day
	The boy with blue eyes seems very happy	*The boy with blue eyes seem very happy
	The girl with golden earrings takes the bus to school	*The girl with golden earrings take the bus to school
	The boy with broken arms tries to read a book	*The boy with broken arms try to read a book
	The book about fast cars makes the girl happy	*The book about fast cars make the girl happy
Main clauses with long distance agreement, plural subjects	The boys in the black car look very scary	*The boys in the black car looks very scary
	The parents with the nice car talk to their kids	*The parents with the nice car talks to their kids
	The girls with short blonde hair like to sit in the park	* The girls with short blonde hair likes to sit in the park

The cats with long white fur drink milk every day	*The cats with long white fur drinks milk every day
Those tourists with the heavy suitcase seem tired	*Those tourists with the heavy suitcase seems tired
The kids with the red bike play in the garden	*The kids with the red bike plays in the garden

Fillers	*Father my went to the shop every Monday
	*Children the played the guitar very well
	*Student the often drank coffee in the morning
	*Thief the stole many paintings yesterday
	*The black shoes are too small for woman the
	*My garden had many different types of flowers
	*Monkey the loved to eat bananas for dinner
	*Many people like to wear shorts in summer the
	*Dogs the barked at the postman with the red jacket
	*Boy the with the blue jacket ate an apple
	*The man enjoyed to drive car his to work
	*The man wrote a book about mother his
	*Mother my read a book about flowers

Appendix 7: Information letter

Forespørsel om deltakelse i forskningsprosjektet

Jeg er en mastergradsstudent i engelsk lingvistikk ved UiT Norges Arktiske Universitet, som søker deltakere til et forskningsprosjekt. Prosjektet er tilknyttet forskningsgruppa LAVA (Language Acquisition, Variation and Attrition) ved Institutt for språkvitenskap. Formålet med studien er å teste kunnskap om engelsk setningsstruktur blant videregående elever og barneskoleelever med norsk som morsmål, og engelsk som andrespråk. Dette danner grunnlag for min mastergradsoppgave, som tar sikte på å identifisere hvilke deler av språk som er mest utfordrende i andrespråkstilegnelse. Formålet med dette er først og fremst å søke kunnskap om den kognitive prosessen som skjer i andrespråkstilegnelse, men vil også kunne gi informasjon som kan være nyttig i utviklingen av skolens språkundervisning.

Med utgangspunkt i dette, ønsker jeg to grupper med elever: en gruppe med barneskoleelever og en gruppe med elever fra videregående skole. Det er ønskelig med omtrent 30 deltakere i hver gruppe, og derfor en eller to klasser, slik at elevene utfører testen samtidig og har mulighet til å stille spørsmål dersom noe er uklart. Det vil være opp til skolen selv å bestemme hvilke elever som blir bedt om å delta i studien.

Hva innebærer deltakelse i prosjektet?

Det er viktig å understreke at resultatene av testen på ingen måte påvirker skoleprestasjoner, og vil kun bli brukt som statistisk data til dette prosjektet.

Datainnsamlingen utføres ved hjelp av en web-basert test som består av tre deler. I den første delen, som også er hoveddelen, får deltakerne se en rekke setninger som de blir bedt om å rangere på en skala fra 1-4. 1 betyr at setningen er ugrammatisk på engelsk, og 4 betyr at setningen er grammatisk. I den andre delen vil deltakerne se setninger som mangler visse ord eller fraser. Under setningen finner de tre alternativer, og de blir bedt om å velge det alternativet som gjør setningen grammatisk. Til slutt vil elevene bli bedt om å fylle ut et skjema med bakgrunnsinformasjon om seg selv, men dette er helt anonymt og inneholder ikke informasjon av sensitiv karakter. Eksempler på spørsmål som vil kan bli stilt her er alder, hvilket språk deltakerne bruker til daglig, og hvor lenge de har lært engelsk. Testen vil ta omtrent 30 minutter å gjennomføre.

Hva skjer med informasjonen?

Alle personopplysninger vil bli behandlet konfidensielt. De utfylte skjemaene vil kun være tilgjengelige for de involverte i mastergradsoppgaven, som vil si undertegnede, samt min veileder, professor Marit Westergaard. Skjemaene vil bli lagret på en passordbeskyttet, privat datamaskin for å ivareta konfidensialitet. Deltakerne vil heller ikke kunne gjenkjennes i publikasjonen. Prosjektet skal etter planen avsluttes 1. juni 2016, og all web-basert data vil da bli slettet.

Frivillig deltakelse

Det er naturligvis frivillig å delta i studien, og elevene kan på hvilket som helst tidspunkt, og uten å oppgi grunn, trekke seg. Deltakelse i studien vil ikke på noen måte innvirke på forholdet med skolen, ei heller dersom noen ikke ønsker å delta, eller trekker seg underveis.

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Ta gjerne kontakt dersom du/dere har spørsmål om prosjektet, eller ønsker mer informasjon.

Med vennlig hilsen,

Isabel Nadine Jensen
Mastergradsstudent
UiT Norges Arktiske Universitet, Institutt for språkvitenskap
Tlf: 92808122
E-post: ije023@post.uit.no

Appendix 8: Form of consent

Forespørsel om deltakelse i forskningsprosjektet

Til elever og foreldre/foresatte

Jeg er en mastergradsstudent i engelsk lingvistikk ved UiT Norges Arktiske Universitet, som søker deltakere til et forskningsprosjekt. Prosjektet er tilknyttet forskningsgruppa LAVA (Language Acquisition, Variation and Attrition) ved Institutt for språkvitenskap. Formålet med studien er å teste kunnskap om engelsk setningsstruktur blant videregående elever og barneskoleelever med norsk som morsmål, og engelsk som andrespråk. Dette danner grunnlag for min mastergradsoppgave, som tar sikte på å identifisere hvilke deler av språk som er mest utfordrende i andrespråkstilegnelse. Formålet med dette er først og fremst å søke kunnskap om den kognitive prosessen som skjer i andrespråkstilegnelse, men vil også kunne gi informasjon som kan være nyttig i utviklingen av skolens språkundervisning.

Med utgangspunkt i dette, ønsker jeg to grupper med elever: en gruppe med barneskoleelever og en gruppe med elever fra videregående skole. Det er ønskelig med omtrent 30 deltakere i hver gruppe, og derfor en eller to klasser, slik at elevene utfører testen samtidig og har mulighet til å stille spørsmål dersom noe er uklart. Det vil være opp til skolen selv å bestemme hvilke elever som blir bedt om å delta i studien.

Hva innebærer deltakelse i prosjektet?

Datainnsamlingen utføres ved hjelp av en web-basert test som består av tre deler. I den første delen blir deltakerne bedt om å rangere en rekke engelske setninger på en skala fra 1-4, basert på om de oppfatter setningen som grammatisk riktig, eller ikke. Den andre delen inneholder spørsmål som gir en indikasjon på elevens kunnskapsnivå i engelsk grammatikk. I den siste delen blir deltakerne bedt om å oppgi alder og hvilke språk de snakker. Testen er med andre ord helt anonym, og ingen sensitive personopplysninger innhentes.Testen vil ta omtrent 30 minutter. Resultatene av testen vil på ingen måte påvirke skoleprestasjoner, og vil kun bli brukt som statistisk data til dette prosjektet.

Hva skjer med informasjonen?

Alle personopplysninger vil bli behandlet konfidensielt. De utfylte skjemaene vil kun være tilgjengelige for de involverte i mastergradsoppgaven, som vil si undertegnede, samt min

veileder, professor Marit Westergaard. Skjemaene vil bli lagret på en passordbeskyttet, privat datamaskin for å ivareta konfidensialitet. Deltakerne vil heller ikke kunne gjenkjennes i publikasjonen. Prosjektet skal etter planen avsluttet 1. juni 2016, og all web-basert data vil da bli slettet.

Frivillig deltakelse

Det er naturligvis frivillig å delta i studien. Dersom eleven ønsker å delta i studien, og får tillatelse fra foreldre/foresatte, kan dere likevel, på hvilket som helst tidspunkt, trekke deres samtykke uten å oppgi noen grunn. Dersom dere trekker dere, vil alle opplysninger bli anonymisert. Deltakelse i studien vil ikke på noen måte innvirke på forholdet med skolen, ei heller dersom du/dere eller eleven ikke ønsker å delta.

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Ta gjerne kontakt dersom du/dere har spørsmål om prosjektet, eller ønsker mer informasjon.

Med vennlig hilsen,

Isabel Nadine Jensen
Mastergradsstudent
UiT Norges Arktiske Universitet, Institutt for språkvitenskap
Tlf: 92808122
E-post: ije023@post.uit.no

Samtykke til deltakelse i studien

Jeg har mottatt informasjon om studien, og er villig til å delta

------------------------------------ ------------------- ---

Elevens navn   Sted/dato   Foresattes underskrift

Appendix 9: The participants' age and proficiency scores

Age	Proficiency score	Proficiency group
12	11	Low intermediate
12	13	
12	13	
11	14	
12	14	
12	14	
12	15	
12	15	
12	17	
12	18	Intermediate
12	18	
12	18	
12	18	
12	19	
12	19	
12	19	
12	20	
12	20	
12	21	
12	22	
17	26	High intermediate
12	27	
16	28	

16	28	
16	28	
16	28	
16	28	
16	29	
16	29	
17	29	
12	30	
15	30	
15	31	
17	31	
18	31	
16	32	
16	32	
16	32	
17	32	
17	32	
16	33	Advanced
15	33	
16	33	
17	34	
18	34	
16	35	
16	35	
17	35	
18	35	

18	35	
17	35	
16	36	
18	36	
17	36	
16	36	
16	37	
18	37	
18	37	
17	37	
16	38	

Appendix 10: Correct judgements

Age	Prof. score	Gr. Agree. (in %)	Ungr. Agree. (in %)	Gr. verb move-ment (in %)	Ungr. verb movement (in %)	Total correct (in nr.)	Total nr. of sentences (in nr.)
12	14	83.33	4.16	91.67	8.33	33	72
12	17	83.33	12.5	75	25	35	72
12	11	41.67	37.5	16.67	91.67	32	72
12	14	45.83	20.83	50	25	25	72
12	13	83.33	12.5	75	25	33	72
12	15	75	4.16	66.67	16.67	29	72
12	13	66.67	8.33	83.33	8.33	29	72
12	15	100	4.16	100	33.33	41	72
11	14	87.5	25	83.33	25	5	72
12	21	66.67	20.83	66.67	58.33	36	72
12	19	91.67	20.83	75	16.67	38	72
12	18	79.17	20.83	66.67	25	35	72
12	20	87.5	12.5	91.67	8.33	36	72
12	18	95.83	4.16	91.67	50	41	72
12	22	100	0	91.67	8.33	36	72
12	20	91.67	8.33	91.67	25	38	72
12	19	100	0	100	8.33	37	72
12	18	95.83	0	91.67	8.33	35	72
12	19	91.67	8.33	75	16.67	38	72
12	18	83.33	12.5	75	50	38	72
12	30	95.83	0	91.67	16.67	36	72
12	27	91.67	4.16	83.33	91.67	44	72
16	28	91.67	8.33	75	83.33	43	72

16	28	100	66.67	50	83.33	56	72
16	32	91.67	16.57	100	75	47	72
16	29	62.5	29.17	83.33	91.67	43	72
16	32	70.83	25	83.33	75	42	72
16	28	91.67	29.17	91.67	75	49	72
16	32	83.33	37.5	100	100	54	72
15	31	87.5	79.17	83.33	58.33	61	72
16	32	100	20.83	100	100	53	72
16	28	95.83	4.16	83.33	58.33	41	72
16	29	100	16.57	83.33	100	50	72
15	30	75	33.33	83.33	100	48	72
17	32	91.67	4.16	100	66.67	43	72
16	28	75	20.83	91.67	66.67	42	72
17	31	100	0	91.67	16.67	37	72
17	29	83.33	95.83	66.67	66.67	59	72
17	26	91.67	16.57	75	83.33	45	72
18	31	58.33	66.67	83.33	100	52	72
17	35	91.67	87.5	100	66.67	63	72
18	35	83.33	37.5	91.67	100	53	72
18	35	95.83	29.17	100	66.67	50	72
17	35	75	33.33	83.33	91.67	47	72
16	37	75	95.83	91.67	50	62	72
18	37	91.67	54.17	83.33	75	53	72
18	36	95.83	41.67	83.33	83.33	53	72
18	37	95.83	91.67	91.67	100	68	72
17	36	87.5	100	100	100	69	72

17	34	66.67	58.33	100	100	54	72
17	37	100	70.83	100	100	65	72
16	36	100	33.33	91.67	100	55	72
16	33	100	25	83.33	100	52	72
18	34	91.67	100	91.67	91.67	68	72
16	36	87.5	70.83	100	83.33	60	72
16	35	95.83	8.33	66.67	100	45	72
15	33	62.5	37.5	75	100	45	72
16	35	100	95.83	75	75	65	72
16	38	95.83	29.17	83.33	100	52	72
16	33	91.67	4.16	83.33	83.33	43	72

Appendix 11: Correlation tests

<table>
<tr>
<td>A</td>
<td>
Residuals:

Min	1Q	Median	3Q	Max
-7.8553	-2.8063	0.0958	2.1937	11.3896

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-17.97715	0.37539	-47.89	<2e-16 ***
age	3.04897	0.02498	122.07	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.801 on 4318 degrees of freedom

Multiple R-squared: 0.7753,   Adjusted R-squared: 0.7753

F-statistic: 1.49e+04 on 1 and 4318 DF, p-value: < 2.2e-16
</td>
</tr>
<tr>
<td>B</td>
<td>
Residuals:

Min	1Q	Median	3Q	Max
-24.6464	-4.3124	0.5505	3.6470	14.3382

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.836	3.300	3.889	0.000262 ***
Proficiency	1.201	0.116	10.352	8.48e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.203 on 58 degrees of freedom

Multiple R-squared: 0.6488,   Adjusted R-squared: 0.6428

F-statistic: 107.2 on 1 and 58 DF, p-value: 8.476e-15
</td>
</tr>
<tr>
<td>C</td>
<td>Residuals:</td>
</tr>
</table>

	Min	1Q	Median	3Q	Max
	-0.57638	-0.06007	0.01001	0.09466	0.22969

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.668054	0.065642	10.177	1.61e-14 ***
Proficiency.score	0.006817	0.002307	2.955	0.00451 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1433 on 58 degrees of freedom

Multiple R-squared: 0.1309,   Adjusted R-squared: 0.1159

F-statistic: 8.732 on 1 and 58 DF, p-value: 0.004513

D

Residuals:

	Min	1Q	Median	3Q	Max
	-0.41416	-0.04437	0.03018	0.08468	0.19148

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.735918	0.059546	12.359	<2e-16 ***
Proficiency.score	0.004840	0.002093	2.313	0.0243 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.13 on 58 degrees of freedom

Multiple R-squared: 0.08444, Adjusted R-squared: 0.06865

F-statistic: 5.349 on 1 and 58 DF, p-value: 0.0243

E

Residuals:

	Min	1Q	Median	3Q	Max
	-0.69467	-0.14492	-0.06379	0.16954	0.60435

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.079962	0.107226	10.072	2.38e-14 ***
Proficiency.score	-0.027451	0.003769	-7.284	9.76e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.234 on 58 degrees of freedom

Multiple R-squared: 0.4778,   Adjusted R-squared: 0.4688

F-statistic: 53.06 on 1 and 58 DF, p-value: 9.761e-10

F

Residuals:

Min	1Q	Median	3Q	Max
-0.56948	-0.14954	0.04651	0.17328	0.41922

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.169399	0.121085	9.658	1.11e-13 ***
Proficiency.score	-0.018008	0.004256	-4.232	8.36e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2643 on 58 degrees of freedom

Multiple R-squared: 0.2359,   Adjusted R-squared: 0.2227

F-statistic: 17.91 on 1 and 58 DF, p-value: 8.363e-05

G

Residuals:

Min	1Q	Median	3Q	Max
-0.54976	-0.16753	0.06633	0.22541	0.48541

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.205833	0.139060	8.671	4.64e-12 ***

	Proficiency.score -0.019750 0.004887 -4.041 0.000159 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.3035 on 58 degrees of freedom Multiple R-squared: 0.2197, Adjusted R-squared: 0.2062 F-statistic: 16.33 on 1 and 58 DF, p-value: 0.0001586

H	Residuals:
	
	    Min 1Q Median 3Q Max
	
	  -0.45519 -0.03445 0.01920 0.06641 0.18071
	
	Coefficients:
	
	                   Estimate Std. Error t value Pr(>|t|)
	
	(Intercept) 0.703925 0.047846 14.712 < 2e-16 ***
	
	Proficiency.score 0.007691 0.001682 4.574 2.57e-05 ***
	
	--- 	
	
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
	
	Residual standard error: 0.1044 on 58 degrees of freedom
	
	Multiple R-squared: 0.2651, Adjusted R-squared: 0.2524
	
	F-statistic: 20.92 on 1 and 58 DF, p-value: 2.569e-05

Appendix 12: Normality test

A	Energy test of multivariate normality: estimated parameters
	data: x, sample size 60, dimension 1, replicates 999
	E-statistic = 0.46132, p-value = 0.3263

Appendix 13: Generalized linear mixed model

A	AIC	BIC	logLik	deviance	df.resid
	2115.2	2137.9	-1053.6	2107.2	2156

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-6.4776	-0.5240	0.1876	0.5369	7.6844

Random effects:

	Groups	Name	Variance	Std.Dev.
	Response.ID	(Intercept)	2.993	1.7300
	item	(Intercept)	0.226	0.4754

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.1404	0.2558	4.458	8.28e-06 ***
typeVerb movement	-2.2486	0.2141	-10.501	< 2e-16 ***

B	AIC	BIC	logLik	deviance	df.resid
	1536.7	1582.1	-760.4	1520.7	2152

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-5.5815	0.1854	0.2681	0.3807	2.6915

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	0.9085	0.9531
item	(Intercept)	0.1509	0.3885

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.1915	0.2401	4.962	6.98e-07 ***
cond3RD-PL-SHORT	1.3978	0.3181	4.394	1.11e-05 ***
cond3RD-SG-LONG	1.5412	0.3235	4.763	1.90e-06 ***
cond3RD-SG-SHORT	1.8386	0.3371	5.453	4.95e-08 ***
condV2-AUX	0.9812	0.3055	3.212	0.00132 **
condV2-MAIN	0.8651	0.3030	2.855	0.00431 **

E

AIC	BIC	logLik	deviance	df.resid
688.4	706.7	-340.2	680.4	716

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.5540	-0.4314	-0.2377	0.4837	4.0346

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	4.632680	2.15237
item	(Intercept)	0.001471	0.03836

	Number of obs: 720, groups: Response.ID, 60; item, 12
	Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.8163	0.3207	-2.546	0.0109 *
condV2-MAIN	-1.0070	0.2111	-4.770	1.84e-06 ***

I

AIC	BIC	logLik	deviance	df.resid
2055.9	2101.4	-1020.0	2039.9	2152

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.1692	-0.4816	0.2183	0.5063	8.6649

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	1.5982	1.2642
item	(Intercept)	0.2398	0.4897

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	zvalue	Pr(>|z|)
(Intercept)	2.2039	0.3360	6.558	5.44e-11 ***
grouphigh.intermediate	-0.9358	0.4456	-2.100	0.035721 *
groupadvanced	-2.4149	0.4415	-5.470	4.49e-08 ***
typeVerb movement	-1.3429	0.2712	-4.951	7.39e-07 ***

grouphigh.intermediate:				
typeVerb movement	-1.6811	0.3190	-5.270 1.36e-07 ***	
groupadvanced:				
typeVerb movement	-1.1979	0.3319	-3.609 0.000307 ***	

J	AIC	BIC	logLik	deviance	df.resid
	2055.9	2101.4	-1020.0	2039.9	2152
	Scaled residuals:				
	Min	1Q	Median	3Q	Max
	-5.1692	-0.4816	0.2183	0.5063	8.6649
	Random effects:				
	Groups	Name	Variance	Std.Dev.	
	Response.ID	(Intercept)	1.5982	1.2642	
	item	(Intercept)	0.2398	0.4897	
	Number of obs: 2160, groups: Response.ID, 60; item, 36				
	Fixed effects:				
		Estimate	Std. Error	z value	Pr(>|z|)
	(Intercept)	-0.2110	0.3188	-0.662	0.507957
	grouphigh.intermediate	1.4791	0.4347	3.403	0.000667 ***
	grouplow.intermediate	2.4149	0.4414	5.470	4.49e-08 ***
	typeWO	-2.5407	0.3114	-8.158	3.39e-16 ***
	grouphigh.intermediate:				
	typeWO	-0.4832	0.3503	-1.380	0.167723

	grouplow.intermediate: typeWO 1.1979 0.3319 3.609 0.000307 ***

K	AIC BIC logLik deviance df.resid
	
	1558.6 1615.4 -769.3 1538.6 2150
	
	Scaled residuals:
	
	   Min 1Q Median 3Q Max
	
	-5.5405 0.1911 0.2715 0.3810 2.6623
	
	Random effects:
	
	   Groups Name Variance Std.Dev.
	
	   Response.ID (Intercept) 0.7326 0.8559
	
	   item (Intercept) 0.5106 0.7145
	
	Number of obs: 2160, groups: Response.ID, 60; item, 36
	
	Fixed effects:
	
	                 Estimate Std. Error z value Pr(>|z|)
	
	(Intercept) 1.47870 0.37225 3.972 7.12e-05 ***
	
	Groupsintermediate 1.13548 0.48165 2.357 0.0184 *
	
	groupshigh.interm. 1.00854 0.42106 2.395 0.0166 *
	
	groupsadvanced 0.98314 0.42016 2.340 0.0193 *
	
	typeVerb movement -0.24806 0.39323 -0.631 0.5282
	
	groupsintermediate:
	typeVerb movement 0.06289 0.45837 0.137 0.8909

	groupshigh.intermediate: typeVerb movement -0.08521 0.38654 -0.220 0.8255 groupsadvanced: typeVerb movement 0.25689 0.39823 0.645 0.5189
L	AIC BIC logLik deviance df.resid 1558.9 1615.6 -769.4 1538.9 2150 Scaled residuals:

Min   1Q   Median   3Q   Max

-5.6395   0.1912   0.2719   0.3799   2.6622

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	0.7329	0.8561
item	(Intercept)	0.5105	0.7145

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.23066	0.43062	2.858	0.00427 **
groupsadvanced	1.24000	0.47562	2.607	0.00913 **
groupshigh.intermediate	0.98602	0.47256	2.087	0.03693 *
groupsintermediate	1.06602	0.52373	2.035	0.04181 *
typeAgreement	0.24806	0.39327	0.631	0.52821

	groupsadvanced: typeAgreement	-0.25689	0.39829	-0.645	0.51893
	groupshigh.intermediate: typeAgreement	0.01297	0.39204	0.033	0.97361
	groupsintermediate: typeAgreement	0.07464	0.44232	0.169	0.86599

M

AIC	BIC	logLik	deviance	df.resid
2058.0	2114.8	-1019.0	2038.0	2150

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.3274	-0.4831	0.2173	0.5003	8.5965

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	1.5391	1.2406
Item	(Intercept)	0.2396	0.4894

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.7578864	0.4692284	3.746	0.000179 ***
groupsadvanced	-1.9676466	0.5468822	-3.598	0.000321 ***
groupshigh.intermediate	-0.4933321	0.5505163	-0.896	0.370185
groupsintermediate	0.8307267	0.6319423	1.315	0.188657

	typeVerb movement	-1.3481876 0.3392357	-3.974	7.06e-05 ***	
	groupsadvanced:				
	typeVerb movement	-1.1894712 0.3890136	-3.058	0.002231 **	
	groupshigh.intermediate:				
	typeVerb movement	-1.6679017 0.3779678	-4.413	1.02e-05 ***	
	groupsintermediate:				
	typeVerb movement	-0.0006976 0.4168913	-0.002	0.998665	

N					
	AIC	BIC	logLik	deviance	df.resid
	2062.9	2119.7	-1021.5	2042.9	2150

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.5163	-0.4869	0.2117	0.5061	8.6819

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	1.6086	1.2683
item	(Intercept)	0.2377	0.4875

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.4100	0.4969	0.825	0.409294
groupsadvanced	-3.1606	0.6044	-5.229	1.70e-07 ***
groupshigh.intermediate	-2.1673	0.5929	-3.655	0.000257 ***

groupsintermediate	0.5794	0.6341	0.914	0.360885
typeAgreement	1.3504	0.3390	3.984	6.79e-05 ***
groupsadvanced: typeAgreement	1.1890	0.3891	3.055	0.002249 **
groupshigh.intermediate: typeAgreement	1.6802	0.3837	4.379	1.19e-05 ***
groupsintermediate: typeAgreement	0.1673	0.4036	0.414	0.678548

O

AIC	BIC	logLik	deviance	df.resid
2062.9	2119.7	-1021.5	2042.9	2150

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.5163	-0.4869	0.2117	0.5061	8.6819

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	1.6086	1.2683
item	(Intercept)	0.2377	0.4875

Number of obs: 2160, groups: Response.ID, 60; item, 36

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.7573	0.3791	-4.635	3.57e-06 ***
groupsintermediate	2.7467	0.5481	5.011	5.40e-07 ***

groupslow.intermediate 2.1673	0.5930	3.655	0.000257 ***
groupsadvanced	-0.9933	0.5086	-1.953	0.050830 .
typeAgreement	3.0306	0.3056	9.916	< 2e-16 ***
groupsintermediate: typeAgreement	1.5130	0.3751	-4.033	5.50e-05 ***
groupslow.intermediate: typeAgreement	-1.6802	0.3838	-4.378	1.20e-05 ***
groupsadvanced: typeAgreement	-0.4913	0.3561	-1.380	0.167667

P

AIC	BIC	logLik	deviance	df.resid
2108.9	2137.3	-1049.5	2098.9	2155

Scaled residuals:

Min	1Q	Median	3Q	Max
-6.2439	-0.5427	0.1907	0.5638	7.7492

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	1.351	1.162

Number of obs: 2160, groups: Response.ID, 60

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.4118	0.2978	-8.098	5.59e-16 ***
grouphigh.intermediate	1.2121	0.3925	3.088	0.00202 **

	grouplow.intermediate and intermediate	2.8291	0.4014	7.048	1.81e-12 ***
	type2Agreement	2.1680	0.1305	16.611	< 2e-16 ***

Q

AIC   BIC   logLik   deviance   df.resid

1366.1   1418.8   -673.0   1346.1   1430

Scaled residuals:

  Min   1Q   Median   3Q   Max

-3.7697 -0.4756 -0.0858   0.4811 6.8352

Random effects:

Groups	Name	Variance	Std.Dev.
Response.ID	(Intercept)	1.4631	1.2096
item	(Intercept)	0.1654	0.4068

Number of obs: 1440, groups: Response.ID, 60; item, 24

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.6615	0.5003	3.321	0.000896 ***
groupsintermediate	0.9856	0.6839	1.441	0.149531
groupshigh.intermediate	-0.2105	0.5837	-0.361	0.718308
groupsadvanced	-2.1986	0.5785	-3.801	0.000144 ***
typeVerb movement	-1.2464	0.3850	-3.237	0.001208 **
groupsintermediate: typeVerb movement	-0.1731	0.5070	-0.341	0.732821

	groupshigh.intermediate:
	typeVerb movement -1.8907 0.4444 -4.255 2.09e-05 ***
	groupsadvanced:
	typeVerb movement -0.9501 0.4441 -2.139 0.032415 *

R	AIC BIC logLik deviance df.resid
	
	1616.8 1650.9 -802.4 1604.8 2154
	
	Scaled residuals:
	
	    Min 1Q Median 3Q Max
	
	-4.2862 0.2394 0.3024 0.3983 1.3209
	
	Random effects:
	
	 Groups Name Variance Std.Dev.
	
	 Response.ID (Intercept) 0.6106 0.7814
	
	Number of obs: 2160, groups: Response.ID, 60
	
	Fixed effects:
	
	                Estimate Std. Error z value Pr(>|z|)
	
	(Intercept) 1.3107 0.3036 4.317 1.58e-05 ***
	
	groupsadvanced 0.9938 0.3685 2.697 0.00701 **
	
	groupshigh.intermediate 0.9170 0.3714 2.469 0.01355 *
	
	groupsintermediate 1.0268 0.4111 2.498 0.01249 *
	
	typeVerb movement -0.1294 0.1364 -0.948 0.34304

Appendix 14: T-tests

A	Welch Two Sample t-test
	t = 16.197, df = 1429.3, p-value < 2.2e-16
	95 percent confidence interval:
	0.3051710 0.3892734
B	Paired t-test
	t = 4.6853, df = 359, p-value = 3.973e-06
	95 percent confidence interval:
	0.07414527 0.18141029
C	Paired t-test
	t = -5.3331, df = 119, p-value = 4.665e-07
	95 percent confidence interval:
	-0.2856847 -0.1309819
D	Paired t-test
	t = 3.0608, df = 119, p-value = 0.002729
	95 percent confidence interval:
	0.05001834 0.23331500
E	Paired t-test
	t = 4.4229, df = 119, p-value = 2.168e-05
	95 percent confidence interval:
	0.1104606 0.2895394

F	Paired t-test t = -5.4827, df = 119, p-value = 2.386e-07 95 percent confidence interval: -0.3743165 -0.1756835
G	Paired t-test t = -2.0257, df = 119, p-value = 0.04503 95 percent confidence interval: -0.0659163552 -0.0007503114
H	Paired t-test t = -2.9155, df = 119, p-value = 0.004245 95 percent confidence interval: -0.11194458 -0.02138876
I	Paired t-test t = 2.7505, df = 239, p-value = 0.006406 95 percent confidence interval: 0.03192523 0.19307477
J	Paired t-test t = 2.9654, df = 119, p-value = 0.003654 95 percent confidence interval: 0.0470699 0.2362634
K	Paired t-test

	t = 3.0491, df = 119, p-value = 0.002829 95 percent confidence interval: 0.05843176 0.27490157
L	Paired t-test t = 2.756, df = 119, p-value = 0.006772 95 percent confidence interval: 0.02815411 0.17184589
M	Paired t-test t = 2.9454, df = 719, p-value = 0.003329 95 percent confidence interval: 0.02361986 0.11804680
N	Welch Two Sample t-test t = 4.0571, df = 177.55, p-value = 7.432e-05 95 percent confidence interval: 0.1141323 0.3303122
O	Welch Two Sample t-test t = 3.9014, df = 192.28, p-value = 0.0001322 95 percent confidence interval: 0.08240748 0.25092585
P	Welch Two Sample t-test t = 14.186, df = 515.25, p-value < 2.2e-16

| | 95 percent confidence interval: |
| | 0.4163959 0.5502708 |

Appendix 15: T-tests when long-distance agreement is excluded

A	Welch Two Sample t-test
	t = 3.6449, df = 1364.2, p-value = 0.0002775
	95 percent confidence interval:
	0.02822115 0.09400108
B	Welch Two Sample t-test
	t = 13.533, df = 1438, p-value < 2.2e-16
	95 percent confidence interval:
	0.2873918 0.3848305
C	Welch Two Sample t-test
	t = 3.2881, df = 207.01, p-value = 0.001185
	95 percent confidence interval:
	0.08156688 0.32584053
D	Welch Two Sample t-test
	t = 4.8764, df = 238.39, p-value = 1.975e-06
	95 percent confidence interval:
	0.1324488 0.3119957
E	Welch Two Sample t-test
	t = 12.459, df = 452.7, p-value < 2.2e-16
	95 percent confidence interval:
	0.4248287 0.5839433

F	Welch Two Sample t-test
	t = 8.1756, df = 395.75, p-value = 4.035e-15
	95 percent confidence interval:
	0.2310237 0.3773096

Appendix 16: T-tests of ungrammatical and grammatical sentences

A	Welch Two Sample t-test t = 11.412, df = 84.29, p-value < 2.2e-16 95 percent confidence interval: 0.4323718 0.6148504
B	Welch Two Sample t-test t = 4.485, df = 81.884, p-value = 2.354e-05 95 percent confidence interval: 0.1058787 0.2746769
C	Welch Two Sample t-test t = 1.118, df = 15.99, p-value = 0.2801 95 percent confidence interval: -0.1244718 0.4022496
D	Welch Two Sample t-test t = 2.34, df = 16.753, p-value = 0.03194 95 percent confidence interval: 0.01757953 0.34353158
E	Welch Two Sample t-test t = 9.8307, df = 26.289, p-value = 2.696e-10 95 percent confidence interval: 0.5030599 0.7688699

F	Welch Two Sample t-test t = 20.691, df = 33.558, p-value < 2.2e-16 95 percent confidence interval: 0.7176306 0.8740361
G	Welch Two Sample t-test t = -0.58076, df = 14.798, p-value = 0.5701 95 percent confidence interval: -0.2380519 0.1362001
H	Welch Two Sample t-test t = 0.083465, df = 21.388, p-value = 0.9343 95 percent confidence interval: -0.08294592 0.08989036
I	Welch Two Sample t-test t = 2.362, df = 23.757, p-value = 0.02672 95 percent confidence interval: 0.02233484 0.33292832
J	Welch Two Sample t-test t = 5.9341, df = 22.661, p-value = 5.044e-06 95 percent confidence interval: 0.2753636 0.5704697

Appendix 17: T-tests of ungrammatical and grammatical sentences when long-distance agreement is excluded

A	Welch Two Sample t-test
	t = 5.3206, df = 73.949, p-value = 1.063e-06
	95 percent confidence interval:
	0.1555058 0.3417164
B	Welch Two Sample t-test
	t = 0, df = 15.941, p-value = 1
	95 percent confidence interval:
	-0.2066311 0.2066311
C	Welch Two Sample t-test
	t = 0, df = 20.946, p-value = 1
	95 percent confidence interval:
	-0.06370869 0.06370869
D	Welch Two Sample t-test
	t = 2.3228, df = 22.211, p-value = 0.02974
	95 percent confidence interval:
	0.01888553 0.33199166
E	Welch Two Sample t-test
	t = 7.2443, df = 20.026, p-value = 5.171e-07
	95 percent confidence interval:

	0.4124125 0.7459208

Appendix 18: Mean scores on the Likert scale (0-1) for each sentence

	Grammatical	Ungrammatical
The.students.sits.in.the.park.after.school		0.70000000
The.students.sit.in.the.park.after.school	0.9333333	
The.brothers.attends.football.practice.every.day		0.60000000
The.brothers.attend.football.practice.every.day	0.8166667	
The.cats.plays.with.the.yellow.and.green.ball		0.73333333
The.cats.play.with.the.yellow.and.green.ball	0.9000000	
The.kids.likes.to.play.in.the.park.every.weekend		0.75000000
The.kids.like.to.play.in.the.park.every.weekend	0.9500000	
The.sisters.loves.to.run.in.the.forest		0.73333333
The.sisters.love.to.run.in.the.forest	0.8833333	
The.teachers.gives.their.students.a.lot.homework		0.9166667
The.teachers.give.their.students.a.lot.of.homework.	0.9166667	
The.student.love.to.read.books.about.football		0.76666667
The.student.loves.to.read.books.about.football	0.9333333	
The.teacher.eat.fish.for.dinner.every.Friday		0.65000000
The.teacher.eats.fish.for.dinner.every.Friday	0.9166667	
The.boy.like.to.go.swimming.in.the.ocean		0.60000000
The.boy.likes.to.go.swimming.in.the.ocean	0.9000000	
The.brown.dog.play.with.the.yellow.footbal		0.63333333
The.brown.dog.plays.with.the.yellow.football	0.9666667	
The.girl.drink.a.lot.of.water.every.day		0.66666667
The.girl.drinks.a.lot.of.water.every.day	0.9666667	
The.girl.drive.to.work.every.Wednesday.morning		0.56666667
The.girl.drives.to.work.every.Wednesday.morning	0.9000000	

The.book.about.fast.cars.make.the.girl.happy		0.66666667
The.book.about.fast.cars.makes.the.girl.happy	0.9500000	
The.boy.with.blue.eyes.seem.very.happy		0.73333333
The.boy.with.blue.eyes.seems.very.happy	0.9500000	
The.boy.with.broken.arms.try.to.read.a.book		0.50000000
The.boy.with.broken.arms.tries.to.read.a.book	0.9500000	
The.girl.with.golden.earrings.take.the.bus.to.school		0.68333333
The.girl.with.golden.earrings.takes.the.bus.to.school	0.9166667	
The.house.with.yellow.and.white.doors.look.nice		0.81666667
The.house.with.yellow.and.white.doors.looks.nice	0.9166667	
The.teacher.with.black.shoes.walk.to.work.every.day		0.65000000
The.teacher.with.black.shoes.walks.to.work.every.day	0.8833333	
The.boys.in.the.black.car.looks.very.scary.		0.75000000
The.boys.in.the.black.car.look.very.scary	0.8333333	
The.cats.with.long.white.fur.drinks.milk.every.day		0.68333333
The.cats.with.long.white.fur.drink.milk.every.day	0.7500000	
The.girls.with.blonde.hair.likes.to.sit.in.the.park		0.78333333
The.girls.with.blonde.hair.like.to.sit.in.the.park	0.85000000	
The.kids.with.the.red.bike.plays.in.the.garden		0.60000000
The.kids.with.the.red.bike.play.in.the.garden	0.6166667	
The.parents.with.the.nice.car.talks.to.their.kids		0.6833333
The.parents.with.the.nice.car.talk.to.their.kids	0.8000000	
Those.tourists.with.the.heavy.suitcase.seems.tired		0.55000000
Those.tourists.with.the.heavy.suitcase.seem.tired	0.7000000	
Yesterday.went.the.teacher.to.the.shop		0.21666667
Yesterday.the.teacher.went.to.the.shop	0.8833333	

Yesterday.celebrated.the.girl.her.birthday		0.26666667
Yesterday.the.girl.celebrated.her.birthday	0.9333333	
Every.Monday.played.the.girls.football.in.the.park		0.28333333
Every.Monday.the.girls.played.football.in.the.park	0.8666667	
Hopefully.passed.the.student.her.exam		0.31666667
Hopefully.the.student.passed.her.exam	0.9000000	
Last.weekend.drank.the.students.lots.of.coffee		0.26666667
Last.weekend.the.students.drank.lots.of.coffee	0.8333333	
Sometimes.ate.the.dog.the.family's.shoes		0.25000000
Sometimes.the.dog.ate.the.family's.shoes	0.6666667	
Tomorrow.might.the.football.player.score.a.goal		0.30000000
Tomorrow.the.football.player.might.score.a.goal	0.7833333	
Tomorrow.must.the.sisters.clean.their.room		0.36666667
Tomorrow.the.sisters.must.clean.their.room	0.8666667	
Very.soon.will.the.band.play.their.favourite.song		0.46666667
Very.soon.the.band.will.play.their.favourite.song	0.9000000	
Every.day.should.the.students.bring.their.books.to.school		0.45000000
Every.day.the.students.should.bring.their.books.to.school	0.7666667	
Every.year.must.the.students.pass.their.exams		0.36666667
Every.year.the.students.must.pass.their.exams	0.9333333	
In.December.will.the.kids.celebrate.Christmas		0.41666667
In.December.the.kids.will.celebrate.Christmas	0.9166667	
Dogs.the.barked.at.the.postman.with.the.red.jacket		0.11666667
Children.the.played.the.guitar.very.well		0.06666667
Boy.the.with.the.blue.jacket.ate.an.apple		0.20000000
Father.my.went.to.the.shop.every.Monday.		0.10000000

Many.people.like.to.wear.shorts.in.summer.the.		0.16666667
Monkey.the.loved.to.eat.bananas.for.dinner		0.08333333
Mother.my.read.a.book.about.flowers		0.08333333
Garden.my.had.many.different.types.of.flowers.		0.01666667
Student.the.often.drank.coffee.in.the.morning		0.16666667
The.black.shoes.are.too.small.for.woman.the		0.06666667
The.man.wrote.a.book.about.mother.his		0.10000000
The.man.enjoyed.to.drive.car.his.to.work		0.18333333
Thief.the.stole.many.paintings.yesterday		0.13333333