

This is a post-print version reflecting changes made in the peer review and editing process, but it is not the publisher's PDF. The paper appeared as:

Jensen, Isabel, Björn Lundquist, Roumyana Slabakova & Marit Westergaard. 2019. The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research*.

The publisher should be contacted for permission to re-use or reprint the material in any form.

The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English

Isabel Nadine Jensen

UiT The Arctic University of Norway, Norway

Roumyana Slabakova

University of Southampton, UK and UiT The Arctic University of Norway, Norway

Marit Westergaard

UiT The Arctic University of Norway, Norway and NTNU Norwegian University of Science and Technology, Norway

Björn Lundquist

UiT The Arctic University of Norway, Norway

Abstract

The Bottleneck Hypothesis (Slabakova, 2008, 2013) proposes that acquiring properties of the functional morphology is the most challenging part of learning a second language. In the experiment presented here, the predictions of this hypothesis are tested in the L2 English of Norwegian native speakers. Two constructions are investigated that do not match in English and Norwegian: One involving functional morphology, Subject–Verb (SV) agreement, which is obligatory in the L2 but non-existent in the L1, and one involving syntax, Verb-Second (V2) word order, which is obligatory in the L1, but restricted to specific contexts in the L2. The results of an acceptability judgement task indicate that the participants struggled more with identifying ungrammatical SV agreement than ungrammatical word order. We conclude that the findings lend tentative support to the Bottleneck Hypothesis.

Key words

The Bottleneck Hypothesis, functional morphology, core syntax, L2 English, L1 Norwegian, word order, verb-second, agreement

Introduction

The Bottleneck Hypothesis (BH, Slabakova, 2008, 2013) addresses the question of why certain properties of language are harder to acquire than others in an additional language (L2). Based on a comparison of linguistic modules and their learnability, the hypothesis holds that functional morphology, typically hosted in functional categories (FC) is the bottleneck of L2 acquisition. It is the most challenging part of acquisition because this is where differences among languages are located, as per the Borer–Chomsky Conjecture (Borer, 1984; Chomsky, 1995; Baker, 2008). This conjecture relies on a division of labor between the formal grammatical features hosted by the FCs and their semantic and syntactic reflexes, or consequences, such as calculating a certain grammatical meaning or executing some movement operation. In comparison, core syntax and semantic operations are putatively universal, in the sense that movement and semantic operations are executed in the same way and obey similar restrictions in languages of the world.

Since recognizing and supplying the overt exponent of a FC and knowing all the features reflected in that FC are distinct processes, there are two possible views with respect to which process precedes and possibly even triggers the other. One view, dubbed *morphology-before-syntax* by White (2003: 184–187), suggests that learning the functional morphology drives acquisition of syntax (Eubank, 1994; Vainikka and Young Scholten, 1994; Clahsen and Hong, 1995). The opposite view, *syntax-before-morphology* (Haznedar and Schwartz, 1997; Lardiere, 1998a,b; White, 2003), argues that syntactic knowledge is available in learner production much earlier than correct suppliance of the overt functional morphology. We adopt this latter view without further argumentation, but see the above works for extended discussion.

In addition, FCs typically comprise *bundles* of semantic, syntactic and morpho-phonological features, which affect the acceptability and the meaning of complete sentences. It is possible that not all the features in a feature bundle are acquired at the same time. To wit, Feature Reassembly (Lardiere, 2009) suggests that mismatches in feature realization are responsible for most of the L2 acquisition challenges. In sum, by assumption, FCs comprise bundles of formal features, which regulate syntactic behaviour and bring forward grammatical meanings; the bundle may have an overt exponent (an affix) or not.

In order to test the BH, several experimental requirements should be obeyed. The rationale is as follows: If we find that a certain property A is acquired faster and more accurately than another property B, this may be due to a number of factors or dimensions along which the

two properties differ. Language learnability is concerned with the nature of the input that is needed in order to acquire a linguistic feature and the extent to which properties of the input determine the process. Rate of acquisition, a proxy for ease or difficulty of acquisition, can be established by monitoring accuracy scores at various levels of proficiency.

First of all, a construction under investigation can be *attested* or *exemplified* in the input available to learners or not. If a property is available to observe in the input, it is learnable from positive evidence. If a property is not attested in the input, negative evidence may be needed for its unavailability to be acquired. Direct negative evidence is not the same thing as absence of positive evidence; it refers to explicit instruction on the ungrammaticality of a certain string. Let us exemplify with the prohibition of a certain order of adjectives and nouns, the so-called *a*-adjectives (Boyd and Goldberg, 2011; Yang, 2015). While strings as in (2) with the adjectives used predicatively are attested in English input, a learner cannot deduce from this that the pre-nominal A–N order in (1) is unavailable.

- (1) *an/the awake child
- (2) The child is awake.

Properties that are exemplified with positive evidence have been argued to be easier to learn than properties whose unavailability should be inferred from the absence of positive evidence for it (Trahey and White, 1993; Mazurkewich and White, 1984; Inagaki, 2006; Gabriele, 2009; Montrul and Yoon, 2009: 308). While direct negative evidence is arguably not taken into account by children learning their native language (Brown and Hanlon, 1970; Braine, 1971; Bowerman, 1988; Marcus, 1993), the situation is more complicated in the second language (White 1989, 1991, 2003).

If a construction is attested in the input, then construction *frequency* becomes a second possible factor that could influence rate of acquisition. This is because a more frequent construction provides learners with more positive linguistic evidence, which can lead to earlier and more accurate acquisition (Ellis, 2002). However, as argued by Yang (2002), mere construction frequency is not enough. The input has to exemplify beyond any doubt that a particular grammatical option is chosen in the language being acquired. For example, as argued by Westergaard (2009c), when children acquire a verb-second (V2) language such as German or

Norwegian, sentences with SVO word order do not unequivocally show the learner that the language they are acquiring is V2. Only sentences with verb movement across the subject resulting in XVSO order point to a V2 option. We discuss this in the next section.

Another possible predictor of difficulty is *perceptual salience*, which “refers to how easy it is to hear or perceive a given structure” (Goldschneider and DeKeyser, 2001: 47). The morpheme order studies (Dulay and Burt, 1973, 1974; Bailey, Madden and Krashen, 1974) identified salience as a significant predictor of functor difficulty, as measured by later acquisition. According to Goldschneider and DeKeyser (2001), functor salience is composed of three further variables: how many phones (sounds) it has, whether it is syllabic (contains a vowel), and how sonorous it is. For example, *-ing* is considered to be more salient than *-s*.

The fourth dimension of acquisition variability is naturalistic exposure versus the need for *instruction*. Scholars of second language acquisition have long debated whether primary linguistic data in the form of comprehensible input that language learners are exposed to is sufficient for acquisition of any grammatical property or whether explicit instruction is needed. This debate is related to “the interface issue” (e.g., Krashen, 1985), referring to the interface between explicit and implicit knowledge.¹ In short, explicit instruction and drilling in a language classroom could be a factor in second language acquisition, particularly if one espouses the position that instruction can affect internalized grammatical competence.

In the next section, we describe the two properties at hand and discuss how they stack up with respect to the acquisition difficulty factors we have examined above.

The two properties under investigation

In this study, we use Subject–Verb (SV) agreement to test learners’ knowledge of functional morphology, whereas word order in declaratives (V2 vs. non-V2) is used to test knowledge of core syntax. These two constructions exhibit a grammatical contrast in English and Norwegian. SV agreement (3rd sg. *-s*) is obligatory in English but not attested in Norwegian, while (X)SVO

¹ Three positions—no-interface, strong-interface and weak-interface—have been articulated. The no-interface position would posit that no explicitly taught knowledge can become internalized (Krashen, 1985; Schwartz, 1993). The strong-interface position argues that language learning being just like any other kind of learning, practice would turn explicit instruction into implicit, proceduralized knowledge (DeKeyser, 2007). Finally, the weak-interface position acknowledges the primacy of naturalistic exposure, but allows space for developmental factors based on instruction (e.g., Ellis, 2015).

(non-V2) word order, which is the typical pattern in English, is generally not allowed in Norwegian, which requires V2 word order in (virtually all) declaratives. The constructions have been chosen on the basis of findings in previous studies, showing that they pose challenges to L2 learners. For example, Westergaard (2003) shows that acquisition of English non-V2 word order is difficult for native speakers of Norwegian at an early stage. Haznedar (2001) and Lardiere (1998a,b) document that SV agreement is problematic for child as well as end-state learners of English L2 (see also Ionin and Wexler, 2002). These previous findings indicate that both verb placement and SV agreement should be challenging for Norwegian learners of English as an L2. However, which one of them is more challenging has not yet been investigated. In the following sections, the constructions are discussed in more detail, including the relevant uninterpretable features and their exponents.

Word order in English and Norwegian

English is an SVO language, while Norwegian displays V2 syntax, in that the finite verb is generally required to occur in the second position of main clauses. There are numerous exceptions to V2 in Norwegian, especially with respect to *wh*-interrogatives across various dialects (see e.g., Westergaard, 2009b; Westergaard, Vangsnes and Lohndal, 2017). However, in main clause declaratives, the finite verb typically appears in second position (with only a few exceptions related to specific adverbs; see e.g., Westergaard, 2008; Bentzen, 2014). This is seen in both non-subject-initial and subject-initial sentences with adverbs, as illustrated in (3) and (4), respectively.

- (3) I går drakk studentene vin. (XVSO)
 Yesterday drank students.DEF wine
 ‘Yesterday the students drank wine.’
- (4) Studentene drakk ofte vin. (SVXO)
 students.DEF drank often wine
 ‘The students often drank wine.’

V2 word order in Norwegian is typically argued to be caused by V-to-C movement, triggered by a feature in C which attracts the finite verb (see e.g., Vikner, 1995). Alternatively,

the process is triggered by a syntactic (micro-)cue, an abstract piece of structure in speakers' L-language grammars (Westergaard, 2009b), illustrated in (5), where the verb appears in second position, immediately following an XP, in declarative main clauses (DeclPs). Under this approach, V2 is the result of verb movement to different syntactic heads in different contexts, e.g., different clause types.

(5) ${}_{\text{DeclP}}[\text{XP}{}_{\text{Decl}^*} \text{V}]$

English does not exhibit verb movement of finite lexical verbs (Pollock, 1989). Thus, the learning task for Norwegian L1 speakers is to *unlearn* this V2 cue. As Westergaard (2003) shows, Norwegian L1 speakers often transfer this structure to declaratives in L2 English.

Functional morphology in English and Norwegian

Whereas Norwegian has no overt SV agreement morphology, as illustrated in (6), English marks present tense verbs with the suffix *-s* when the subject is 3rd person singular. The agreement marker is not only frequent in English input (see Appendices A and B), but also subject to intensive instruction from an early stage in L2 English classrooms (see Appendix D).

- (6) a. Lars og Mari snakker norsk.
 Lars and Mari speak Norwegian
 ‘Lars and Mari speak Norwegian.’
- b. Lars snakker norsk.
 Lars speaks Norwegian
 ‘Lars speaks Norwegian.’

There is a significant bundle of abstract morphosyntactic features to which *-s* is the overt exponent: [tense], [aspect] and [agreement]. In addition, features that ensure that the subject in English is overt, that it is in the Nominative case, and that the verb stays in the VP (i.e., there is no verb movement), are also captured in the FC of TP. Importantly, all of these features have to be acquired for this FC to be completely native-like. We do not test knowledge of all these features in L2 English, because such experiments have already been done, see White (2003: 187–

193 for review). In a nutshell, child and adult L2 learners are much more accurate, even at ceiling, with the core syntax features of verb movement and subject expression, while at the same time the accuracy of the morphological marker is between 4.5% and 46.5%. Such findings support the syntax-before-morphology view we introduced above (White, 2003).

It is not the case that the same features are missing in Norwegian; to the contrary, they are also reflected in the same functional category. However, whatever syntactic and semantic feature values it comprises, the Norwegian bundle does not have an overt morphological reflex. According to the Micro-cue Model of Westergaard (2009a), the micro-cue for English SV agreement could be illustrated as in (7).

(7) ${}_{TP} [XP]_{3sg} [V-s]$

While English-learning children typically acquire this morpheme conservatively (i.e., only provide it where necessary), both under- and overuse has been found in L2 acquisition (e.g., Dröschel, 2011; Vettorel, 2014; Garshol, 2018).

We now discuss how these two properties stack up with respect to the dimensions affecting difficulty, as discussed in the introduction. The first dimension along which the two properties differ is learnability. As mentioned in the introduction, previous studies in second language acquisition (Trahey and White, 1993; Mazurkewich and White, 1984; Inagaki, 2006; Gabriele, 2009; Montrul and Yoon, 2009: 308) have established that it is easier for learners to add new features for which the L2 input provides positive evidence than to subtract or unlearn L1 features from the L2 grammar. This is because negative evidence would presumably be required to inform learners of the features' unacceptability. Trahey and White (1993) called this unlearning process "preemption" and argued that it is particularly difficult, since learners cannot rely on positive evidence to determine which features or constructions are ruled out by the L2. In our case, English provides ample evidence of SV agreement (see Appendix A, appendix B and the next factor). On the other hand, the ungrammaticality of English declaratives with V2 word order cannot be positively established. Since we know from previous studies that such structures are transferred from the L1 into the L2 (e.g., Westergaard, 2003), negative evidence is needed for the preemption of these V2 word orders.

In a traditional (macro-)parameter approach to V2 word order, an L1 Norwegian learner

of L2 English would simply reset the parameter on exposure to non-V2 input, and no preemption would be necessary. However, given the micro-variation that exists with respect to verb movement, both in Norwegian and in English (see e.g., Roeper, 1999; Westergaard, 2007, 2009b) and children's early acquisition of this variation (e.g., Westergaard, 2009a, 2009c), V2 cannot be a parameter. Furthermore, the most common word order in English, SVO, is in fact ambiguous and can be interpreted as either SVO or V2. According to Westergaard (2003), in order for an L1 Norwegian learner of English to acquire non-V2, she must be exposed to sentences with the structure XSV (e.g., *Yesterday John went to the cinema*), which are ungrammatical in Norwegian. However, this input would only be evidence for the learner that the L2 is *different* from the L1 in that it allows such structures. It would not be enough to tell the learner that the L1 structure (i.e., V2 word order) is ungrammatical, since there is abundant evidence that there is variation with respect to verb movement both in the L1 and the L2. That is to say, both word orders could in principle be grammatical in the L2. In this case, the learner would only be exposed to positive evidence for one of them, the non-V2 word order that is predominant in English. This means that the transferred V2 word order would still need to be preempted. Since this evidence is not directly observable in the available input, non-V2 word order should be a relatively difficult property to acquire for L1 Norwegian learners of L2 English.²

This brings us to the second factor of comparison, input frequency (Ellis, 2002). There should be no doubt that the *-s* morpheme marking SV agreement is very frequent. It has to appear in all simple present tense sentences that have a third person subject. In the 520-million-word Corpus of Contemporary American English (COCA, Corpus.byu.edu, 2018a), the third-person singular *-s* occurs a total of 6,198,523 times. That is, it appears in as much as 37.5% of all relevant contexts, i.e., present tense lexical verbs (see Appendix A). When only spoken sources are considered, the *-s* occurs 944,638 times, i.e., in 23.4% of all occurrences of simple present tense lexical verbs (see Appendix B). This very high frequency means that every learner of English must have substantial exposure to the marker and any acquisition delay could not be due to lack of input.

² Yang (2015) shows that, in certain specific cases, negative constraints (i.e., absence of a certain word order) can be learned from indirect positive evidence. However, it is not clear how the V2 word order could be preempted in the framework of micro-cues.

With respect to the syntactic property, the only cue to unlearn V2 in non-subject-initial declaratives would be declaratives with non-V2 in English, i.e. XSV word order. Westergaard (2003) argues that, while non-subject-initial declaratives are relatively frequent in V2 languages, approximately 30% (see also Lightfoot, 1999; Westergaard 2008), XSV word order is rare in English. Some numerical evidence for this comes from Yang (2001), who has investigated modern English in the Penn Treebank, finding that only about 10% of all declaratives have the verb in the third position (SXV or XSV), thus indicating that English is not a V2 language. But SXV word order cannot be a cue to unlearn V2 in non-subject-initial declaratives, as word order in subject-initial and non-subject-initial declaratives are arguably not directly related (a language may have short verb movement across an adverb, but no movement across the subject, e.g., French), as argued in e.g., Westergaard (2009a) and Westergaard, Lohndal and Alexiadou (2016). For our current purposes, it is therefore prudent to identify the proportion of XSV. Charles Yang has kindly carried out a search in the Pearl Sprouse Corpus, which offers Penn TreeBank-style parses for a number of corpora of child-directed speech in CHILDES (see Appendix C). The finding shows that XSV word order makes up only 6.8% (3,114/36,117) of all declaratives. These findings show that, while the word order evidence for non-V2 is relatively sparse in English (6.8%), the frequency of the *-s* agreement marker is considerably higher (37.5%).

Thirdly, since grammatical functors are the most frequent elements of a language, they tend to become abbreviated and phonologically fused with the surrounding material (Zipf, 1949; Bybee, 2000). They are low in psychophysical salience, compared to verbs and nouns. Furthermore, the agreement marker *-s* does not contain a vowel, does not constitute a syllable, and is very often part of consonant clusters, so it is even less salient. Although *-s* has three allomorphs [s], [z] and [əz], one of which is syllabic, this variability in exponence likely makes it harder to acquire in comparison with one-allomorph affixes. However, salience cannot be the whole explanation to morpheme acquisition order: For example, plural *-s* is acquired considerably faster than agreement *-s* by English-learning children (Brown, 1973; in fact, the plural is the second fastest morpheme acquired by children, after *-ing*).

Given this, can we even compare word order and functional morphology head-to-head with respect to salience? While we are not aware of any such direct comparison, Cintrón-Valentín and Ellis (2016) compare prepositional phrases, temporal adverbs, and lexical linguistic

cues to verb agreement, and argue that the former are more salient since they are longer and stressed in the speech stream, hence more noticeable.³ Assuming this perspective, we would like to conclude that agreement *-s* is less salient than word order.

Finally, while agreement is explained and drilled in Norwegian language classrooms, word order is usually not explicitly taught. To illustrate this, we examined the learning material used in the classrooms of the current study's participants. The material included textbooks and workbooks. In addition, we examined NDLA (Norwegian Digital Learning Area), an online compilation of educational resources designed for students in the Norwegian equivalent to upper secondary school (11th to 13th grade).⁴ NDLA's English subject pages have a total of 1644 tasks and 750 pieces of reading material (NDLA, 2018b). Our investigation showed that none of the textbooks taught or addressed word order explicitly (see Appendix D). In contrast, agreement (typically referred to as concord) was described in all of the textbooks, although to different degrees, as seen in Appendix D. NDLA describes word order in one out of 750 pieces of reading material, while agreement, or concord, is addressed ten times.

Why might instruction be an important factor in our learning situation? Two reasons can be considered. First, explicit instruction draws the attention of learners to the linguistic feature. Secondly, such instruction provides concentrated evidence for the feature. To cite just one recent study, Cintrón-Valentín and Ellis (2016) showed that the low salience of morphosyntactic cues can be overcome by form-focused instruction, which leads learners to attend to cues that might have otherwise been ignored.

³ These authors found that L2 learners were less sensitive to morphological cues than to adverbial cues in an artificial language experiment on temporal reference. They also established that in learning English morphological cues, learners from rich-morphology languages have an advantage over learners whose native languages do not have verbal morphology (e.g., Chinese, which has no tense marking).

⁴ This online learning area is frequently used by schools across Norway, with 100 million page views since 2009 and 28.8 million page views in 2016 (NDLA, 2016: 11; NDLA, 2018a).

Table 1: Factors affecting the ease or difficulty of the two constructions in L1 Norwegian-L2 English acquisition. Bold font indicates a prediction of easier acquisition; italics stand for relatively more difficult acquisition.

	Grammatical module	Possible L1 Transfer	Learnability	Frequency of evidence	Saliency	Instruction
SV agreement	<i>Functional morphology</i>	No overt affix to transfer	Learning from positive evidence	Very frequent	<i>Low saliency</i>	Taught in language classrooms
Lack of V2	Core syntax	Yes, V2 can transfer and has to be preempted	<i>Unlearning by the presence of XSV plus negative evidence</i>	<i>Signature XSV less frequent</i>	Higher saliency	<i>Not instructed regularly</i>

Table 1 summarizes the factors affecting difficulty of acquisition. *A priori*, the BH predicts that SV agreement would be more difficult to acquire (see the second column of Table 1). However, it is possible that the other factors listed in the table go in consort with or work against the predictions of the BH. On surveying columns 4–7, it can be established that not all predictions based on factors affecting difficulty go in the same direction. The preemption of V2 word order depends on exposure to non-V2 in non-subject-initial declaratives (XSV), while agreement simply has to be learned. In addition, frequency and the non-availability of instruction suggest that acquiring the lack of V2 in English would be more difficult for learners. On the other hand, the low saliency of the functional affix may work against its easier acquisition. Thus, if agreement turns out to be more difficult than non-V2, it will be *despite* the factors of learning vs unlearning based on positive evidence, higher frequency, and explicit instruction. In the next section, we report on a study that tests these predictions for ease and difficulty of acquisition and checks them against the BH.

The Present study

Research questions and predictions

This is the first experimental study designed to directly test the BH. The factors listed in Table 1 above lead to the following predictions: If frequency, the additive nature of the property and instruction are crucial factors in acquisition, we expect Norwegian learners at the same level of proficiency to make fewer errors with SV agreement than with non-V2 syntax. If, on the other hand, the BH is on the right track and functional morphology is indeed the bottleneck of acquisition, we expect the situation to be reversed. Furthermore, knowledge of core syntax would improve faster than suppliance of functional morphology as the speakers become more advanced. We address these questions by comparing performance across general proficiency and across properties (see research questions 1 and 2 below).

We also acknowledge that the BH does not preclude property-internal variation based on concrete L1–L2 contrasts and processing considerations. Therefore, within the two main conditions, SV agreement and non-V2 word order, we include the following additional variables in our research design. First, we use different verb types in the non-subject-initial declarative clauses, auxiliaries and lexical verbs. We predict that non-V2 syntax should be more challenging when the test sentences contain auxiliary verbs. This expectation is based on the observation that auxiliary verbs move in English (e.g., in questions, see 8a and 9a), whereas lexical verbs never do (see 8b, 9b). Instead *do*-support is triggered in questions, see (9c).

- (8) a. Peter has always liked coffee.
b. *Peter liked always coffee.
- (9) a. What is Peter reading?
b. *What reads Peter?
c. What does Peter read?

In other words, lexical verbs have a more consistent behavior in English than auxiliaries. As a result, more variability is expected with sentences containing auxiliaries, as learners may be more inclined to move these verbs, even in contexts where there should be no verb movement, e.g., in non-subject-initial declaratives (see example (11)).

This contention is based on previous findings in L1 and L2 acquisition. Results from

studies on L1 (Håkansson and Collberg, 1994) and L2 (e.g. Colliander, 1993) acquisition of Swedish suggest that the target word order negation–modal is acquired later than negation–lexical verb in Swedish subordinate clauses. The target word order in Swedish is exemplified in (10) (Håkansson and Collberg, 1994: 96). Westergaard (2003) found that Norwegian learners at early stages of L2 English (7- to 12-year-olds) move auxiliary verbs more frequently compared to lexical verbs (Westergaard, 2003: 94). That is, the learners made more errors with English word order when the sentences contained an auxiliary, such as example (11), than when they contained lexical verbs only. In contrast to these studies, Waldmann (2008: 229) found no significant correlation between verb type and L1 Swedish children’s word order.

- (10) Vi vet [att barn **inte kan** tala på samma sätt som vuxna].
We know that children not can speak in the same way as adults
‘We know that children cannot speak in the same way as adults.’
- (11) *There will we eat a Big Mac and lots of French fries.

Nevertheless, according to the BH, both syntactic constructions are expected to cause fewer problems for the learners, in comparison to agreement.

Secondly, we look more closely at the type of subject that gives rise to problems in agreement. In order to check that, we manipulated the number of the subject NP in all subject-initial declaratives by using plural as well as 3rd person singular subjects. Previous studies on L2 subject–verb agreement have found both over- and underproduction of the morpheme –s. According to most studies, dropping the –s where it is needed is more common than oversupplying –s where it is not needed (see e.g., Dröschel, 2011; Vettorel, 2014). In contrast, a recent corpus study of agreement errors in L2 English by native speakers of Norwegian, Garshol (2018), showed that overproduction of the –s was more common than erroneous omissions. The written corpus examined in Garshol’s study consisted of 430,000 words produced by 15–16-year-old Norwegian learners of English (N=185) over a one-year period. Garshol reports that 45% of the total agreement errors are omissions of the –s, while 55% are oversuppliance. Based on this recent finding, in the current experiment we check 3rd person singular as well as plural agreement, and expect the learners to overuse the –s as well as drop it.

Lastly, we varied the distance between the subject NP and the agreeing verb in the

subject-initial declaratives by testing both long-distance and local agreement. The distance was increased by adding a prepositional phrase between the subject and the verb. In order to investigate whether the speakers rely on the subject NP or the closer NP within the PP, the number feature always differs on these two NPs (see examples 14 and 15). Incorrect agreement with the NP closest to the verb is called an *attraction error* and is argued to be triggered by the misleading number feature on that NP (Bock and Miller, 1991; see Cunnings, 2017 for an updated approach). A large literature on attraction errors in native speakers suggests that more problems with long-distance agreement than with local agreement can be expected.

Just as native speakers, L2 speakers also demonstrate reduced sensitivity to agreement violations. Research on real-time language processing by non-native speakers has documented that agreement is more challenging when the distance between the agreeing elements in a construction is increased; see e.g., Foote, 2011, who offers two explanations: First, it may be caused by a working memory deficiency (Myles, 1995; Keating, 2005), more specifically, the difficulty of maintaining a feature of the subject (e.g., number or gender) in working memory across the element that separates the subject and the agreeing verb. According to Foote (2011), this working memory limitation also affects native speakers. The second explanation was originally offered by Almor et al. (2001), who argue that the distance effect could be caused by speakers' expectations about where the verb should occur in a sentence. That is, speakers expect the verb to immediately follow the subject, because local agreement is more frequent. Consequently, speakers are generally more sensitive to agreement violations in sentences where the subject and the verb are adjacent. A recent explanation, Cunnings (2017), invokes not just distance but intervention: the author argues that the long-distance dependencies contain an intervening distractor item bearing features which allow this item to be retrieved, thus slowing the processing and increasing error rates.

As suggested by all three explanations, however, problems with long-distance agreement do not necessarily indicate that the speakers have not acquired agreement.

The following research questions and predictions sum up the discussion of this section:

- **Research question 1:** Is functional morphology more difficult than core syntax?
Prediction 1: It should be less problematic to identify ungrammatical word order than ungrammatical agreement.

- **Research question 2:** Is functional morphology a more persistent problem than core syntax?
Prediction 2: The ability to identify ungrammatical word order should improve faster than the ability to identify ungrammatical agreement.
- **Research question 3:** Which syntactic and morphological sub-conditions are more difficult?
Prediction 3:
 - a. Non-V2 syntax is more challenging when the test sentences contain an auxiliary verb.
 - b. Long-distance agreement is more challenging than local agreement.
 - c. We expect oversuppliance and underuse of the *-s*, but the most common error should be oversuppliance.

Participants

Sixty students in two different age groups (11–12 and 15–18) participated in the experiment. The younger students attended 7th grade of a Norwegian primary school (N = 25) and the older students were in their first (N = 27) and second (N = 17) year of upper secondary school (11th and 12th grade). The participating classes were chosen by the schools. All participants were native speakers of Norwegian, with Norwegian as their only L1 and English as their L2. They were all introduced to English as a second language in their first year of primary school at the age of 6.

The participants' general proficiency level in English was measured by means of a subset of the Oxford Proficiency test, consisting of 40 items.

Experimental conditions and test items

We created different experimental conditions to test SV agreement and word order. Two conditions tested knowledge of core syntax: Non-subject-initial declarative main clauses with auxiliaries and lexical verbs, illustrated in (12) and (13). Four conditions represented functional morphology: Long-distance and local agreement in present tense subject-initial declarative main clauses with singular and plural subjects, illustrated in (14)–(17).

(12) Non-subject-initial declaratives with lexical verbs

- a. *Yesterday went the teacher to the shop.
- b. Yesterday the teacher went to the shop.

(13) *Non-subject-initial declaratives with auxiliary verbs*

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

(14) *Long-distance agreement with plural subjects*

- a. *The kids with the red bike plays in the garden.
- b. The kids with the red bike play in the garden.

(15) *Long-distance agreement with singular subjects*

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

(16) *Local agreement with plural subjects*

- a. *The teachers gives their students a lot of homework.
- b. The teachers give their students a lot of homework.

(17) *Local agreement with singular subjects*

- a. *The brown dog play with the yellow football.
- b. The brown dog plays with the yellow football.

There were 36 sentence pairs in the acceptability judgement task, with one ungrammatical and one grammatical version of the same sentence, i.e., six sentence pairs in each condition. We also included 13 fillers, which were all subject-initial declaratives with non-target-like placement of possessive determiners and definite articles, see examples (18) and (19), respectively. We expected that it would be unproblematic for the participants to detect these violations. This expectation was confirmed, since most participants rejected all fillers. The choice to use only ungrammatical fillers was based on the response bias known as acquiescence (also known as ‘yea-saying’ or ‘agreement tendency’), i.e., that sentences are more often endorsed than rejected in acceptability judgement tasks, regardless of their grammaticality (Mohan, 1977; see also Schütze, 2016).

- (18) *Mother my read a book about flowers.
(19) *Dogs the barked at the postman with the red jacket.

All test sentences were of equal length (10–12 syllables), and the lexical items were selected from COCA's (Corpus of Contemporary American English) top 5000-word frequency list (Corpus.byu.edu, 2018b). The purpose was to ensure that the sentences were similar in terms of length and vocabulary, and that the lexical items were familiar to the participants. In addition, all sentence subjects were full NPs. The non-subject-initial declaratives had past tense lexical verbs and modal auxiliaries to avoid judgments being based on SV agreement. The present tense declaratives that were used in the agreement test had no irregular verbs, and, as previously mentioned, the ones that tested long-distance agreement all contained prepositional phrases modifying the subject NP.

Research design and procedure

The study employed three tasks: an acceptability judgement task, a proficiency test and a background questionnaire. All parts were presented and recorded on Survey Gizmo, an independent survey service. Testing took place during school hours. The participants took approximately 25–35 minutes to complete the whole battery of tasks.

In the acceptability judgement task, the participants were asked to evaluate English sentences on a Likert scale from 1 to 4, where 1 signified that the sentence was completely unacceptable and 4 marked a completely acceptable sentence. There was also an additional option "I don't know". The instructions were given to the participants both orally and in writing. The oral instructions were given in Norwegian.

As illustrated in Figure 1, five sentences were presented on each page of the survey. Once the participants had judged all the sentences and moved on to the next page, they were not able to return to previous pages. Thus, they were not able to change any of their previous judgements. In addition, the test was pseudo-randomized to ensure that a sentence pair (acceptable and unacceptable) never appeared on the same page or immediately followed one another from one page to the next. Pseudo-randomizing made it possible to ensure that there was always at least one sentence that represented core syntax, one that represented functional morphology and no

more than one filler on each page.

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

19%

Next, the participants saw a 40-item subset of the Oxford Proficiency Test. This was a multiple-choice task in which they had to pick one out of three words or phrases to fill in a blank spot in a sentence, see example (20).

(20) Very _____ people can travel abroad.

- less
- little
- few

The third and final part of the test was a background questionnaire. Here, the participants were asked to answer questions about their age and linguistic background. This section was in Norwegian to avoid any misunderstanding.

Results

The proficiency test

The participants' scores on the multiple-choice proficiency test ranged from 11 to 38 (mean =

27.3, median = 29.5, mode = 35) out of 40 possible proficiency points. There was a strong positive correlation between age and proficiency scores (adjusted $r^2 = .7753$), and for that reason, only proficiency scores were used as a variable in the analysis of the data (see Appendix E).

The acceptability judgement task

To analyze the data, we fitted two linear mixed models: the first one for the two main conditions, SV agreement and non-V2, and the second one for all six sub-conditions, i.e., local and long-distance agreement with plural and singular subjects (morphological conditions) and non-V2 with lexical and auxiliary verbs (syntactic conditions), using the *lme4* package in R (Bates et al., 2015). We were interested in the participants' ability to identify ungrammaticality, and for that reason, the dependent variable was the difference in judgement (1-4 on a Likert scale) between the grammatical sentences and their ungrammatical counterparts. This gives us a 7-point scale that ranges from -3 to 3. To exemplify, we get the value -3 when a participant has rated a grammatical sentence as 1 and the ungrammatical counterpart as 4; the value 0 occurs when both sentences of a sentence pair are given the same score on the Likert scale, and a positive value occurs when a grammatical sentence is rated higher than the ungrammatical version of the sentence. Every "I don't know" option was treated as a missing value, and both versions of the sentence pair were excluded from the analysis. In total 56 pairs were removed, corresponding to 2.5% of the data. The predictors were proficiency scores (ranging from 11 to 38 out of 40 possible points), conditions (main conditions and sub-conditions) and the interaction between proficiency scores and conditions. We included random intercepts for participant and items (the sentence pairs) as well as a by-participant slope for condition in both models. We obtained p-values for the main effects and interactions of the two factors (proficiency and condition) by likelihood ratio tests (Pinheiro and Bates, 2000; Winter, 2013), starting off with an intercept-only model and thereafter adding the variables proficiency and condition and the interaction between them. In addition, p-values for the coefficients of the individual sub-conditions in the second model were obtained through the *lmerTest* package (Kuznetsova et al., 2017).

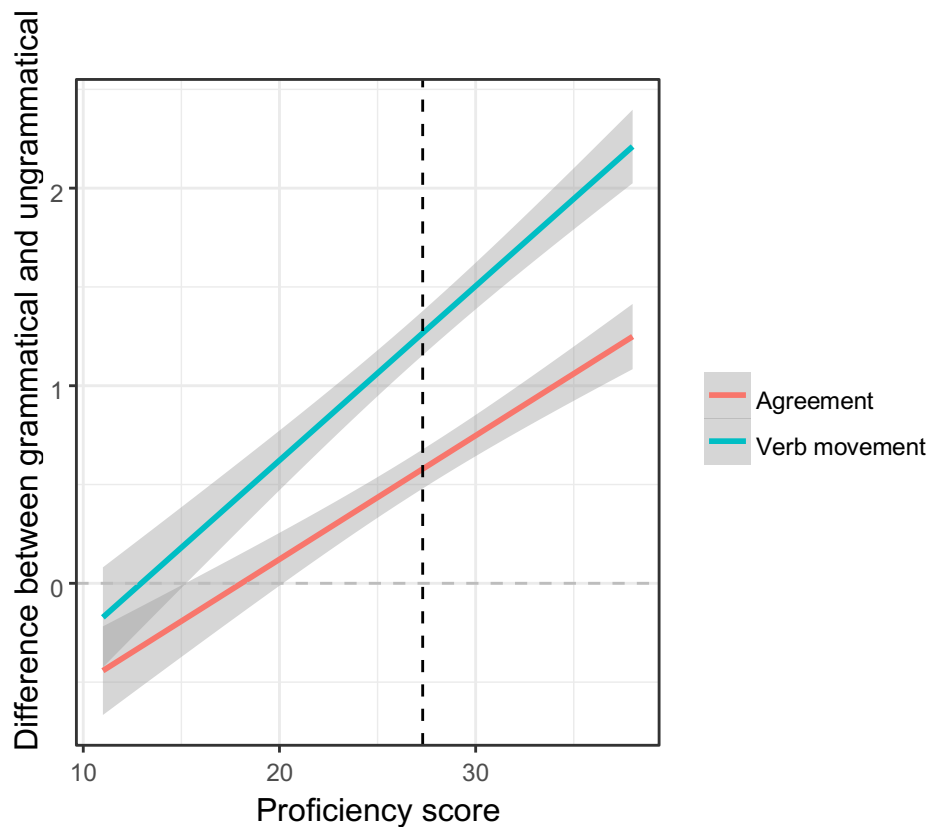
We found significant main effects of main condition ($\chi^2(1) = 39, p < .001$), sub-condition ($\chi^2(5) = 68.4, p < .001$), and proficiency score ($\chi^2(1) = 58.4, p < .001$). We also found a significant interaction of proficiency scores and main conditions ($\chi^2(3) = 159.0, p < .001$), and proficiency scores and sub-conditions ($\chi^2(5) = 58.7, p < .001$). We now unpack these results.

Model 1: Main conditions

Figure 2 visualizes the results for the main conditions, SV agreement and non-V2 word order. The values on the Y-axis represent the difference between Likert scale judgements (1–4) of grammatical and ungrammatical sentence pairs (from –3 to 3). The values on the X-axis represent the proficiency scores (11–38).

We found a significant main effects of main condition ($\chi^2(1) = 39, p < .001$, see Appendix F for the full model). As seen in Figure 2, non-V2 word order is consistently higher on the Y-axis than SV agreement. That is, the participants are overall more accurate at identifying ungrammatical word order than ungrammatical agreement, and this holds across proficiency levels. It should also be pointed out that the average difference between the grammatical and ungrammatical sentences is well above zero for both the word order and agreement pairs, which shows that the participants overall have a preference for the grammatical sentences over the ungrammatical ones.

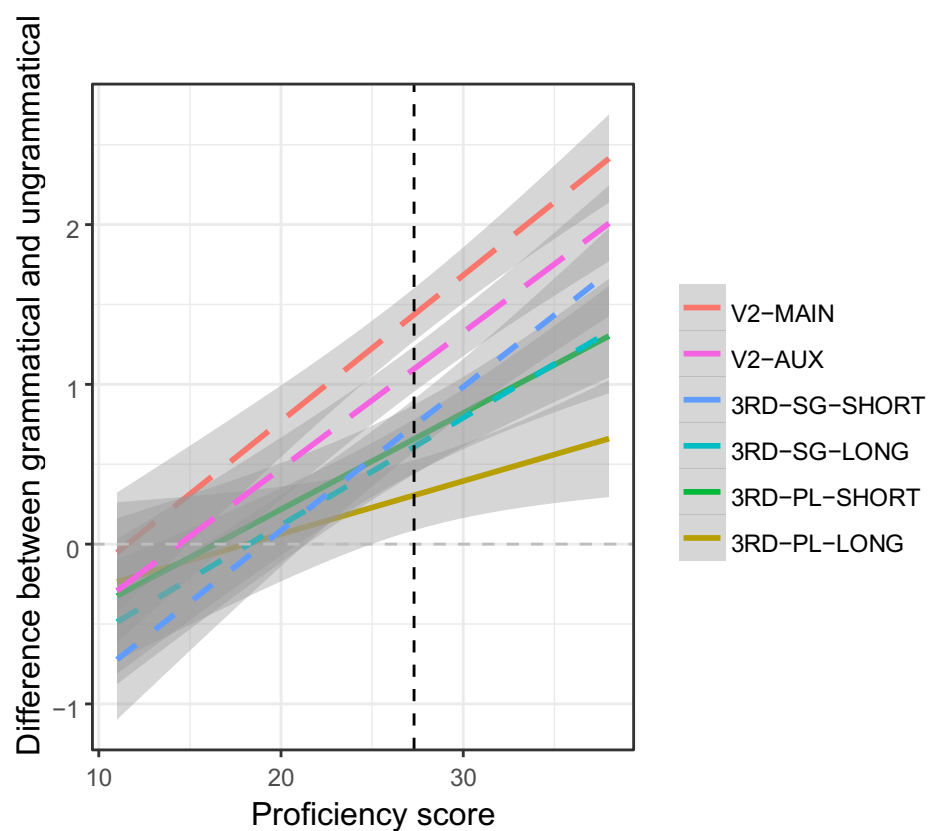
However, in investigating if a condition is a bottleneck, we are not only interested in an accuracy differential at a specific point in time, but also in how those accuracies develop over time in relation to general proficiency. We found a main effect of proficiency score ($\chi^2(1) = 58.4, p < .001$). As Figure 2 shows, the participants develop a stronger preference for the grammatical over the ungrammatical sentences as their general proficiency increases. More importantly, we found an interaction between proficiency and condition ($\chi^2(1) = 21.2, p < .001$). As Figure 2 and the model reveal, the slope for verb movement is significantly steeper than the slope for agreement ($\beta = .03, SE = .01, t > 2, p < .05$). That is, the L2 learners improve their accuracy for verb movement faster than for agreement, as predicted by the Bottleneck Hypothesis.



Model 2: Sub-conditions.

Figure 3 illustrates the results when the two main conditions, SV agreement and non-V2, are divided into sub-conditions. As previously discussed, there are four morphological and two syntactic sub-conditions: Local and long-distance agreement in subject-initial declaratives with singular and plural subjects, and non-V2 word order in non-subject-initial declaratives with lexical and auxiliary verbs. Again, in addition to the main effect of proficiency score, we find a main effect of condition ($\chi^2(5) = 68.4, p < .001$), and an interaction between proficiency and condition ($\chi^2(5) = 58.7, p < .001$). The full model is found in appendix G. Figure 3 shows that the lines that represent non-V2 word order (dashed purple and red lines online) are significantly higher at all proficiency levels than the agreement conditions. The mean difference is highest for the V2-Main condition (non-subject-initial declaratives with lexical verbs), followed by V2-Aux (significantly lower: $\beta = .35, SE = .1, t > 2, p < 0.05$). All the agreement conditions have significantly lower mean differences than the word order conditions (all p -values < 0.05). We see that the values for SG-SHORT (local agreement in declaratives with singular subjects), SG-LONG and PL-SHORT cluster together, while the difference value for PL-LONG is lower. That is, the participants are overall better at spotting word order violations than agreement violations,

and the hardest violation to spot is the long-distance plural agreement. The slopes differ as well: the SG-SHORT condition patterns with the two word order conditions, i.e., the acquisition path is equally steep for the local third person agreement as for the word order patterns. The long-distance agreement with plural subjects (yellow line) is significantly gentler than the word order slopes ($\beta = -.06$, $SE = .02$, $t > 2$, $p < 0.05$). The plural local and singular long-distance agreement conditions also have less steep slopes than the word order and singular local agreement conditions, both of them approaching significance ($p = 0.07$ for plural short, and $p = 0.065$ for singular long). Overall, the plural conditions are harder than the singular ones.



Discussion

Is functional morphology more difficult than core syntax?

The *a priori* prediction of the BH was that SV agreement would be more difficult than non-V2 syntax. Opposing predictions were based on learnability, frequency, and instruction. Predictably, each participant group made fewer errors in both morphology and syntax when the test sentences were grammatical. In other words, they accepted the grammatical sentences with high ratings,

attesting to the fact that the comprehensible input these learners have been exposed to has provided them with evidence of these grammatical features. However, just accepting grammatical sentences with high ratings is not sufficient. In assessing the learners' competence, we were looking for performance that accepts the grammatical sentences and rejects the ungrammatical ones in equal measure. In order to gauge this dual knowledge, we calculated difference scores, as explained in the previous section. As visible from Figure 2, the average difference between the grammatical and ungrammatical sentences started below zero for the youngest and low-proficiency learners, but climbed above zero in the more advanced participants, for both the word order and agreement conditions. In short, there was development with both properties.

Furthermore, we were interested in establishing whether one of the tested properties, as a whole, was easier to acquire than the other, or whether they were of equal difficulty. As Figure 2 illustrates, the two plot lines do not run together. We found that learners were more accurate in identifying ungrammatical word order than ungrammatical agreement across proficiency levels. These observations reveal that unlearning the V2 cue may be less problematic than learning the rules of SV agreement in English L2 acquisition. This is true even when the more difficult morphological property is more frequent, learnable from positive evidence and widely taught in language classrooms. Such findings lend support to our main prediction.

Is functional morphology a more persistent problem than core syntax?

In order to test the Bottleneck Hypothesis, it is necessary to not only investigate the difficulty of agreement and non-V2 syntax at a point in time, but also the way in which acquisition of these two properties develops. As this is a cross-sectional and not a longitudinal study, we use proficiency as a proxy for development and measure the slope of development with increased proficiency. The significant interaction between condition and proficiency suggests that accuracy on one property, verb movement, develops faster than the other, SV agreement. This is also visible from the different slopes of the lines in Figure 2. This finding supports our second prediction, that SV agreement would be a more persistent problem than verb movement in L2 acquisition of English.

Which syntactic and morphological sub-conditions are more difficult?

Looking at Figure 3 more generally, we can establish that the two sub-conditions of verb movement have higher mean difference scores across the proficiency spectrum, as compared to the four agreement sub-conditions. This is important because it attests that no agreement score is higher than a verb movement score. Still, meaningful differences across sub-conditions are also attested.

Morphological conditions. Since difference scores for agreement are lower than zero at the lowest proficiency levels, this suggests that ungrammatical agreement is frequently accepted. This situation changes substantially at the higher proficiency levels as difference scores become positive. Local agreement with singular subjects is attested to be the easiest agreement sub-condition, as expected. Our results correspond to the pattern found in Garshol (2018), which represents one of the rare investigations of young Norwegian speakers' agreement errors (15–16-year-olds). Together with Garshol's, our findings suggest that this pattern is specific to Norwegian learners of English. For the time being, this pattern remains unique to Norwegian learners of English in the literature and is not directly predicted by the BH. We can only speculate that this pattern may be due to massive overlearning and effective overgeneralization.⁵ What is more, the attested higher difficulty in noticing the wrong *-s* marker with plural subjects tentatively suggests that the low salience of the *-s* does not preclude acquisition and even overgeneralization to other types of subjects.

The next two sub-conditions, long-distance agreement with singular subjects and local agreement with plural subjects, seem to be developing in tandem (see Figure 3). In other words, plural agreement and the additional material between the subject and the verb cause roughly equal problems for these learners. The hardest sub-condition to develop is long-distance agreement with plural subjects, where the difference scores are lower and the development slope is less steep. All these findings very likely reflect processing effects, with two elements of difficulty in the structure (distance between the agreeing elements and the number feature of the subject) causing the lowest scores. They add to findings in previous studies, such as Foote (2011), which shows that more structural complexity in a sentence makes agreement more

⁵ We are grateful to an anonymous reviewer for discussing this point.

challenging, or Cunnings (2017), which attributes the effect to retrieval interference. Furthermore, since plural subjects cause more difficulty than singular ones (see above), it appears that these L2 learners have more trouble rejecting an overt morpheme next to a singular noun (e.g., *The kids with the red bike **plays** in the garden*) than a null one next to a plural noun (e.g., *The teacher with black shoes **walk** to work every day*). This increased difficulty with plural subjects is consistent with the findings in local agreement and suggests a general preference for the morpheme – s to be present.

Syntactic conditions. Sentences with auxiliary verbs proved to be more problematic than sentences with lexical verbs, as expected. That is, sentences with verb movement were accepted more often with auxiliaries than with lexical verbs. This supports our contention that learners would be more inclined to accept an auxiliary verb in a moved position, as opposed to a lexical verb, as had been found in some previous studies (e.g. Westergaard, 2003). A possible account of this would be to follow Pollock's (1989) analysis of auxiliaries as lighter and carrying less information and therefore having a higher propensity to move. Furthermore, as mentioned above, this result may also be due to auxiliaries in English behaving differently from lexical verbs in other contexts: Auxiliaries do move in certain contexts, whereas lexical verbs always stay in the VP (cf. examples in (8) and (9) above). This may provide learners with conflicting information and cause them to be less sensitive to ungrammatical auxiliary verb movement in non-subject-initial declaratives (e.g., **Yesterday was I thinking of you*).

Is the Bottleneck Hypothesis supported?

We would like to argue that the results in the current study lend initial support to the Bottleneck Hypothesis, as the participants' performance with functional morphology was weaker than with syntactic movement, as measured by accuracy at a point in time. Furthermore, performance with functional morphology lingered at lower accuracy levels, as measured by development across participants with increasing general proficiency. The functional morphology was found to be more difficult despite several factors working in its favor, such as being learnable from positive evidence, being more frequent, and being explicitly taught (see Table 1).

At the same time, it is also clear from our complex findings that there is internal variation within the syntactic and morphological properties that is beyond the predictions of the BH. In

this sense, the hypothesis remains on a general level of prediction with respect to whole language modules. It does not preclude other difficulty factors such as L1–L2 contrasts and processing or interference considerations from further modifying the actually attested ease or difficulty of acquisition. More concretely, learners in this study found it harder to identify the non-targetlike presence of overt agreement with plural subjects than its absence with singular subjects, and even harder to do this at a distance from the subject. In this sense, our findings point to the fact that the BH provides general module-based predictions, which may be further sharpened based on the individual properties under investigation and a host of other factors, some of which we identified in Table 1.

Conclusion and suggestions for future research

The purpose of the study presented here is to contribute to current knowledge of the cognitive process of L2 acquisition by testing the predictions put forward by the BH. We carried out an experiment on the L2 acquisition of English, focusing on learners' knowledge of functional morphology and core syntax. We also tested learners at different levels of proficiency in order to be able to investigate possible development over time. In a nutshell, our experimental results support the BH. However, these are initial findings and many questions remain. In order to further investigate the hypothesis, it is necessary to investigate functional morphology in direct comparison to other domains beyond core syntax, such as semantics or the syntax-discourse interface. Comparing different functional morphemes to different core syntactic constructions might also provide a more nuanced understanding of linguistic complexity, as in e.g., M. Jensen's (2017) study, which has also investigated L1 Norwegian learners of L2 English and found that there is a clear difference in difficulty between different types of morphological exponents and different types of syntactic structures. For example, it is almost certain that its low salience and functional redundancy (not being necessary for subject identification) are two factors that make the *-s* morpheme more difficult than certain other functional morphemes which are more salient and more useful for communication purposes, e.g., past tense *-ed* in English. Furthermore, different language combinations will clearly expand the inquiry profitably. Finally, in further studies it will also be essential to add on-line experimental techniques such as eye-tracking or speeded acceptability judgement tasks in order to investigate L2 learners' real-time language processing.

The general results provide implications for language teaching, as they illustrate that even at advanced proficiency levels, learners struggle with SV agreement, although this property is explicitly and continually taught in Norwegian schools. It seems that more practice beyond explicit instruction is needed in order to make the nascent competence more established in performance.

References

- Almor A, MacDonald M, Kempler D, Andersen ES and Tyler LK (2001) Comprehension of long distance number agreement in probable Alzheimer's disease. *Language and Cognitive Processes* 16: 35–63.
- Bailey N, Madden C and Krashen SD (1974) Is there a 'natural sequence' in adult second language learning? *Language Learning* 24: 235–243.
- Baker MC (2008) *The Syntax of Agreement and Concord*. Cambridge: Cambridge University Press.
- Bates D, Maechler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48.
- Bentzen K (2014) Verb placement in clauses with initial adverbial 'maybe'. *Nordic Atlas of Language Structures (NALS) Journal* 1: 225–239.
- Bock K and Miller CA (1991) Broken agreement. *Cognitive Psychology* 23(1): 45–93.
- Borer H (1984) *Parametric syntax. Case studies in Semitic and Romance Languages*. Dordrecht: Foris.
- Bowerman M (1988) The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In: Hawkins J (ed) *Explaining language universals*. Oxford: Basil Blackwell, pp. 73–101.
- Boyd JK and Goldberg AE (2011) Learning what not to say: the role of statistical pre-emption and categorization in a-adjective production. *Language* 87(1): 55–83.
- Braine MDS (1971) On two models of the internalization of grammars. In: Slobin DI (ed) *The ontogenesis of grammar. A theoretical symposium*. New York: Academic Press, pp. 153–186.
- Brown R (1973) *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- Brown R and Hanlon C (1970) Derivational complexity and order of acquisition in child speech.

- In: Hayes JR (ed) *Cognition and the development of language*. New York: Wiley.
- Bybee J (2000) *Mechanisms of Change in Grammaticalization: The Role of Frequency*. Albuquerque, NM: University of New Mexico.
- Chomsky N (1995) *The Minimalist Program*. Cambridge, MA: MIT Press.
- Cintrón-Valentín MC and Ellis NC (2016) Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus. *Frontiers in Psychology* 7: 1–21.
- Clahsen H and Hong U (1995) Agreement and null subjects in German L2 development: new evidence from reaction-time experiments. *Second Language Research* 11: 57–87.
- Colliander G (1993) Profiling Swedish second language acquisition: a method for assessing L2 proficiency. In: *Problem, process, product in language learning* (ed B Hammarberg), Stockholm, Sweden, 21–22 October 1992, pp. 32–47. Stockholm: University of Stockholm.
- Corpus.bya.edu (2018a) Corpus of Contemporary American English. Available at: <https://corpus.byu.edu/coca/> (accessed 11 May 2018).
- Corpus.bya.edu (2018b) Word frequency data. Available at: <https://www.wordfrequency.info/sample.asp#simple> (accessed 11 May 2018).
- Cunnings I (2017) Parsing and Working Memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20(4): 659–678.
- DeKeyser R (2007) Skill Acquisition Theory. In: VanPatten B and Williams J (eds) *Theories in Second Language Acquisition: An Introduction*. Mahwah, NJ: Erlbaum, pp. 94–112.
- Dröschel Y (2011) *Lingua Franca English: The Role of Simplification and Transfer*. Bern: Peter Lang.
- Dulay HC and Burt MK (1973) Should we teach children syntax? *Language Learning* 23(2): 245–258.
- Dulay HC and Burt MK (1974) Natural Sequences in Child Second Language Acquisition. *Language Learning* 24: 37–53.
- Ellis NC (2002) Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143–188.
- Ellis NC (2015) Implicit and explicit learning: Their dynamic interface and complexity. In: Rebuschat P (ed) *Implicit and explicit learning of languages*. Amsterdam: John

- Benjamins, pp. 1–24.
- Eubank L (1994) Optionality and the initial state in L2 development. In: Hoekstra T and Schwartz BD (eds) *Language acquisition studies in generative grammar*. Amsterdam: John Benjamins, pp. 369–88.
- Foote R (2011) Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics* 32: 187–220.
- Gabriele A (2009) Transfer and transition in the SLA of aspect. *Studies in Second Language Acquisition* 31: 371–402.
- Garshol L (2018) We Agreed to Disagree: Subject-Verb (Dis)Agreement Patterns in L2 English of Norwegian learners. In: *3rd Norwegian Graduate Student Conference in Linguistics and Philology: Abstracts*, Oslo, Norway, 15–16 February 2018, pp. 13.
- Goldschneider JM and DeKeyser RM (2001) Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A Meta-analysis of Multiple Determinants. *Language Learning* 51(1): 1–50.
- Haznedar B (2001) The Acquisition of the IP system in child L2 English. *Studies in second Language Acquisition* 23, 1–39.
- Haznedar B and Schwartz BD (1997) Are there optional infinitives in child L2 acquisition? *Proceedings of the 21st Annual Boston University Conference on Language Development (BUCLD)*. Somerville, MA: Cascadilla Proceedings Project.
- Håkansson G and Collberg SD (1994) The preference for Modal + Neg: an L2 perspective applied to Swedish L1 children. *Second Language Research* 10(2): 95–124.
- Inagaki S (2006) Manner-of-motion verbs with locational/directional PPs in L2 English and Japanese. In: Slabakova R, Montrul S and Prevost P (eds) *Inquiries in Linguistic Development*. Amsterdam: John Benjamins, pp. 41–68.
- Ionin T and Wexler K (2002) Why is ‘is’ easier than ‘-s’? Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research* 18(2): 95–136.
- Jensen M (2017) *Investigating the Bottleneck Hypothesis in Second Language Acquisition: The acquisition of narrow syntax and functional morphology among Norwegian L2 learners of English*. MA Thesis, University of Tromsø – The Arctic University of Norway, Norway.

- Keating G (2005) *Processing gender agreement across phrases in Spanish: Eye movements during sentence comprehension*. PhD Thesis, University of Illinois at Chicago, USA.
- Krashen S (1985) *The Input Hypothesis: Issues and Implications*. London: Longman.
- Kuznetsova A, Brockhoff PB and Christensen RHB (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13): 1–26.
- Lardiere D (1998a) Case and tense in the ‘fossilized’ steady state. *Second Language Research* 14: 1–26.
- Lardiere D (1998b) Dissociating syntax from morphology in a divergent end-state grammar. *Second Language Research* 14: 359–375.
- Lardiere D (2009) Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25(2): 173–227.
- Lightfoot D (1999) *The development of language: Acquisition, change, and evolution*. Wiley-Blackwell.
- Marcus GF (1993) Negative evidence in language acquisition. *Cognition* 46(1): 53–85.
- Mazurkewich I and White L (1984) The acquisition of dative alternation: unlearning overgeneralizations. *Cognition* 16: 261–283.
- Mohan BA (1977) Acceptability testing and fuzzy grammar. In: Greenbaum S (ed) *Acceptability in Language*. The Hague: Mouton, pp. 133–148.
- Montrul S and Yoon J (2009) Putting parameters in their proper place. *Second Language Research* 25(2): 291–311.
- Myles F (1995) Interaction between linguistic theory and language processing in SLA. *Second Language Research* 11(3): 235–266.
- NDLA (2016) *Årsmelding for 2016*. Report for Nasjonal Digital Læringsarena. Report no. 16.
- NDLA: Open Educational Resources For Secondary Schools (2018a) About NDLA. Available at: <http://om.ndla.no/about-ndla/> (accessed 11 May 2018).
- NDLA: Open Educational Resources For Secondary Schools (2018b) NDLA: Engelsk. Available at: <https://ndla.no/en/node/42> (accessed 11 May 2018).
- Pinheiro JC and Bates DM (2000) *Mixed-Effects Models in S and SPLUS*. New York: Springer.
- Pollock JY (1989) Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20(3): 365–424.
- Roeper T (1999) Universal bilingualism. *Bilingualism: Language and Cognition* 2(3): 169–186.

- Rugset A and Ulven E (2011) *Gateways*. Oslo: Aschehoug.
- Røkaas F, Kagge J, Kaspersen EC, Skanke JL and Haugum E (2011) *Worldwide*. Oslo: Aschehoug.
- Schütze CT (2016) *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Berlin: Language Science Press.
- Schwartz BD (1993) On explicit and negative data effecting and affecting *competence* and *linguistic behavior*. *Studies in Second Language Acquisition* 15(2): 147–63.
- Slabakova R (2008) *Meaning in the Second Language*. Berlin: Mouton de Gruyter.
- Slabakova R (2013) What is easy and what is hard to acquire in a second language: A generative perspective. In: García Mayo M, Gutiérrez Mangado MJ and Martínez-Adrián M (eds) *Contemporary approaches to second language acquisition*. Amsterdam/Philadelphia: John Benjamins, pp. 5–28.
- Solberg C and Unnerud HD (2015a) *Stairs 7: Textbook*. Oslo: Cappelen Damm.
- Solberg C and Unnerud HD (2015b) *Stairs 7: Workbook*. Oslo: Cappelen Damm.
- Trahey M and White L (1993) Positive Evidence and Preemption in the Second Language Classroom. *Studies in Second Language Acquisition* 15(2): 181–204.
- Vainikka A and Young-Scholten M (1994) Direct access to X'-theory: evidence from Korean and Turkish adults learning German. In: Hoekstra T and Schwartz BD (eds) *Language acquisition studies in generative grammar*. Amsterdam: John Benjamins, pp. 265–316.
- Vettorel P (2014) *English as a Lingua Franca in Wider Networking: Blogging Practices*. Berlin: De Gruyter Mouton.
- Vikner S (1995) *Verb Movement and Expletive Subjects in the Germanic Languages*. New York/Oxford: Oxford University Press.
- Waldmann C (2008) *Input och output: Ordföljd i svenska barns huvudsatser och bisatser*. PhD Thesis, Lunds Universitet, Sweden.
- Westergaard M (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 M (2007) English as a mixed V2 grammar: Synchronic word order inconsistencies from the perspective of first language acquisition. *Poznan Studies in Contemporary Linguistics* 43(2): 107–131.

- Westergaard M (2008) Acquisition and change: On the robustness of the triggering experience for word order cues. *Lingua* 118(12): 1841–1863.
- Westergaard M (2009a) *The acquisition of word order: Micro-cues, information structure and economy*. Amsterdam: John Benjamins.
- Westergaard M (2009b) Microvariation as diachrony: A view from acquisition. *Journal of Comparative Germanic Linguistics* 12: 49–79.
- Westergaard M (2009c) Usage-based vs. Rule-based Learning: The Acquisition of Word Order in Wh-Questions in English and Norwegian. *Journal of Child Language* 36(5): 1023–1051.
- Westergaard M, Lohndal T and Alexiadou A (2016) The asymmetric nature of V2: Evidence from learner languages. Talk given at CGSW 31, Stellenbosch, December 2.
- Westergaard M, Vangsnes ØA and Lohndal T (2017) Variation and change in Norwegian *wh*-questions: The role of the complementizer *som*. *Linguistic Variation* 17(1): 8–43.
- White L (1989) *Universal grammar and second language acquisition*. Amsterdam: John Benjamins.
- White L (1991) Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research* 7(2): 133–16.
- White L (2003) *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- Winter B (2013) Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499*.
- Yang C (2001) Internal and external forces in language change. *Language Variation and Change* 12(3): 231–250.
- Yang C (2002) *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.
- Yang C (2015) Negative knowledge from positive evidence. *Language* 91(4): 938–953.
- Zipf GK (1949) *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.

Appendix

Appendix A: Occurrences of lexical verbs in COCA (Corpus.byu.edu, 2018a)

Verb type	Count	Percentage of all present tense lexical verbs (N=16,535,156)	Percentage of all lexical verbs (N=61,188,902)
3 rd person singular	6,198,523	37.5%	10.1%
Non-3 rd person singular	10,336,633	62.5%	16.9%

Notes: Non-3rd person singular = 1st and 2nd person singular and plural, and 3rd person plural.

Appendix B: Occurrences of lexical verbs in spoken COCA (Corpus.byu.edu, 2018a)

Verb type	Count	Percentage of all present tense lexical verbs (N=4,042,909)	Percentage of all lexical verbs (N=13,175,959)
3 rd person singular	944,638	23.4%	7.2%
Non-3 rd person singular	3,098,271	76.6%	23.5%

Notes: Non-3rd person singular = 1st and 2nd person singular and plural, and 3rd person plural.

Appendix C: Raw counts and percentages of declaratives in child-directed speech in the Pearl Sprouse Corpus (offering Penn TreeBank style parses on data from the Brown, Soderstrom, Suppes, and Valian corpora in CHILDES); searches carried out by Charles Yang, who is hereby thanked for his important contribution.

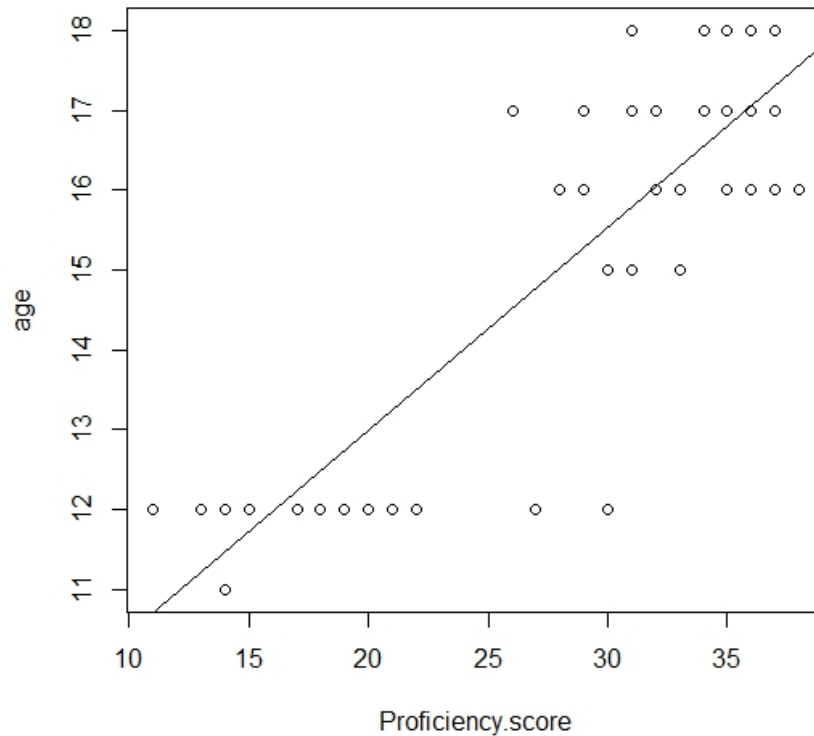
Sentence type	Count	Percentage of all declaratives (N=46,117)
V3 sentences	4449	9.6%
SXV	1115	2.4%
XSV	3114	6.8%

Notes: The numbers include declarative sentences with a verb (but no questions, imperatives or interjections). V3 sentences are those where the verb is the third constituent, both SXV and XSV (the latter being our focus of interest).

Appendix D: Number of exercises and articles in the learning material that focus on subject-verb agreement and word order.

Material	Grade	Agreement			Word order		
		Exercises	Articles	Total	Exercises	Articles	Total
Røkaas et al. (2011)	12 th	17	3	20	0	0	0
Rugset and Ulven (2011)	11 th	0	1	1	0	0	0
NDLA (2018b)	11 th – 13 th	10	10	20	1	1	2
Solberg and Unnerud (2015a,b)	7 th	2	1	3	0	0	0

Appendix E: Correlation between proficiency scores and age



Appendix F: Main conditions, Model coefficients for average difference in judgements (based on the 1-4 Likert scale) between the grammatical sentences and their ungrammatical counterparts. There are 2106 observations, 60 participants and 36 items. The model includes random intercepts for Participant and Item, and a by-participant random slope for condition. The intercept is the overall mean for SV agreement. The variance for the random effects is given in the right-hand part of the table.

Fixed effects	Estimate	SE	df	<i>t</i>	<i>p</i>	Random effects	Variance
Intercept	.576260	.087273	77.62	6.6	<.001	Participant Intercept	.32351
				03			
Word order	.681819	.118075	72.58	5.7	<.001	Participant Word order	.34496
				74			
Proficiency	.062857	.009705	60.66	6.4	<.001	Item Intercept	.03808
				77			
Word order: Proficiency	.026901	.012026	61.59	2.2	.0289		
				37			

Appendix G: Sub-conditions, Model coefficients for average difference in judgements between the grammatical sentences and their ungrammatical counterparts. There are 2106 observations, 60 participants and 36 items. The model includes random intercepts for Participant and Item, and a by-participant random slope for condition. The intercept is the overall mean for V3 with lexical verbs. The variance for the random effects is given in the right-hand part of the table.

Fixed effects	Estimate	SE	df	<i>t</i>	<i>p</i>	Random effects	Variance
Intercept	1.430755	.090646	56.74	15.784	< .001	Participant Intercept	.25377
Proficiency	.092891	.010193	61.51	9.113	<.001	Participant 3rd-pl-long	.93549
3rd-pl-long	-1.125235	.153390	64.38	-7.336	<.001	Participant 3rd-pl-short	.81075
3rd-pl-short	-.779870	.146438	63.18	-5.326	<.001	Participant 3rd-sg-long	.49580
3rd-sg-long	-.815434	.127230	56.59	-6.409	<.001	Participant 3rd-sg-short	.49214
3rd-sg-short	-.687718	.126922	57.42	-5.418	<.001	Participant Word order-aux	.16072
Word order-aux	-.346288	.103222	42.07	-3.355	<.01	Item Intercept	.01006
Proficiency: 3rd-pl-long	-.060291	.017818	61.75	-3.384	<.01		
Proficiency: 3rd-pl-short	-.031107	.016892	61.91	-1.842	.0703		
Proficiency: 3rd-sg-long	-.026738	.014247	61.22	-1.877	.0653		
Proficiency: 3rd-sg-short	-.003344	.014182	62.67	-.236	.8144		
Proficiency: Word order-aux	-.006495	.010850	66.54	-.599	.5515		