Original Article

second language research

Second Language Research
1–27

DOI: 10.1177/02676583221128520
journals.sagepub.com/home/slr
SAGE

Testing the Bottleneck Hypothesis: Chinese EFL learners' knowledge of morphology and syntax across proficiency levels

Shiyu Wu
Shanghai Jiao Tong University, China

Dilin Liu
The University of Alabama, USA; Dalian University of Foreign Language, China

Zan Li
Shanghai Jiao Tong University, China

Abstract

This study tests the Bottleneck Hypothesis (BH) that functional morphology presents the greatest difficulty in second language acquisition by examining Chinese English as a foreign language (EFL) learners' knowledge of both functional morphological properties and core syntactic properties across three language proficiency levels. Specifically, this study compares Chinese EFL learners' grasp of subject–verb agreement (a functional morphological property) vs. their grasp of the syntactic formation rules (properties) of WH-questions including WH-movement, WH-*do*-insertion (i.e. the insertion of the auxiliary *do*), and WH-*do*-inversion (subject–auxiliary inversion): three core syntactic transformation rules. Analyses of the experimental results using generalized mixed-effects models yield complex results that generally support the BH. While subject–verb agreement was found to be persistently more difficult than WH-movement and WH-*do*-inversion, it was shown to pose essentially the same level of difficulty as WH-*do*-insertion due to the fact that the latter also involves the application of the functional features of SV-agreement and tense, i.e. the high level of difficulty of WH-*do*-insertion likely lies in its accompanying application of functional features. Possible explanations for the complex results and their implications are discussed.

Keywords

Bottleneck Hypothesis, core syntax, functional morphology, learning difficulty, SLA

Corresponding author:
Dilin Liu, School of Foreign Languages, The University of Alabama, Box 870244, Tuscaloosa, AL 35487, USA.
Email: dliu@ua.edu

I Introduction

Understanding what features or properties of a second language (L2) are particularly difficult to acquire is very important in second language acquisition (SLA) research (DeKeyser, 2005). Many studies have found that functional properties (e.g. inflectional morphemes) are especially challenging (Hawkins and Chan, 1997; Jiang, 2004; Lardiere, 1998a, 1998b, 2009; Prévost and White, 2000; Slabakova, 2008, 2013). Based on such findings, Slabakova (2008, 2013, 2016) has proposed the Bottleneck Hypothesis (hereafter BH), which postulates that acquiring functional morphology of an L2 constitutes the greatest challenge in SLA, greater than acquiring syntactic structures and semantics. The hypothesis has important implications for both SLA research and L2 pedagogy because, if supported by strong research evidence, it may help SLA researchers and teachers gain a better understanding of the challenges in SLA and in turn allow them to zero in on the key issues in their research and teaching effort. However, so far, little empirical research has been conducted to test the BH except for Jensen et al.'s (2020) recent study, which, according to the authors, was the first one that tested the BH directly via an experiment. This lone study dealt with Norwegian speakers learning English as an additional language comparing their acquisition of English subject–verb (SV) agreement (a functional morphological property) with their acquisition of word order, specifically verb movement in declarative sentences (a core syntactic property). It would be very helpful and important to conduct more empirical studies testing this hypothesis, especially those that involve L2 English learners from other L1 backgrounds or L2 learners of other languages.

Against this backdrop, the present study aims to test the BH by examining Chinese English as a foreign language (EFL) learners' grasp of subject–verb (S–V) agreement (a functional morphological property) and three syntactic properties involved in the formation of WH-questions: WH-movement, the insertion of the auxiliary *do*, and subject–auxiliary inversion. The rest of the article is organized as follows. Section II provides the theoretical background and motivations for the present study. Section III states the research purposes and questions. Section IV describes the methodology. Section V presents the data analysis, results, and discussion. Section VI concludes the article with a summary of the main findings and their implications.

II Background

1 The BH and its implications

According to the BH, functional morphology – which covers a variety of functional categories (FCs) such as case, gender, and number – is very important albeit especially challenging in SLA (Slabakova, 2008, 2013). It is important because 'most grammatical meanings are captured in functional morphology', a fact that makes it a central issue in language acquisition (Slabakova, 2013). It is challenging because, based on the Borer–Chomsky Conjecture (Baker, 2008; Borer, 1984; Chomsky, 1995), functional morphology, more than syntax, is where cross-language differences lie and hence where learning difficulties may most likely be found (Jensen et al., 2020). Furthermore, the Borer–Chomsky Conjecture assumes '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' (Jensen et al., 2020: 4). Unlike the workings of functional morphology, the operations of core syntax and semantics are believed to be largely universal, i.e. they adhere to the same execution rules and constraints across languages (Jensen et al., 2020).

The challenges of functional morphology are exacerbated by the fact that it 'bundles a variety of semantic, syntactic and morphophonological features that have an effect on the acceptability and the meaning of the whole sentence' (Slabakova, 2019: 319). This implies that to acquire functional morphology, learners need to grasp not only specific morpho-lexical forms and their allomorphic variations but also the various relevant conditioning factors governing their uses (Hwang and Lardiere, 2013). Added to the challenges is the fact that a bundle of formal features may or may not show overtly with an inflectional morpheme (Jensen et al., 2020). In short, the complexity of the information involved in the use of FCs makes the acquisition of functional morphology especially onerous.

One related issue that is of debate but worth noting for the present study is that there have been two opposing views regarding whether morphology or syntax is learned first: the 'morphology-before-syntax' hypothesis versus the 'syntax-before-morphology' postulation (White, 2003: 183–87). The former view holds that the acquisition of functional morphology precedes and propels the acquisition of syntax (Clahsen and Hong, 1995; Eubank, 1994; Vainikka and Young-Scholten, 1994). In contrast, the latter view contends that syntactic knowledge, as evidenced by right word order and correct structural movement, emerges much earlier than the knowledge of functional morphology manifested by correct provision of obligatory functional morphological features (Haznedar and Schwartz, 1997; Lardiere, 1998a, 1998b; White, 2003). Like Jensen et al. (2020), we embrace the latter view in this study.

2 Key linguistic/input variables affecting the acquisition of an L2 property

The reason we examine the key linguistic/input variables that have been found to affect the level of success or difficulty in the acquisition of an L2 property is that, in order to test the BH, we must understand these variables to ensure that when a language property (being it a morphological or syntactic one) is found to be acquired faster and more successfully than another, it is not mainly due to one or more of these influencing variables. Jensen et al. (2020: 4) have identified four such variables or 'dimensions along which the two properties [in comparison] vary', including (a) learnability, (b) frequency, (c) salience, and (d) instruction, all of which will be introduced below. We will also discuss L1–L2 similarity/difference as an additional variable that has been found to affect L2 acquisition (Ellis, 2016).

a Learnability. Language learnability deals with the kind of input needed to acquire a given target language property and the acquisition process involved. It centers on the issue of the availability of positive evidence. Generally, if a target language property is available in the input, it serves as positive evidence, making it learnable; on the other hand, if a target property is unattested in the input, then negative evidence in the form of

explicit instruction about ungrammatical uses of this property may be required in order for it to be acquired (Jensen et al., 2020; Prawatmuang and Yuan, 2020). Research has suggested that language properties that are available in the input are easier to grasp than those that are unavailable in the input (Gabriele, 2009; Juffs, 1996; Slabakova, 2006; White, 1991).

b Frequency. If a given target language property is available in the input, how frequently it occurs in the input can significantly determine its acquisition rate (Christiansen and Chater, 2001; Ellis, 2002; Madlener, 2015). Generally, with other factors being constant, the higher the frequency a linguistic property boasts, the higher its acquisition rate will be. Yet, research has also found that frequency itself may not suffice for a given linguistic usage to be acquired unless the input is unmistakably clear that this specific usage is the correct one in the target language (Yang, 2002), a point we will illustrate with examples below.

c Salience. Salience has been used mainly to refer to how noticeable a linguistic property is acoustically or perceptually (Ellis, 2016; Goldschneider and DeKeyser, 2001). For example, a stressed vowel is more salient than an unstressed one. The more salient a target language structure/item is, the easier it is to acquire (Ellis, 2016; Goldschneider and DeKeyser, 2001).

d Instruction. This variable deals with whether and to what extent instruction may assist the acquisition of a given linguistic property compared with naturalistic exposure. Some scholars argue that formal instruction does not assist acquisition because what is learned from instruction is explicit knowledge and cannot become implicit, proceduralized knowledge needed in language acquisition, i.e. there is no interface between language learning and acquisition (e.g. Krashen, 1981, 1985). However, other scholars contend that there is a strong interface between the two since explicit knowledge can become proceduralized and unconscious via practice (McLaughlin, 1987; DeKeyser, 2007). There are also some scholars who believe in a weak interface between the two while simultaneously recognizing the dominant role naturalistic exposure plays in SLA (Ellis, 2015).

e L1–L2 similarity (or congruence). Although cross-linguistic similarities do not always mean ease for L2 acquisition and differences do not necessarily cause difficulties, ample research has shown that some inter-linguistic differences, including those in functional morphology, pose serious challenges (Liu and Gleason, 2002; Liu et al., 2020; Montrul, 2001; Odlin, 1989). In other words, while complete congruence between learners' L1 and L2 can lead to positive L1 transfer, incongruence may often cause negative transfer, resulting in learning difficulties.

3 The target linguistic properties of English: Rationales and description

This study elects to use (a) SV-agreement to measure learners' knowledge of functional morphology just as Jensen et al. (2020) did and (b) the formation of WH-questions to test

their knowledge of core syntax because WH-question formation, with the exception of when the WH-word/phrase is the subject (e.g. '*Who* said so / *What* happened?'), involves the application of two or more of the following core syntactical transformation rules 'WH-movement', 'insertion of the auxiliary *do* or *do*-support', and 'subject–auxiliary inversion' (Adger, 2002; Chomsky, 1965; Radford, 2004). It is important to note that the *do*-insertion property when applied also involves the application of the functional features of SV-agreement (e.g. *Does Tom speak Chinese?*) and tense (e.g. ***Did*** *Tom go?*). In other words, the application of *do*-insertion is not purely a syntactic operation.

Our rationales for choosing SV-agreement and WH-question formation in our study are twofold. First, SV-agreement and WH-question formation-related transformational properties are absent in the Chinese language, an issue that we will elaborate on below when we describe the differences between English and Chinese related to these target properties. Second, previous studies have shown that Chinese EFL/ESL (English as a foreign language / English as a second language) learners, including advanced learners, often have difficulty with these properties (e.g. Chen et al., 2007; Chuang and Nesi, 2006; Darus and Ching, 2009; Jiang, 2004; Jung, 2010; Lee, 2012, 2016; Wei et al., 2015).

Regarding SV-agreement, several experimental studies have found that Chinese EFL/ESL learners were insensitive to SV-agreement issues in reading (Chen et al., 2007; Jiang, 2004) and in elicited sentence production (Wei et al., 2015). Furthermore, some error-analysis studies on Chinese learners' writings (e.g. Chuang and Nesi, 2006; Darus and Ching, 2009) have identified SV-agreement errors as one common type of error. In short, existing research has demonstrated that acquiring SV-agreement constitutes a persistent challenge for Chinese EFL/ESL learners even at the advanced proficiency level (Chen et al., 2007).

Concerning WH-questions, several studies have found that Chinese EFL/ESL learners have substantial difficulty in processing and producing WH-questions correctly (Jung, 2010; Lee, 2012, 2016).[1] Jung's (2010) study compared the performance in reading/comprehending English WH-questions between Chinese/Korean EFL/ESL learners and native English speakers. The results showed that the EFL/ESL learners' accuracy scores were much lower than native English speakers', and their reading profiles were also noticeably different from native English speakers', revealing the former's considerable difficulty in processing and comprehending English WH-questions (Jung, 2010). Lee (2012, 2016) examined the WH-questions that they elicited from Chinese EFL/ESL learners and found many errors, highlighting the challenges that Chinese EFL/ESL learners face in learning these questions.

a Functional morphology (SV-agreement) in Chinese and English. English is an inflectional language whereas Chinese is essentially noninflectional, i.e. the latter has extremely limited morphology for number, case, or gender (Packard, 2000). As an inflectional language, English requires SV-agreement, i.e. the present-tense 3rd person singular *-s* is obligatory for verbs whose subject is singular. In contrast, there is no SV-agreement in Chinese as illustrated in 1a and 1b since the Chinese verb *xihuan* ('love') is uninflected as all Chinese verbs.

(1) a. Ming xihuan faguo shipin.
Ming love French food.
'Ming loves French food.'
b. Ming he Lan xihuan faguo shipin.
Ming and Lan love French food.
'Ming and Lan love French food.'

On the surface, the use of the third person singular *-s* for expressing SV-agreement looks simple and straightforward, but, as Jensen et al. (2020) explain clearly, the English FC for SV-agreement is actually quite complex.

First, the overt *-s* entails a number of underlying abstract morphosyntactic features, such as tense, aspect, and agreement. Second, this FC also includes morphosyntactic '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)' (Jensen et al., 2020: 7). To use this FC correctly, learners must grasp all the abstract and overt features involved. Following Jensen et al. (2020), we do not test learners' knowledge of these related features because ample research has been carried out (see review by White, 2003: 187–93). The findings of the previous research on the acquisition of syntax and functional morphology have shown that '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%' (Jensen et al., 2020: 7). Such findings support the syntax-before-morphology position.

b Syntax (WH-question formation) in Chinese and English. As noted above, the formation of WH-questions in English, with the exception of when the WH-word/phrase is the subject, requires two or more core syntactic transformational properties (moves) as shown in examples 2 and 3.

(2) *What* is this? (WH-movement and inversion involved)
(3) *What did* he say? (WH-movement, *do*-insertion, and inversion involved)

No such movement and insertion properties are found in the formation of Chinese WH-questions as illustrated in 4 and 5.

(4) zhe shi shenme?
This is what?
'What is this?'
(5) ta shuo [le] shenme? (*le* is an aspect marker)
He say [perfective marker] what?
'What did he say?'

As shown in the above examples, the formation of WH-questions in English is much more complex than that in Chinese thanks to its required WH-movement, *do*-insertion, and/or auxiliary–subject inversion. Of these required English syntactic movement properties, the latter two are much more difficult for Chinese EFL/ESL learners, for research has shown that these learners make more lack-of-*do*-insertion and/or lack of inversion

errors (e.g. *'What he said?'/'Where it is?') than lack-of-movement errors (e.g. *'He said what?'/ 'Where I buy tickets?') (Lee, 2012, 2016). Because of these seemly different levels of difficulty posed by the different syntactical properties for forming WH-questions, it will be of both interest and importance to test them separately in our study.

c Comparison of the target properties in light of influencing factors. As discussed above, some variables may affect the rate of success or level of difficulty in the acquisition of language properties. Hence, to compare the level of difficulty of the functional morphological and syntactic properties, it is necessary to understand how they each measure up against the five key influencing variables mentioned above in Section I.2.

Let us first look at the *learnability* variable (i.e. whether positive evidence alone suffices or negative evidence is needed for the learning of a linguistic feature). Regarding SV-agreement, Jensen et al.'s (2020: 8) analysis of the present tense uses of lexical verbs in Davies's (2008–) Corpus of Contemporary American English (COCA) suggests that 'English provides ample evidence for SV-agreement', for, there were 31,798 present-tense lexical verb tokens per million words (PMWs) in COCA and 11,920 (37.5%) of them were 3rd person singular uses.[2] In contrast, English does not provide ample clear evidence for WH-questions based on our query/analysis of a four million-word representative sample of COCA: there were only 469 WH-questions PMWs, which accounted for only 0.7% of the total 69,682 sentences PMWs. In other words, the 469 PMW frequency of WH-questions is only 4% of the 11,920 PMW frequency of SV-agreement. This limited evidence of WH-questions is further exacerbated by the aforementioned problem that not all WH-questions require movement and/or insertion of the auxiliary *do*. The two problems together will likely make it difficult for learners to know clearly from the input which WH-questions call for movement/insertion of *do*, hence necessitating negative evidence for learners to gain a clear understanding of the issue. Regarding the second variable *frequency*, based on the frequency information of the two features just mentioned, SV-agreement boasts a frequency 25 times that of WH-questions (11,920 PMWs/469 PMWs = 25.4). In short, as far as learnability and frequency are concerned, WH-question formation appears to be much more difficult than SV-agreement due to its limited and unclear evidence as well as its low frequency.

Concerning the third variable *salience*, WH-movement and the insertion/inversion of *do* in WH-questions appear to be much more perceptually salient than SV-agreement. This is because the fronting movement of the WH word and the insertion/inversion of *do* are much more perceptually noticeable than the 3rd person singular *-s*, which is an unstressed consonant attached to the end of a verb. The high perceptual salience of WH-questions will suggest that they are easier to acquire than SV-agreement.

As far as the variable *instruction* is concerned, English textbooks used by Chinese schools generally contain a few units designed specifically for teaching SV-agreement, but they offer no such units for teaching WH-questions. The latter are merely included as part of the units on question formations, i.e. there is no specific isolated unit on WH-questions. In other words, while WH-questions are taught, they do not receive the type of intensive instructional focus given to SV-agreement. Finally, regarding the variable *L1–L2 similarity*, both the SV-agreement FC property and the three syntactic

Table 1. Summary of variables affecting the difficulty of the target properties in Chinese EFL/ESL acquisition.

	Grammatical module	Learnability	Frequency of evidence	Salience	Instruction	Possible L1 interference
SV-agreement	Functional morphology	Clear positive evidence	High	Low	Focused instruction	Yes
WH-question formation	Core syntax	Inadequate positive evidence	Low	High	No focused instruction	Yes

properties involved in WH-question formation are absent in Chinese. Hence, they should be similarly challenging since L1 interference has been found in the application of these properties as evidenced by ample research showing Chinese EFL/ESL students' persistent failure to provide the 3rd person *-s* in required contexts, frequent nonuse of the auxiliary *do* when needed, and occasional failure to make required WH-movement (Chuang and Nesi, 2006; Darus and Ching, 2009; Jiang, 2004; Lee, 2012, 2016; Wei et al., 2015).

Table 1 summarizes the information concerning the influencing variables related to the two SV-agreement and WH-question formation. According to the BH, SV-agreement (an FC) should be more difficult than WH-question formation. Yet, most of the variables related to the target properties appear to work against this BH-based prediction. This is because SV-agreement seems to enjoy more favorable acquisition conditions than WH-question formation in three out of the five variables (learnability, frequency, and instruction) while the latter has an advantage in only one variable: salience. The two fare the same in the L1–L2 interference variable since they both suffer from L1 interference.

III Research purposes and questions

As has just been noted, based on the information in Table 1, SV-agreement should be easier to acquire thanks to its clearer and more frequent positive evidence and the more focused instruction it receives compared to WH-question formation. However, if the BH is correct, then SV-agreement should be more difficult to acquire than WH-question formation. This is because the BH assumes that 'knowledge of core syntax would improve faster than suppliance of functional morphology as the speakers become more advanced' (Jensen et al., 2020: 10). Hence, to empirically test the BH, this study compares Chinese EFL/ESL students' acquisition of the target functional morphological and core syntactic properties across different proficiency levels.

Because, as noted earlier, the three syntactic properties involved in the formation of English WH-questions have been found to present different levels of difficulty, we will compare the level of difficulty of each of the three properties with that of SV-agreement. Furthermore, we will also test whether the subconditions of SV-agreement (i.e. 'singular vs. plural subject' and 'local- vs. long-distance agreement' sentences to be described in Section IV.2) pose different levels of difficulty to Chinese EFL learners because previous

Table 2. Mean ages, lengths of English study and proficiency test scores by group (standard deviations in parentheses).

Group ($N = 124$)	Mean age (years)	Mean length of English study (years)	Mean language proficiency score
Middle-school ($n = 46$)	13.90 (0.42)	5.00 (0)	23.10 (6.36)
High-school ($n = 33$)	16.80 (0.58)	8.07 (0.32)	26.50 (3.84)
College ($n = 45$)	18.90 (0.64)	10.9 (0.88)	29.50 (2.71)

studies (Foote, 2011; Jensen et al., 2020) have found such subconditions did present different levels of difficulty for EFL/ESL learners with other L1s.

Specifically, this study attempts to answer the following research questions.

- Research question 1: Does learning the FC property of *SV-agreement* pose greater and more persistent difficulty for Chinese EFL learners than learning the core syntactical properties of *Wh-movement*, *WH-do-insertion*, and *WH-do-inversion*?
- Research question 2: Do the subconditions of SV-agreement present different levels of difficulty?

Our reason for treating Wh-movement, *do*-insertion, and auxiliary–subject inversion as three separate syntactic transformational properties but treating singular vs. plural subject and local- vs. long-distance agreement (in SV-agreement) as one FC property is that while the former three are each a truly distinct syntactic property or operation, the four subconditions of SV-agreement are each not a distinct FC property because all the four subconditions are governed by the same FC property: adding the third person singular *-s* to the verb in the present tense when the subject is a third person singular noun. The four subconditions have no bearing on this single FC property although the degree of success in applying this property may or may not vary across the four subconditions.

IV Methodology

1 Participants

One-hundred and twenty-four Chinese students from three different age groups participated in this experiment (for a summary of their demographics, see Table 2). The first group consisted of 46 eighth grade students aged 14–16 years and recruited from a middle school in a province in eastern China. The second group included 33 11th grade students aged 16–18 years from a high school in the same province. The third group was composed of 45 second-semester college freshmen aged 18–20 years from a key university in China. All the students were native speakers of Chinese. Because the length of English study varied across the three groups with the college freshmen boasting the longest and the 8th graders having the least, we expected the English proficiency level also varied across the groups. We confirmed this by having the participants take a proficiency test described below. Hereafter, the three groups are labeled ‘middle-school’, ‘high-school’, and ‘college’ respectively.

2 Materials

The background questionnaire (provided in supplementary material part 1). This questionnaire consisted of questions concerning the participants' demographics, including their age, gender, length of English learning, and whether and how long they had stayed in an English-speaking country.

The English proficiency test (provided in supplementary material part 2). Following Jensen et al.'s (2020) practice, we used a 40-item subset of the Oxford Proficiency test to test the participants' English proficiency. The format of the test was multiple-choice where the participants had to select one out of three options to fill in a blank as illustrated in the following example (7):

(6) He is very well known _____ the world.
A. all in B. all over C. in all

The acceptability judgement test (provided in supplementary material part 3). Altogether, 40 sets of sentences were created for this test. They were evenly divided into two groups according to the distance between the subject and the verb of the sentence with the verbs in one group being contiguous with the subject (i.e. local-distance) and the verbs in the other group interrupted by a modifying propositional phrase (i.e. long-distance). Each sentence set consists of two pairs and eight related sentences as shown in (7). The first pair include four versions of a declarative sentence in the present tense used to test the participants' knowledge of SV-agreement, with two – (7a) and (7c) – being grammatical and two – (7b) and (7d) – being ungrammatical (where either the required third person singular *-s* is missing or is supplied erroneously when the subject is plural). The second pair contains two versions of an interrogative sentence designed to test the participants' knowledge of WH-movement, with one (7e) being grammatical and one (7f) being ungrammatical for making no WH-movement. The sixth and seventh sentences were also WH-questions designed to test the participants' knowledge of the required auxiliary *do* insertion (7g) and subject–auxiliary (*do*) inversion (7h) respectively, for, whereas WH-movement is made in both questions, (7g) missed the required auxiliary *do* and (7h) failed to invert *do* with the subject.

(7) a. The little child plays football each Saturday.
b. *The little child play football each Saturday.
c. The little children play football each Saturday.
d. *The little children plays football each Saturday.
e. What did the little children play last Saturday?
f. *The little children played what last Saturday?
g. *What the little children played last Saturday?
h. *What the little children did play last Saturday?

In addition, to answer research question 2 about the effects of the four subconditions of SV-agreement, we included these conditions evenly in the two groups of SV-agreement questions (i.e. local- vs. long-distance) with twenty sentence pairs in each sub condition (see supplementary material part 3). Sentences (8–11) are examples of the four

conditions, with (8) being singular local-distance, (9) being plural local-distance, (10) being singular long-distance, and (11) being plural long-distance.

(8) a. The lady walks her dogs in the park every day.
b. *The lady walk her dogs in the park every day.
(9) a. The students play games in the park every Sunday.
b. *The students plays games in the park every Sunday.
(10) a. The girl in the red coat speaks French very well.
b. *The girl in the red coat speak French very well.
(11) a. The little frogs in the rice field jump from time to time.
b. *The little frogs in the rice field jumps from time to time.

Measures were taken to ensure the comparability of the sentences both within each sentence set and across the main conditions (SV-agreement vs. WH-question formation) and subconditions. First, sentences in each set were created from the same sentence base although they were in different types (see examples in 7). Second, all the sentences were of essentially equal length (11–13 syllables; see example 7–11). It is important to note that the long-distance SV-agreement sentences included some that contained an intervening prepositional phrase ending with a singular noun (e.g. 'in the rice field' in [11]) and some that contained an intervening phrase ending with a plural noun (e.g. 'in white T-shirts' in [13]). To prevent this difference from becoming a confounding variable, we had equal number of the two types of intervening phrases in the two long-distance SV agreement conditions.

Furthermore, because the sentences in each set were all derived from the same base sentence, the creation of the subconditions for SV-agreement questions also resulted in two different types of WH-questions: those with a subject containing no post-nominal modifier shown in (12) and those with a subject containing a post-nominal modifier shown in (13).

(12) a. What did the students do in the park last Sunday?
b. *The students did what in the park last Sunday?
c. *What the students did in the park last Sunday?
d. *What the students did do in the park last Sunday?
(13) a. What did the boys in white T-shirts do last Monday?
b. *The boys in white T-shirts did what last Monday?
c. *What the boys in white T-shirts did last Monday?
d. *What the boys in white T-shirts did do last Monday?

Moreover, to ensure that the vocabulary in the test was known to the participants, all the lexical items were selected from the textbooks used by the middle school students who took part in this study. In addition, the verbs in the SV questions were all in the present tense so SV-agreement could be tested. On the other hand, the verbs in the WH-questions were all in the past tense so that no application of subject–verb agreement was involved. Furthermore, we included only *do*-inversion (i.e. no *be*/*have*/modal–subject-inversion) in the test. The inclusion of only the past-tense and *do*-inversion in the WH-questions would help avoid confounding variables of SV-agreement and tense-agreement, which

are FC properties not to be tested by the WH-questions. It is also worth noting that the use of the past tense should not pose any additional difficulty to the participants for the following two reasons. First, the past tense was taught early in the first year of English classes in Chinese schools (i.e. all the participants were familiar with this tense). Second, roughly three fourths of the main verbs used in the 40 sets of sentences were simple regular verbs (e.g. *jump*/*play*/*walk*). For each irregular verb, its past-tense form was provided and explained in Chinese at the end of the sentence it appeared in.

The 320 sentences (40 × 8) used for the acceptability judgment test were grouped into four counterbalanced lists (with each list containing 80 sentences: 40 SV-questions and 40 WH-questions) to ensure that participants would not read the different versions of the same sentence in the same type. More specifically, each list contained 20 grammatical and 20 ungrammatical SV-agreement sentences, and 10 grammatical and 30 ungrammatical WH-question sentences, with half of the total 80 sentences including a local-distance modifying phrase between the subject and the verb and the other half consisting of a long-distance modifying phrase. To balance the number of grammatical and ungrammatical WH sentences, we created 20 additional grammatical WH-sentences as fillers. We ensured that these filler sentences matched their counterpart ungrammatical experimental sentences in length and structure. Then, in order to have adequate fillers for each test list, we developed 80 additional filler sentences (50 SV sentences and 30 WH sentences) with half of them being grammatical and another half being ungrammatical, hence a total of 50 SV filler sentences and 50 WH filler sentences (30 here + the 20 mentioned above). All the filler sentences were similar to the experimental sentences in length, type, and structure. Participants were randomly assigned to either list of the instrument. Altogether, each participant would read 180 sentences, i.e. 80 experimental sentences plus 100 filler sentences.

3 Procedure

The Chinese Survey Star online platform was used in this study. We sent all the participants each an email and a link to the three aforementioned instruments on the platform and asked them to complete the instruments in the following order: first the demographics survey and the proficiency test in one sitting and then the acceptability judgment test a week after the completion of the first two instruments. It took the participants about 15–20 minutes to finish both the questionnaire and the proficiency test and approximately 30–40 minutes to complete the acceptability judgment test.

In completing the acceptability judgement questions, the participants were instructed to evaluate each sentence on a Likert scale from 1 to 4, where 1 meant completely unacceptable and 4 indicated completely acceptable. There was also an additional option ‘I don’t know.’ The instructions were given to the participants both orally (in Chinese) and in writing.

On each page, only one sentence was presented. Once the participants had judged the sentence and moved onto the next page, they were not able to return to the previous pages to change any of their previous answers. In addition, the test was pseudo-randomized to ensure that the different versions of the same sentence (though in different types) never immediately followed one another on two consecutive pages. For every 3

pages, there was at least one sentence dealing with WH-questions and one concerning subject–verb agreement questions (with filler sentences included).

V Data analysis, results, and discussion

1 The proficiency test

We first tabulated the descriptive statistics regarding the basic demographics of the three groups of participants (reported in Table 2). Regarding the scoring of the English proficiency test, one point was given for a correct answer for each of the 40 questions, which means the highest possible score was 40. Then, we conducted an independent one- way ANOVA across the three groups in age, length of English study, and English proficiency. The results indicated a significant difference across the three groups in all the three variables: age ($F(2,121) = 846.31$, $p < .001$, $\eta^2 = .933$), length of study ($F(2,121) = 356.00$, $p < .001$, $\eta^2 = .855$) and English proficiency ($F(2,121) = 21.08$, $p < .001$, $\eta^2 = .258$). Post hoc Tukey's tests indicated that the college group's mean scores in all three variables were significantly higher than those of the high-school group, whose mean scores were, in turn, significantly higher than those of the middle-school group ($ps < .0001$). We then carried out a correlational test between the participants' scores of the English proficiency test, ages, and lengths of English study. The result indicated a very strong positive correlation among these three factors (between the proficiency and the age: $r = .49$, $p < .0001$; between the proficiency and the length of learning: $r = .49$, $p < .0001$; between age and length of learning: $r = .95$, $p < .0001$). Given the strong significant correlation and the fact that English proficiency has often been found to be a significant predictor (more significant than lengths of study) in English learners' grammatical performance (Martirosyan et al., 2015), only English language proficiency (hereafter simply 'proficiency') was used as a variable in the data analysis shown below. The three proficiency levels arguably resembled a longitudinal L2 English proficiency developmental pattern. Hence, entering this variable into the analysis model would enable us to test whether EFL learners' difficulty with the tested lexicogrammatical properties persisted over time across proficiency levels.

2 The acceptability judgment test

a Main conditions. First, it is necessary to note that, in our tabulation of the participants' judgment responses, the selected values on the Likert scale were treated as a binary variable,[3] i.e. as accurate or inaccurate. For the grammatically correct sentences, selections of both scales 1 and 2 were considered inaccurate while selections of both scales 3 and 4 were accurate. For the ungrammatical sentences, however, the practice was reversed, with scales 1 and 2 being accurate while scales 3 and 4 being inaccurate. This is because judging an ungrammatical sentence acceptable (scale 3 or 4) was clearly an inaccurate response while judging such a sentence unacceptable (scale 1 or 2) was an accurate response.

All the data analyses were carried out by using R (R Core Team, 2019). Generalized linear mixed-effects models were used to fit the binary data i.e. whether the participants'

responses on the acceptability judgment test were accurate or inaccurate) via the *glmer* function in the *lme4* R package (Bates et al., 2015). The independent variable in the present study was *type* of grammatical properties (subject–verb agreement vs. WH- questions with or without *do*-insertion/inversion). The scores of the participants' English proficiency test were entered into the model as a covariate. However, before entering the model, these scores were centered and scaled to overcome potential converging problems and reduce the possibility of multicollinearity of the model. The variable *type* (which includes SV-agreement and the three WH-question related syntactic properties, with SV-agreement being the reference level), the covariate *proficiency*, and the *type* × *proficiency* interaction were thus the fixed-effects factors in the model. The participants and the sentences they read (i.e. *items* as shown in the model below) were random-effect factors in the model. We followed the 'keep it maximal' rule proposed by Barr et al. (2013) when fitting the random-effects structures by including both by-participants and by-items random slopes and their intercepts for all the relevant fixed effects. We obtained *p*-values for the main effects and interactions of the two factors (*type* and *proficiency*) by using likelihood ratio tests via the mixed function.

To answer research question 1, we conducted two separate analyses to help gain a complete but also nuanced understanding of this research question: (1) an overall comparison between the participants' performance on SV-agreement and their performance on all the WH-questions combined (i.e. with their performance on all the three types of WH-questions: WH-movement, WH-*do*-insertion, and WH-*do*-inversion, collapsed) and (2) a detailed analysis comparing the participants' performance on the SV-agreement sentences with their performance on each of the three WH-questions respectively.

The results of the mixed-effects model for the overall comparison indicate that *proficiency* was a significant predictor (χ^2 (1) = 16.97, $p < .001$). With the increase of *proficiency*, the participants' accuracy also rose. However, *type* (of grammatical property) did not emerge as a significant predictor (χ^2 (1) = 2.46, $p = .117$). More importantly, a significant *proficiency* × *type* interaction was found (χ^2 (1) = 4.32, $p = .038$). Furthermore, as clearly illustrated in Figure 1, in this interaction, the effect of *proficiency* on the participants' response accuracy in the WH-questions (core syntactic properties) was larger than its effect on their accuracy in the SV-agreement (FC) questions. Hence, the increase of the participants' English proficiency led to a much steeper rise of their response accuracy in the WH-questions than in the SV-agreement questions. This difference is also shown in the summary of the model in Table 3.

The results of the detailed separate comparison also indicate that *proficiency* was a significant predictor (χ^2 (1) = 7.76, $p = .005$). In addition, *type* (of grammatical property) also had a significant main effect (χ^2 (3) = 37.01.06, $p < .001$). More importantly, a significant *proficiency* × *type* interaction was found (χ^2 (3) = 6.62, $p < .001$). The interaction between these two factors can be clearly seen in the profile plot displayed in Figure 2. The profile plot exhibited that the effect of *proficiency* on the participants' response accuracy was different across the four grammatical properties. As the participants' *proficiency* increased, their response accuracy exhibited a sharp increase in WH-movement and a slightly less steep increase in WH-*do*-inversion, but much less increase in SV-agreement and little increase in WH-*do*-insertion. As shown in the summary of the model in Table 4, while one-unit increase of *proficiency* led to an increase of

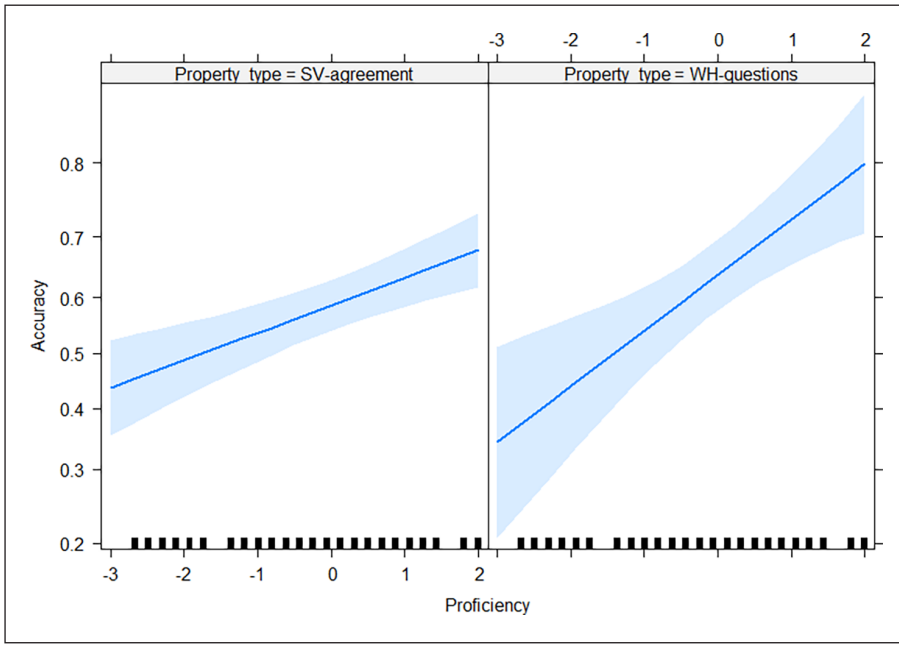

Figure 1. The interaction between *type* (SV-agreement/WH-question) and *proficiency*.

0.20 unit in the participant's response accuracy in the SV-agreement questions (which served as the reference), an increase of one unit in *proficiency* led to a much greater increase of 0.55 (0.20+0.35) unit in the participants' response accuracy in the Wh-movement questions ($\beta = 0.35$, SE = 0.09, $z = 3.75$, $p = .0002$). Furthermore, the effect of *proficiency* was larger on the SV-agreement sentences than on the WH-*do*-insertion questions (0.20 vs. 0.16 (0.20–0.04)), but not significantly larger statistically ($\beta = -0.04$, SE = 0.19, $z = -0.22$, $p = .823$). On the other hand, the effect of *proficiency* was smaller on the SV-agreement sentences than on the WH-*do*-inversion questions (0.20 vs. 0.47 (0.20+0.27)), but not significantly smaller either ($\beta = 0.27$, SE = 0.20, $z = 1.39$, $p = .166$).

In short, while the results of the overall comparison indicate clearly that SV-agreement was significantly more difficult than WH-questions, the results of the detailed comparison appear to suggest that WH-*do*-insertion was the most difficult followed closely by SV-agreement in absolute terms of the participants' response accuracy, but the lack of statistically significant difference between the two suggests, however, that the two properties were essentially of equal difficulty. Similarly, while WH-movement was the easiest of the four followed closely by WH-*do*-inversion in absolute terms of accuracy, these two were also of equal difficulty due to the lack of significant difference between the two. In other words, the results show that SV-agreement and WH-*do*-insertion were much more difficult than WH-*do*-inversion and especially WH-movement. We will further discuss these results by focusing on why WH-*do*-insertion was as difficult as SV-agreement so as to offer a valid overall interpretation in Section V.3.

Table 3. Summary of the mixed-model fitted to the acceptability judgment test (an overall comparison).

	Estimate	Standard error	z value	Pr(>|z|)
Intercept	0.35	0.09	3.69	.0002
PropertyWH-questions	0.22	0.14	1.57	.116
Proficiency*	0.20	0.05	3.99	< .0001
PropertyWH-questions: Proficiency	0.21	0.10	2.10	.0361

*Proficiency effect on SV-agreement as reference level.

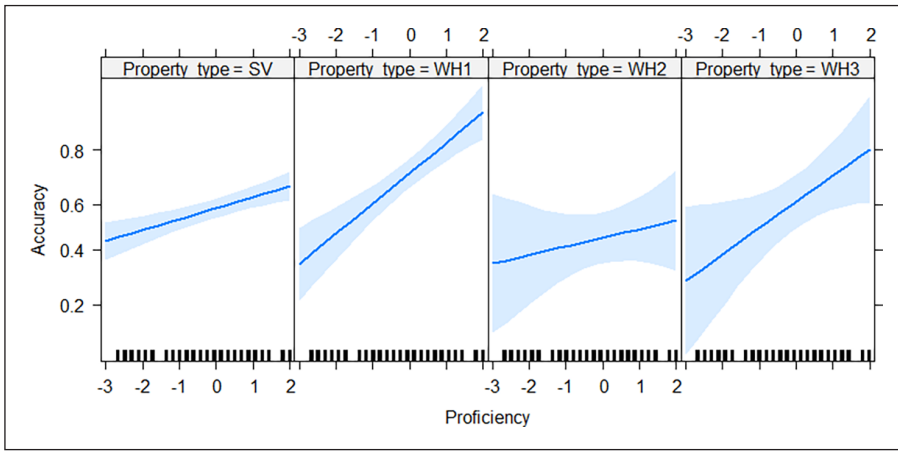

Figure 2. The interaction between *property type* and *proficiency*.
Notes. SV = SV-agreement; WH1 = WH-movement; WH2 = WH-do-insertion; WH3 = WH-do-inversion.

Table 4. Summary of the mixed-model fitted to the acceptability judgment test: (a detailed comparison).

	Estimate	Standard error	z value	Pr(>|z|)
Intercept	0.34	0.09	4.00	< .0001
TypeWH1	0.63	0.14	4.59	< .0001
TypeWH2	−0.53	0.23	−2.34	0.019
TypeWH3	0.11	0.23	0.48	0.633
Proficiency (SV as reference)	0.20	0.05	3.96	< .0001
TypeWH1: Proficiency	0.35	0.09	3.75	0.0002
TypeWH2: Proficiency	−0.04	0.19	−0.22	0.823
TypeWH3: Proficiency	0.27	0.20	1.39	0.166

It is important to note that in the above analyses, the SV-agreement and the WH-movement properties each included both a grammatical and an ungrammatical question in each set. To ascertain whether the participants performed significantly

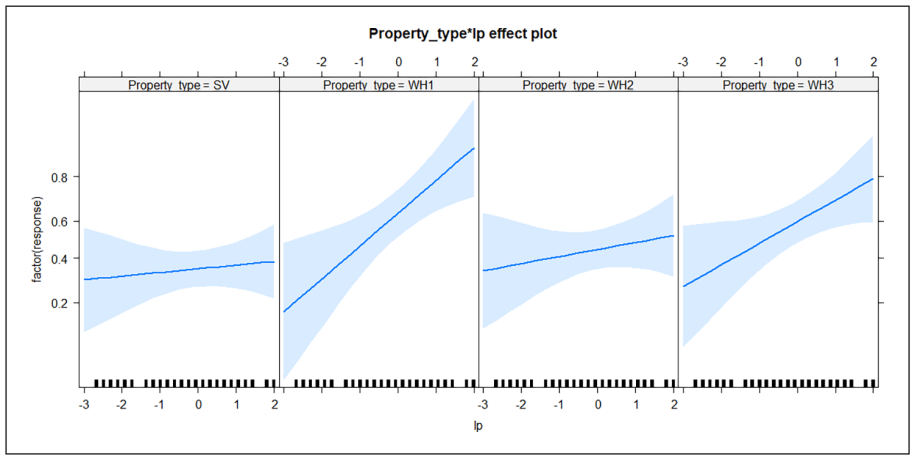

Figure 3. Effect of *property type* and *proficiency* on ungrammatical items only.

differently between the two different question types, we fitted a mixed-effects model comparing their overall scores on the grammatical questions with their overall scores on the ungrammatical questions. The result indicated that the participants indeed performed significantly worse in the ungrammatical sentences than in the grammatical sentences (SV-agreement sentences: $\beta = -2.10$, SE $= 0.35$, $z = -5.95$, $p < .0001$; WH-movement sentences: $\beta = -1.09$, SE $= 0.35$, $z = -3.11$, $p = .002$). Then, to control the effect of the grammatical test items, we excluded these items from our data analysis, i.e. we calculated the participants' response accuracy in only the ungrammatical questions of both the SV agreement and the WH-movement test items along with the WH-*do*-insertion and WH-*do*-inversion items (all being ungrammatical as noted earlier) and refitted a mixed-effects model using the same procedure as described above. The results of this mixed-effects model (illustrated in Figure 3 with summary information reported in Table 5) were quite similar to those from the analysis of all the test items including the grammatical ones (shown in Figures 1 and 2 and Tables 3 and 4 above). As in the results from all the test items, a significant interaction between *proficiency* and *type* was also found here (χ^2 (3) $= 15.39$, $p = .002$), indicating that the effect of *proficiency* was different for the different types of lexicogrammatical properties. The interaction between these two factors can be clearly seen by the profile plot displayed in Figure 3. Similarly, as in the results involving all the test items, WH-movement and WH-*do*-inversion here also enjoyed a much higher response accuracy and also a much higher increase of accuracy across proficiency levels than SV-agreement and WH-*do*-insertion. In contrast, the latter two properties here also showed not only a lower accuracy but also an extremely low accuracy increase across proficiency levels.

One other point worth noting is that, as mentioned in Section IV.2, the creation of the sentences for the four subconditions of SV-agreement sentences resulted in the existence of two types of WH-questions (i.e. subjects without a post-nominal modifier vs. subjects with a post-nominal modifier as shown in examples 12–13 in Section IV.2). To make sure that these two different sentence types did not cause any difference in the participants'

Table 5. Results from the mixed-model fitted to the acceptability judgment test involving ungrammatical items only.

	Estimate	Standard error	z value	Pr(>|z|)
Intercept	−0.64	0.19	−3.36	0.000777
TypeWH1	1.22	0.21	5.93	< .0001
TypeWH2	0.42	0.16	2.63	0.008
TypeWH3	1.03	0.19	5.47	< .0001
Proficiency (SV as reference)	0.08	0.18	0.44	0.6615
TypeWH1: Proficiency	0.64	0.19	3.30	0.001
TypeWH2: Proficiency	0.07	0.15	0.50	0.6165
TypeWH3: Proficiency	0.39	0.18	2.23	0.026

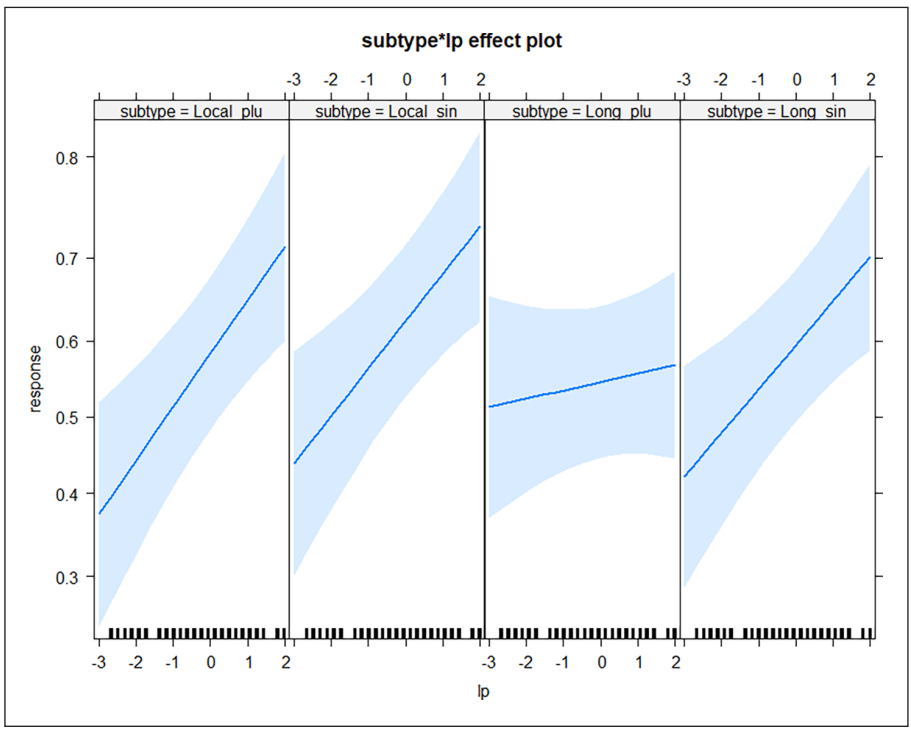

Figure 4. Effect of *proficiency* and *subtype* (i.e. the subconditions of SV-agreement).

performance on the judgement test, we fitted a mixed effects model using *type* (subjects without a post-nominal modifier vs. subjects with a post-nominal modifier) as the fixed-effects factor. The results indicated no significant differences in the participants' performance between the two subconditions ($\beta = 0.07$, SE $= 0.11$, $z = 0.61$, $p = .54$), suggesting that the two different WH-question types were not a potential cause for the

previously reported differences in the participants' response accuracies across the lexico-grammatical properties and proficiency levels.

b Subconditions. As mentioned in our description of the judgment test instrument in Section IV.2, for SV-agreement questions, there were four subconditions (singular vs. plural subjects and local- vs. long-distance agreement). To test the effects of the subconditions of SV-agreement, we fitted a generalized mixed-effects model on the participants' response accuracy in the four SV-agreement subconditions using the same statistical procedure as described above. We treated *proficiency* and *subtype* (i.e. the subconditions of SV-agreement) as the fixed-effects factor and the subjects and the items (i.e. sentences) as random-effects factors. The results indicated that *proficiency* had a main effect (χ^2 (1) = 15.02, $p < .0001$). An increase in language proficiency was accompanied by an increase in the participants' response accuracy to all the subtypes of the SV-agreement questions. However, *subtype* did not emerge as a significant predictor (χ^2 (3) = 1.29, $p = .73$). It means that, without taking into consideration the effect of the participants' language proficiency, the participants' performances in the four subconditions were quite similar.

However, a significant *subtype* by *proficiency* interaction effect was found (χ^2 (3) = 8.91, $p = .03$), indicating that participants' performance in the four different subtypes of the SV-agreement questions was dependent on the effect of *proficiency*. The interaction between these two factors can be clearly seen by the profile plot displayed in Figure 4: while the participant's accuracy increased quite steeply in the local-singular, local-plural, and long-distance-singular conditions, it rose little in the long-distance-plural condition with the increase of their English proficiency as is shown in the summary of the model in Table 6. The participants' response accuracy was quite similar among the other three subconditions (ps > 0.57). The result that the long-distance-plural SV-agreement subcondition was found significantly more difficult than the other three subconditions echoes those of Jensen et al. (2020) where the long-distance-plural subcondition was also shown to be the most difficult one of the four subconditions, significantly more than the rest three.

3 General discussion

In this section, we first discuss the results concerning research question 1: whether the SV-agreement FC property persistently poses significantly more difficulty than the three core syntactic properties involved in WH-question formation. This question directly tests Slabakova's BH and is hence the focus of our discussion. Then, we will discuss the results regarding research question 2: whether the four subconditions of SV-agreement present different levels of difficulty.

For research question 1, as shown in the previous section, the results from the general comparison analysis for this question indicate a clear yes answer that the SV-agreement FC property was persistently more difficult than the three core syntactic properties related to WH-questions, hence supporting the BH. However, the results from the detailed comparison analysis do not appear, at first glance, to be entirely or crystal clear: whereas the SV-agreement FC property was indeed persistently more difficult than the

Table 6. A summary from the mixed-model fitted to the acceptability judgment test on subconditions of SV-agreement.

	Estimate	Standard error	z value	Pr(>|z|)
Intercept	0.34	0.21	1.64	0.100
SubtypeLocal_singular	0.17	0.29	0.60	0.549
SubtypeLong_plural	−0.15	0.29	−0.53	0.597
SubtypeLong_singular	0.05	0.28	0.16	0.873
Proficiency (SubtypeLocal_plural as reference)	0.28	0.08	3.76	0.0002
SubtypeLocal_singular: Proficiency	−0.03	0.09	0.37	0.713
SubtypeLong_plural: Proficiency	−0.24	0.09	−2.69	0.007
SubtypeLong_singular: Proficiency	−0.05	0.09	−0.57	0.570

Notes. Local_singular = Local-distance singular subject, Long_singular = Long-distance singular subject, Long_plural = Long-distance plural subject.

Wh-movement and WH-*do*-inversion syntactic properties respectively, SV-agreement was not persistently more difficult than WH-*do*-insertion, a result that seems to contradict the BH. Yet, this seeming contradiction will not hold if we recall and look closely at the fact mentioned in the background section (Section II.3) that the application of this syntactic property actually involves the application of the functional features of SV-agreement and tense inflection. In other words, the high level of difficulty involved with WH-*do*-insertion might have resulted from the accompanying simultaneous application of FC features, rather than from the application of WH-*do*-insertion itself. Hence, when this factor and the results from both the general and the detailed analyses are considered together, it becomes clear that the SV-agreement FC property was persistently more difficult than the three WH-question related syntactic properties. This finding is similar to that of Jensen et al.'s (2020) study where SV-agreement was found to be persistently more difficult for Norwegian learners of English to grasp than word order (i.e. verb movement) in declarative sentences. Furthermore, their study also found that in the learning of the syntactic property of verb movement (word order), auxiliary verbs appeared to be more difficult to grasp than lexical verbs.

While the BH appears to constitute the main reason for the persistent difficulty in the acquisition of SV-agreement and WH-*do*-insertion with its simultaneous application of FC features, some of the SLA variables discussed earlier (e.g. salience and frequency) might have also been contributing factors, a point that was also explored by Jensen et al. (2020) and should be worth discussing here as well. In terms of salience, since our judgment test was done in writing, two aspects of perceptual salience should be considered: 1) the degree of perceptual noticeability of a feature and 2), in some cases, the presence/absence of a feature (i.e. whether it is visible at all). Of the four properties, SV-agreement and WH-*do*-insertion have lower perceptual salience than WH-*do*-inversion and WH-movement, with the latter two boasting high salience. This is because while the latter two properties require a movement from one position to another in a sentence, to the very front of a sentence in the case of Wh-movement (a fact that might make it more salient than WH-*do*-inversion), SV-agreement and WH-*do*-insertion entail only the

insertion of a morpheme, the unstressed bound morpheme *s* in the case of SV-agreement and the auxiliary *do* in WH-*do*-insertion. What makes SV-agreement and WH-*do*-insertion even less salient is the fact that before either the *s* or *do* is added by a speaker/writer, it is not present in the sentence, i.e. not visible at all. These varied degrees of salience among the four properties might explain why the results of the analysis of only the ungrammatical questions showed the order of difficulty level as SV-agreement > WH-*do*-insertion > WH-*do*-inversion > WH-movement (although the difference between the first two and that between the last two were not statistically significant).

Frequency might also have been a factor that interacted with the variance of salience mentioned above. Although SV-agreement is highly frequent and hence should be easy to learn, the effect of its high frequency is likely negated by its extremely low perceptual and acoustic salience. Of the three WH-question related syntactical properties, WH-movement is clearly the most frequent one because all WH-questions require a Wh-word while not all WH-questions require *do*-insertion/inversion. The high salience and higher frequency of Wh-movement help explain why it is the easiest property of the four. Then, how do we explain why WH-*do*-insertion is more difficult than WH-*do*-inversion when the two share essentially the same frequency (for the inserted do in WH-questions should always be inverted with the subject)? There are three plausible reasons. The first is the higher salience of WH-*do*-inversion mentioned above, for WH-*do*-inversion requires that the auxiliary move from one position to another, which is likely more noticeable than the insertion of *do*, a morpheme/word with only two letters that is absent before insertion (hence not very salient perceptually). Second, according to the derivational complexity theory (also known as derivational theory of complexity or DTC) that the more rules are required to be applied in the formation of a linguistic structure, the more difficult the application becomes (Fodor, 1974), WH-movement and WH-*do*-inversion, each involving the application of only one rule (movement), are thus easier than WH-*do*-insertion, which requires the application of not only insertion but also SV-agreement and tense inflections. Third, the salience of WH-movement and WH-*do*-inversion might have entrenched the inversion movement rule more than the insertion rule in the learner's consciousness and L2 system so that whenever they see an inserted auxiliary *do*, they generally remember the need of its inversion. In other words, in forming WH-questions, Chinese EFL learners are more likely to forget to insert *do* than to forget to invert an inserted *do*.

Finally, instruction does not appear to have had much influence, for, as mentioned earlier, Chinese EFL students generally receive more instruction on SV-agreement than on WH-question formation, but such instruction did not appear to have helped the participants in their performance on the SV-agreement questions. On the one hand, this lack of instructional effect might support the BH in that functional categories are not particularly conducive to instruction. On the other hand, such lack of instructional effect might have been caused by inappropriate or ineffective instructional practices.

Now, we turn to the results regarding research question 2: whether the four subconditions of SV-agreement present different levels of difficulty. The answer to the question, based on our results, is yes because the long-distance-plural condition is found to present significantly more difficulty than the other three subconditions, a finding also reported by Jensen et al. (2020). How do we account for this result, i.e. why long distance+plural

subjects cause L2 learners more difficulty in SV-agreement marking? Drawing on previous research, Jensen et al. (2020) mentioned, as the main reason, the extra processing challenges that the two linguistic features might engender, including the extra processing burden that increased structural complexity (i.e. a plural subject + a longer modifying phrase compared with a singular subject + a shorter modifying phrase) tends to create (Foote, 2011) and the processing interference that long-distance intervening modifiers may cause, a factor that is further modulated or constrained by working memory and other linguistic factors (Alemán Bañón et al., 2014; Alemán Bañón et al., 2017; Cunnings, 2017; Jiang, 2004; Keating, 2010; Lopez-Prego and Gabriele, 2014).

For example, with the use of ERPs (event-related potentials), Alemán Bañón et al. (2014) and Alemán Bañón et al. (2017) investigated L1 English speakers' processing of both number agreement (nouns–adjectives and demonstratives–nouns) and gender-agreement in L2-Spanish in comparison with that of native Spanish speakers. Their results show that L1–L2 similarities have a facilitating effect in L2 agreement processing and that the distance between the agreeing elements affects even advanced L2 learners' and native speakers alike in their processing of agreement structures. Lopez-Prego and Gabriele (2014) examined the impact of task demands (e.g. speeded grammatical judgment task vs. untimed judgement task) on both native and nonnative Spanish speakers' processing of number and gender agreement. Two of their findings were (1) language proficiency was often a significant factor for the nonnative speakers' performance in both tasks and (2) nonnative speakers performed better with number-agreement than gender agreement (the latter being a property absent in their L1). The results of the various studies all point to the complexity in L2 acquisition of agreement.

It is also necessary to note that Jensen et al. (2020) did not explain why long-distance and plural subjects each alone did not lead to significantly more difficulty. We would like to add that, based on the results from our study, while the presence of only one of the two linguistic features in a sentence may increase the level of difficulty in L2 SV-agreement marking, the increase may not be significant. However, when the two features appear together in a sentence, they will jointly engender significantly more difficulty.

VI Conclusions: Main findings and research implications

Via an experiment, this study has tested the BH and produced some complex results. First, the comparative analysis of the persistent difficulty levels between the FC property of SV-agreement and the three core syntactic properties for WH-question formation shows that, in general, the FC property was indeed persistently and significantly more challenging than the three syntactic properties, providing supporting evidence for the BH, just as Jensen et al.'s (2020) did. However, the results also reveal that the application of a syntactic property may sometimes simultaneously involve the application of FC features as in the case of WH-*do*-insertion, making the syntactic property particularly difficult to apply, far more difficult than a syntactic property whose application does not simultaneously involve FC properties. Furthermore, varying contexts in which a given FC property appears may also affect the difficult level in applying the property as demonstrated by the results from the comparison of the effects of the four subconditions of SV-agreement on the participants' success in marking this agreement.

Furthermore, although item frequency and salience, L1–L2 differences, and instruction have all been found in previous research to be important factors in determining the difficulty level of a lexicogrammatical property, in this study, only salience and frequency appeared to be influencing factors because, as noted earlier, all the four English grammatical properties tested in this study are absent in Chinese and the participants actually received more instruction on SV-agreement (although the ineffectiveness of instruction might have been due to inappropriate instructional practices). In addition, as reported above, salience seems to have a higher impact than frequency.

Finally, while these complex results shed some new light and provide support for previous research findings regarding the BH and several important SLA variables (such as frequency and salience), they should, however, be interpreted with caution for the following reasons. First, only one functional morphological property and three core syntactic properties were examined. Second, the participants of the study were all from the same L1 background. Hence, more studies will be needed to test the BH by involving the learning of different morphosyntactic features by EFL/ESL learners of different L1 backgrounds or by learners of different L2s. In addition, as suggested by Jensen et al. (2020), it will be important for future research to compare the learning of functional morphology with the learning of other linguistic domains, including semantics and the interface between syntax and discourse. In short, the present study has made contribution to a topic that merits more research in the future that should involve a broader variety of L2 languages, grammatical properties, contexts, and methodologies.

Acknowledgements

We would like to thank *Second Language Research* editor, Professor Montrul, and three anonymous reviewers for their very helpful constructive comments on the earlier versions of this article. We also want to thank the students who participated in this study.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Social Science Foundation of China, project number 22BYY096.

ORCID iDs

Shiyu Wu https://orcid.org/0000-0003-1043-506X

Dilin Liu https://orcid.org/0000-0002-8290-3734

Supplemental material

Supplemental material for this article is available online.

Notes

1. It is necessary to note that there have been several other studies related to Chinese EFL/ESL learners' grasp of WH-movement, but they focused exclusively on a very small aspect of the issue not directly relevant to the present study: the understanding of the Subjacency Principle, a UG principle that permits WH-movement out of only one bounding category, i.e. not across bounding categories as shown by the impermissible movement in *'Who did Sam claim that had stolen his car?' (Juffs, 2005; White and Juffs, 1998). In these studies, Chinese EFL/ESL learners performed similarly to native English speakers in their rejection of WH-sentences, showing the availability of the Subjacency UG Principle for EFL/ESL learners whose L1 does not involve WH-movement.
2. What Jensen et al. (2020) reported were the total frequency numbers of the present-tense verb and 3rd person present-tense uses in the 520 million-word version of COCA. For easier comparison purposes and per the typical frequency reporting practice in corpus linguistics, we have computed/reported the PMW frequencies.
3. We treated these values as a binary variable because Likert scale data may be coded either as ordinal (which is more common) or binary data depending on one's research purpose and the binary coding has the advantages of being easier to compile, calculate, and process (Grassi et al., 2007). Furthermore, we also run the statistical analysis with the data coded as ordinal data and the overall results were identical with a few miniscule differences in actual values.

References

Adger D (2002) *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Alemán-Bañón J, Fiorentino R, and Gabriele A (2014) Morphosyntactic processing in advanced L2 learners: An event related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research* 30: 275–306.

Alemán Bañón J, Miller D, and Rothman J (2017) Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43: 1509–36.

Baker MC (2008) *The syntax of agreement and concord*. Cambridge: Cambridge University Press.

Barr DJ, Levy R, Scheepers C, and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78.

Bates DM, Maechler M, Bolker B, and Walker S (2015). *lme4: Linear mixed-effects models using Eigen and S4*. Available at: http://CRAN.R-project.org/package=lme4 (accessed September 2022).

Borer H (1984) *Parametric syntax: Case studies in Semitic and Romance languages*. Dordrecht: Foris.

Chen L, Shu H, Liu Y, Zhao J, and Li P (2007) ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition* 10: 161–74.

Chomsky N (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky N (1995) *The minimalist program*. Cambridge, MA: MIT Press.

Christiansen MH and Chater N (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences* 5: 82–88.

Chuang FY and Nesi H (2006) An analysis of formal errors in a corpus of L2 English produced by Chinese students. *Corpora* 1: 251–71.

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.

Cunnings I (2017) Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20: 659–78.

Darus S and Ching KH (2009) Common errors in written English essays of form one Chinese students: A case study. *European Journal of Social Sciences* 10: 242–53.

Davies M (2008–) Corpus of Contemporary American English: 520 million words, 1990–2018. Available at: https://corpus.byu.edu/coca (accessed September 2022).

DeKeyser R (2005) What makes learning second-language grammar difficult? A review of issues. *Language Learning* 55: 1–25.

DeKeyser R (2007) *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press.

Ellis NC (2002) Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143–88.

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. 3–24.

Ellis N (2016) Salience, cognition, language complexity, and complex adaptive systems. *Studies in Second Language Acquisition* 38: 341–51.

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.

Fodor J (1974) *The psychology of language*. New York: McGraw-Hill.

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: A bidirectional study of learners of English and Japanese. *Studies in Second Language Acquisition* 31: 371–402.

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: 1–50.

Grassi M, Nucera A, Zanolin E, et al. (2007) Performance comparison of Likert and Binary formats of SF-36 version 1.6 across ECRHS II adult populations. *Value in Health* 10: 478–488.

Hawkins R and Chan CYH (1997) The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research* 13: 187–226.

Haznedar B and Schwartz BD (1997) *Are there optional infinitives in child L2 acquisition?* In: Hughes E, Hughes M, and Greenhill A (eds) *Proceedings of the 21st Annual Boston University Conference on Language Development*. Somerville. MA: Cascadilla Press, pp. 257–68.

Hwang SH and Lardiere D (2013) Plural-marking in L2 Korean: A feature-based approach. *Second language Research* 29: 57–86.

Jensen IN, Slabakova R, Westergaard M, and Lundquist B (2020) The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research* 36: 3–29.

Jiang N (2004) Morphological insensitivity in second language processing. *Applied Psycholinguistics* 25: 603–34.

Juffs A (1996) *Learnability and the lexicon: theories and second language acquisition research*. Amsterdam: John Benjamins.

Juffs A (2005) The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research* 21: 121–51.

Jung YJ (2010) Syntax–phonology interface of WH-questions. *Studies in generative grammar* 20: 549–76.

Keating G (2010) The effects of linear distance and working memory on the processing of gender agreement in Spanish. In: VanPatten B and Jegerski J (eds) *Research in second language processing and parsing*. Amsterdam: John Benjamins, pp. 113–34.

Krashen SD (1981) *Second language acquisition and second language learning*. Oxford: Oxford University Press.

Krashen S (1985) *The Input Hypothesis: Issues and implications*. London: Longman.

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: 359–75.

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

Lee JFK (2012) Teaching Hong Kong L2 learners WH-questions: Using a learning study approach. *The Journal of Asia TEFL* 9: 171–97.

Lee JFK (2016) 'Why you can't ask a proper question?': The learning difficulties of Hong Kong ESL students. *RELC Journal* 47: 295–311.

Liu D and Gleason J (2002) The acquisition of the article the by nonnative speakers of English: An analysis of four nongeneric uses. *Studies in Second language Acquisition* 24: 1–26.

Liu Y, Qi R, and Di Biase B (2020) Cross-linguistic influence of L2 on L1 in late Chinese-English bilinguals: The case of subject realisation. *Journal of Second Language Studies* 3: 290–315.

Lopez-Prego B and Gabriele A (2014) Examining the nature of morphological variability in native and non-native Spanish. *Linguistic Approaches to Bilingualism* 4: 191–222.

Madlener K (2015) *Frequency effects in instructed second language acquisition*. Berlin: De Gruyter Mouton.

Martirosyan NM, Hwang E, and Wanjohi R (2015) Impact of English proficiency on academic performance of international students. *Journal of International Students* 5: 60–71.

McLaughlin B (1987) *Theories of second language learning*. London: Edward Arnold.

Montrul S (2001) First-language-constrained variability in the second-language acquisition of argument structure-changing morphology with causative. *Second Language Research* 17: 144–94.

Odlin T (1989) *Language transfer*. Cambridge: Cambridge University Press.

Packard JL (2000) *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Prawatmuang W and Yuan B (2020) Roles of positive and indirect negative evidence in L2 feature reassembly. *Journal of Second Language Studies* 3: 205–32.

Prévost P and White L (2000) Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16: 103–33.

Radford A (2004) *English syntax: An introduction*. Cambridge: Cambridge University Press.

R Core Team (2019) *R: A Language and Environment for Statistical Computing* [software]. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org (accessed September 2022).

Slabakova R (2006) Learnability in second language acquisition of semantics: A bidirectional study of a semantic parameter. *Second Language Research* 22: 498–523.

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 M, and Martínez-Adrián M (eds) *Contemporary approaches to second language acquisition*. Amsterdam / Philadelphia, PA: John Benjamins, pp. 5–28.

Slabakova R (2016) *Second language acquisition*. Oxford: Oxford University Press.

Slabakova R (2019) The Bottleneck Hypothesis updated. In Ionin T and Rispoli M (eds) *Three streams of generative language acquisition research: Selected papers from the 7th meeting*

of generative approaches to language acquisition – North America, University of Illinois at Urbana-Champaign. Amsterdam: John Benjamins, pp. 319–45.

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.

Wei X, Chen B, Liang L, and Dunlap S (2015) Native language influence on the distributive effect in producing second language subject–verb agreement. *The Quarterly Journal of Experimental Psychology* 68: 2370–83.

White L (1991) Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. *Second Language Research* 7: 133–61.

White L (2003) *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

White L and Juffs A (1998) Constraints on wh-movement in two different contexts of non-native language acquisition: Competence and processing. In: Flynn S, Martohardjono G, and O'Neill W (eds) *The generative study of second language acquisition*. Hillsdale, NJ: Lawrence Erlbaum, pp. 111–30.

Yang C (2002) *Knowledge and learning in natural language*. Oxford: Oxford University Press.