Spurious Be in Thai Learners' L2 Grammars

Pornsiri Singhapreecha
Language Institute
Thammasat University
E-mail: pornsiri@tu.ac.th

Abstract

This paper is data oriented; it attempts to determine the status of *be* preceding bare VP and NegP complements, in data obtained from written expressions of 32 Thai EFL students. As this spurious *be* has not been examined intensively and its status is crucial to understanding L2 acquisition of tense and agreement, this study conducted a systematic investigation of this issue. Having employed data related to IP and CP, it claims that spurious *be* is restricted to IP, and is inserted into Infl, in the Do-support fashion. Spurious *be* is triggered by *not* and finiteness. Supporting evidence involves the fact that spurious *be* occurs in the absence of *do*, in conjunction with uninflected lexical main verbs. It is claimed that L2 learners are not influenced by their L1, have access to functional categories available from UG, and apply insertion of *be* exclusively for IP.

1. Introduction

This study is small-scale and data-oriented. It focuses on identifying instances of inflected *be* preceding bare VP and NegP complements in Thai EFL learners' data, e.g., *(many schools) are allow mobile devices and *(public transportation) is not spread in every city*. In the grammatical counterparts, it is necessary that *are* be absent and that *is* be replaced by *does*. In the former the lexical main verb *allow,* itself, serves as the element carrying tense and number agreement; in the latter, the auxiliary *does* inflected for tense and agreement is required, not *is*. The spurious *be* data under investigation involve those that must be absent in lexical main verb declaratives and those that occur instead of *do* in lexical main verb negatives. Lexical main verb declaratives refer to sentences in which main verbs have semantic content (e.g., *allow*). In lexical main verb negatives, lexical main verbs cannot be directly negated; inflected *do*'s are resorted to and main verbs are bare. For simplicity's sake, I will use the terms

lexical main verb declaratives and negatives to refer to both types of clauses hereafter.

In the Minimalist framework (Chomsky, 1995b), inflected *do* is derived by Do-support, a language specific rule employed exclusively for the overt expression of tense and agreement in the presence of *not*. Lexical main verbs are fully inflected and affixation does not apply.[1] The inflected *be* is derived differently; it raises in a hierarchical structure to associate with tense and agreement. Thus, *do* and *be* are derived differently in theory but they both serve as overt elements for tense and agreement in the output.

The overuse of *be* in lieu of *do* in L2 data is conceivable in lexical main verb negative constructions, given that an auxiliary is obligatory and both *be* and *do* are capable of expressing tense and number agreement. Negation aside, the use of *be* in lexical main verb sentences is puzzling. In prior L2 studies, *be* in unaccusative contexts, e.g. *was happened*, has been examined intensively (Hirakawa, 1995; Oshita, 2000; Simargool, 2006). Different accounts have been proposed to account for *Be+V3 errors (also termed overpassivized unaccusative), e.g., overgeneralization of adjectival passives, the role of input, and L2 learners' interpretation of the subject as incapable of initiating the event denoted by the predicate (cf. Simargool, 2006).

While occurrences of *be* along with overpassivized unaccusatives have been accounted for, spurious *be* instances remain unresolved. The pattern of spurious *be* occurs moderately in the current study and is quite consistent across transitive and intransitive (unergative and unaccusative) verbs. Such consistency deserves a systematic investigation to yield insights into L2 development of tense and agreement. Minimalism (Chomsky, 1995b), the framework from which I developed my analysis, specifically provides theoretical views to the kinds of data to be taken into account along with viable mechanisms. In my proposal, *be* is identified as the overt element inflected for tense and agreement. It is derived by insertion of *be* in I, on a par with *do*, and is triggered by the presence of *not* and the awareness of finite tense and agreement of IP.

The remaining part of the paper is organized as follows. Section 2 presents theoretical background, previous L2 studies on *be*, L1 Thai, and research questions. Section 3 presents data obtained from 32 Thai EFL learners. Section 4 reports results in response to the research questions. I propose the analysis for the presence of *be* in main verb negative and declarative clauses in section 5 and discuss the findings in relation to SLA theory and previous studies in section 6. Recommendations for future research and a conclusion constitute section 7.

2. Background and Research Questions

This section begins with theoretical background underpinning the morphosyntax of tense and agreement in English in 2.1. This part is a basis for characterizing spurious *be* and is crucial for the upcoming analysis. I will review L2 research on English *be* in 2.2.1. As it is evident that *be* was overused with unaccusative forms, section 2.2.2 covers this area of study. In section 2.3, I will discuss Thai sentences that correspond to grammatical English counterparts of spurious *be* sentences. Research questions follow in section 2.4.

2.1 Theoretical Background

I employ the derivational approach in Minimalism (cf. Hornstein et al., 2005) in constructing IP (Inflection Phrase), i.e. sentence structure. As tense, modality, and (person and number) agreement are associated with I, to provide a comprehensive view of English IP, this section presents the derivations of sentences that contain elements that occupy I in the output. Specifically, in 2.1.1 a finite lexical main verb and *do* in lexical main verb negatives are derived, with additional *do* in interrogatives, as a guide for CP data. Section 2.1.2 shows the derivation of copula and auxiliary *be*, via V-to-I. Finally, as I also hosts modality and a hypothetical mood, section 2.1.3 discusses the derivations of modal and infinitival clauses, which complements the theoretical discussion and is helpful for the understanding of the data.

2.1.1 Lexical Main Verbs, Negatives, and Interrogatives

Prior to the detailed structure and derivation, examples of English lexical main verbs and negatives in (1) are relevant.

(1) a. John uses a mobile phone in class.
   b. John does not use a mobile phone in class.
   c. *John not uses a mobile phone in class.

Descriptively, in the lexical main verb sentence (1a), the finite verb *uses* is inflected for third person and singular, and present tense features, while in the negative sentence (1b), *does* carries these features. (The person and number feature associated with the inflection of a lexical main verb will be referred to as *phi*, hereafter.) In (1c), when *not* precedes the inflected lexical main verb *uses*, the sentence becomes ungrammatical, and English uses *do* to express the inflection.

As a basis for the derivations, a number of theoretical assumptions are in order. Firstly, IP is a structure whose I (Infl) head hosts tense, (optional) modality, and *phi* features in finite clauses. Secondly, *uses* in its fully inflected form (cf. Chomsky 1995b) enters the numeration. That *uses* is fully inflected follows from Pollock's (1989) paradigm in which English has weak Agr (i.e., tense and *phi*) features, while those in French are strong. Such a

paradigm attempts to explain why in French lexical main verb negative sentences, verbs raise into I (aka V-to-I) in overt syntax, whereas in English counterparts, verbs do not raise, and Do-support for tense and *phi* is required. Do-support is considered a language specific rule, not provided by Universal Grammar (UG), and is used as a last resort operation when a UG rule fails to apply.[2] Thirdly, following the VP-Internal Subject Hypothesis (Koopman & Sportiche, 1991), the sentential subject originates in VP, consistently with the level at which it is assigned a thematic role. Fourthly, given that English sentences require overt subjects, it is established that English has a strong Case feature, which resides in I, and is checked when the subject raises in overt syntax into the specifier position of IP (Spec,IP). Finally, strong features raise to have their features checked in a certain domain at PF (Phonetic Form) for pronunciation, while weak features do not raise at PF; they do so at the Logical Form (LF) for interpretability reasons.

The derivation of (1) is shown in (2). For clarity on the part of inflection, VP has been derived. (Prior merging operations to the VP stage are omitted.)

(2) a. VP is merged with I; IP projects
→ [$_{IP}$ [$_{VP}$ John uses a mobile phone in class]]
b. finite tense and *phi* in I attract *John* to Spec, IP; Case is checked
→ [$_{IP}$ John$_i$ [$_{VP}$ t$_i$ uses a mobile phone in class]]

As shown in (2), after the VP is merged with I, as tense and *phi* in I are weak, the verb *uses* does not raise into it. However, the sentential subject *John* raises into Spec,IP for Case checking.

With respect to negatives, (3) illustrates crucial steps.

(3) a. VP is merged with Neg; NegP projects
→ [$_{NegP}$ not [$_{VP}$ John use a mobile phone in class]]
b. NegP is merged with I, IP projects; I carries *phi* of *John*, present tense, and Case
→ [$_{IP}$ [$_{NegP}$ not [$_{VP}$ John use a mobile phone in class]]]
c. *do* is inserted into I; *do* carries the features of I
→ [$_{IP}$ do+I [$_{NegP}$ not [$_{VP}$ John use a mobile phone in class]]]
d. *John* raises into Spec,IP; Case is checked in the Spec,IP domain
→ [$_{IP}$ John$_i$ does [$_{NegP}$ not [$_{VP}$ t$_i$ use a mobile phone in class]]]

In the resources for the numeration of (1b), unlike (1a) where *uses* is fully inflected, *use* in (1b) is not, but there exist tense and *phi* in I, while *do* is not

available in the resources. After NegP, where *not* occupies its specifier, is merged with I, the tense and *phi* remain unattached (3b). As I is a functional head and is phonetically null; the I notation is not shown in (3b), but is attached to *do* in (3c) for explicitness. In (3c), *do* is inserted via Do-support to carry tense and *phi*. Subsequently, *John* raises into Spec,IP; Case is checked in this domain, and do+I becomes *does* at PF, in (3d).

       In interrogatives such as (4), *does* appears in sentence initial position. Assuming a complementizer phrase (CP), a functional category whose head C, can host a complementizer, e.g., *that* or *if*. As *if*, which introduces an indirect question, is overt, it follows that a null Q (question) feature, which introduces a direct question, can fill C as well (cf. Ouhalla, 1994). The Q feature is strong and attracts tense and *phi* from I to be checked in Spec,CP. As shown in the derivation of (4) in (5), after *do* is inserted into I, do+I raises into C overtly, resulting in a grammatical output.

(4)   Does John use a mobile phone in class?

(5)   a. VP is merged with I; IP projects
    → [$_{IP}$ [$_{VP}$ John use a mobile phone in class]]
    b. finite tense and *phi* in I attract *John* to Spec,IP; Case is checked
    → [$_{IP}$ John$_i$ [$_{VP}$ t$_i$ use a mobile phone in class]]
    c. Q is introduced, IP is merged with Q; QP projects
    → [$_{QP}$ [$_{IP}$ John$_i$ [$_{VP}$ t$_i$ use a mobile phone in class]]]
    d. Do is inserted into I; *do* carries tense and *phi*
    → [$_{QP}$ [$_{IP}$ do+I John$_i$ [$_{VP}$ t$_i$ use a mobile phone in class]]]
    e. Do+I raises into Q
    → [$_{QP}$ Does$_i$ [$_{IP}$ John$_i$ [$_{VP}$ t$_i$ use a mobile phone in class]]]

2.1.2 Auxiliary and Main Verb *Be* Sentences

       The auxiliary *be* in (6a) and copula *be* in (6b) are both inflected for tense and *phi*. The auxiliary *be* takes a VP complement *using a mobile phone in class* while the copula *be* a noun phrase *a technocrat*. I adopt Trask's (1993) definition of copula, referring to "a form linking a subject NP to a predicate which either identifies or characterizes the subject."

       In terms of theoretical assumptions, *be* originates in V and is available in the lexical resources; the form *using* enters the numeration inflected for progressive aspect, and *be*, according to Chomsky (1991/1995a), is semantically vacuous, serving grammatical purposes exclusively.[3] Pollock (1989) generalizes the V-to-I operation that applies in French lexical main verbs to English *be*. By virtue of serving a grammatical role, *be* raises from V to I to

carry tense and *phi*. V-to-I applies to *be* in copular sentences such as (6b) as well. The derivation of (6a) is displayed in (7).

(6) a. John is using a mobile phone in class.
     b. John is a technocrat.

(7) a. VP is merged with I; IP projects
     → [$_{IP}$ [$_{VP}$ John be using a mobile phone in class]]
     b. Be raises into I; *be* carries tense and *phi*
     → [$_{IP}$ be$_i$ [$_{VP}$ John t$_i$ using a mobile phone in class]]
     c. *John* raises into Spec,IP, a result of inflected *be*; Case is
         checked
     → [$_{IP}$ John$_j$ is [$_{VP}$ t$_j$ t$_i$ using a mobile phone in class]]

Essentially, *be* originates as the main verb, taking VP complement *using a mobile phone in class* as its complement. It raises into I, as in (7b), unlike *do*, which is inserted into I, as in (3c). Sentence (6b) is derived in the same fashion, assuming *John a technocrat* as VP and *be* undergoing V-to-I movement. In addition, main verb and auxiliary *have*, another semantically empty element (cf. Chomsky, 1991/1995b), undergoes V-to-I, parallel to *be*.

2.1.3 Modal and Infinitival Clauses

Sentences (8) and (9) contain a modal verb and an infinitival clause, respectively.

(8) John can use a mobile phone in class.

(9) John wants to use a mobile phone in class.

Descriptively, *use* is uninflected when it follows the modal *can* (8) and the infinitive marker *to* (9). Theoretically, I carries modality, which is expressed by mood elements, e.g., *will*, *can*, and *should*, in addition to tense and *phi.* In English, when I is specified by a certain mood such as *can,* the following verb is uninflected. The I position can be filled by the infinitive marker *to,* which marks a hypothetical clause *John to use a mobile phone in class.* When a hypothetical sense is indicated by *to*, the following verb *use* is uninflected.

Up to this point, I have shown that in English, I is essentially the position where tense, *phi*, and modality are expressed. In English lexical main verb declaratives, the verbs are inflected while in lexical main verb negatives and interrogatives, tense and *phi* appear in *do*. In auxiliary and copula *be* sentences, *be* is inflected. In modal and infinitival clauses, modals and *to* carry mood and a hypothetical sense, leaving the verbs uninflected. It can be seen that this complex morphosyntax system of English can pose problems for

L2 learners, and this study attempts to examine one of these problems, particularly the overuse of the auxiliary *be* in contexts that do not require it. As *be* has been investigated in L2 research, and is addressed as an item that is acquired earlier among functional elements, we turn to the literature on this issue next.

2.2 Research on L2 Acquisition of *Be*
2.2.1 Morpheme Ordering and Functional Categories

Be has been a subject of interest to researchers since the 70s. Bailey et al. (1974) and Larsen-Freeman (1975) investigated (free and bound) grammatical morphemes, e.g. copula and auxiliary *be*, V-ing, regular past *-ed*, third person *-s*, and possessive *-s* with second language learners, to determine if there was an acquisition order. Bailey et. al. (1974) employed the Bilingual Syntax Measure (BSM) instrument developed by Burt et al. (1973) to elicit participants' expressions of grammatical morphemes. Larsen-Freeman (1975) conducted multiple tasks along with the BSM instrument. Bailey et al. (1974) and Larsen-Freeman (1975) found high and moderate accuracy in their participants' production of V-ing, copula *be* and auxiliary *be*, and relatively low accuracy on regular past-*ed*, possessive-*s* and third person *-s*. Bailey et al. (1974) view these findings as supporting evidence for the presence of an acquisition order where copula and auxiliary *be* precede the bound grammatical morphemes, while Larsen-Freeman (1975), who administered multiple tasks, remarks that the invariant morpheme order may apply to individuals, not across tasks.

In the 90s, grammatical morphemes were subsumed by the establishment of functional categories and queries have shifted to the availability of functional categories in initial states of L2 grammars. These queries center around accessibility to Universal Grammar (UG), a human language faculty with invariant principles that apply to all languages and parameters that individual languages opt for. Questions under investigation involve whether functional categories, i.e., IP and CP, are available from UG from the onset (Epstein et al., 1996), or develop in successive steps (Vainikka & Young-Scholten, 1994, 1996), or are available from L1 (Schwartz & Sprouse, 1996). The auxiliary *be* in progressive negative and interrogative sentences, for instance, has been used as a tool to assess if L2 learners have acquired IP and CP in a study by Epstein et al. (1996). In a more recent study by Singhapreecha (2000), who investigates whether or not the acquisition of L2 structures is guided by a UG pattern of clustering, *be* and *have* in progressive and perfective negatives as a cluster for the raising of auxiliaries are employed.[4] Notably, the *be* instances in Epstein et al. (1996) and Singhapreecha (2000) are accurately inflected, consistent with the findings in Bailey et al. (1974) and Larsen-Freeman (1975). The next section is concerned with another usage of *be*, which occurs with unaccusative verbs, a condition that does not require *be* in L2 English.

2.2.2 Be in Unaccusative Contexts

There has been substantial L2 research on instances of *be* that co-occur with unaccusative verbs (Hirakawa, 1995; Oshita, 2000; Simargool, 2006). A brief note on the background of unaccusatives is worthwhile. Unaccusative is often defined in relation to unergative; in terms of the argument structure, they both are one-place predicates. They differ in that the only argument of unaccusative is internal, while that of unergative is external (Hornstein et al., 2005).[5] In Italian, unaccusative and unergative predicates are distinguished by means of auxiliaries, as shown in (10), reproduced from Hornstein et al.'s (2005, p. 106).

(10) a. Giovanni ha/*è comprato un libro
Giovanni has bought a book
"Giovanni bought a book."
   b. Giovanni ha/*è telefonato
Giovanni has/is called
"Giovanni called."
   c. Giovanni è/*ha arrivato
Giovanni is/has arrived
"Giovanni arrived."

Sentence (10a) is a transitive construction, requiring the auxiliary *ha*. If *è* appears instead, the sentence becomes ungrammatical. The predicate in (10b) is unergative, and that in (10c) is unaccusative, marked by the auxiliary *ha* and *è*, respectively. It is commonly assumed that an unaccusative structure such as (10c) is a derived one. According to Burzio's (1986) generalization, "verbs that do not have external arguments do not assign accusative Case." It follows that in (10c), *arrivato* does not assign accusative Case to *Giovanni*, which, as Theme, originates in the post-verbal position, resulting in the raising of *Giovanni* into the subject position to receive nominative Case, assigned by I in Spec, IP. From an L2 point of view, the fact that *be* appears in unaccusative contexts might lead one to think of a possibility for L1 transfer. As a hypothetical example, if an Italian learner of English produced some form of *be* with *arrived*, his/her expression would be thought of as an L1 influence.

Such a condition notwithstanding, there is counter-evidence of *be* in unaccusatives from data of learners whose L1s do not have auxiliaries in unaccusative predicates (Hirakawa, 1995; Oshita, 2000; Simargool, 2006). In Hirakawa's (1995) study, Japanese learners' performance on English passive constructions was more accurate than their performance on unaccusative constructions. In addition, although Japanese learners' judgment on unaccusatives was moderately accurate, they accepted passivized unaccusatives

such as *the snow was melted* at a greater rate than the native controls. Simargool (2006) conducted a written test, eliciting English unaccusative sentences, with 38 Thai undergraduate students. She found that most of the unaccusative verbs were passivized and *happen* was most problematic for the students.[6] Oshita (2000) analyzed passivized unaccusative data of Italian, Spanish, Japanese, and Korean speakers in the Longman Learners Corpus. Oshita proposes his approach, called Movement-Marker, to account for the data. According to this approach, the Be+V3 is considered a morphosyntactic signal for the logical subject while NP-Movement, applying to passives and unaccusatives, is absent.

Researchers have different accounts for the presence of overpassivized unaccusatives. Hirakawa (1995) disagrees with overgeneralization, and argues for L2 learners' knowledge of NP-Movement. Oshita (2000) has a different viewpoint, attributing this error type to a morphosyntactic means, which Simargool (2006) addresses as applicable. In addition, according to Simargool's interviews with student participants, overpassivized unaccusatives were used because the students thought that the sentential subjects were unable to initiate the events denoted by the predicates. The sense that the subject has to be capable of initiating the event can be tied in with Kellerman's Reasonable Entity Principle (Kellerman 1983 as cited in Oshita, 2000).[7] In this respect, L2 learners consider a subject in an unaccusative structure as incapable of being an initiator and apply the passive form as a reasonable means.[8]

In sum, evidence in the reviewed studies suggests that L2 learners of English find the copula and auxiliary *be* easy and correctly inflect it. On the other hand, *be* is used extensively in unaccusative constructions. While there are contending approaches, Kellerman's (1983 as cited in Oshita, 2000) Reasonable Entity Principle addresses the concept in line with verbal reports of L2 learners.

The next section will investigate another issue, L1 Thai, to determine whether or not Thai plays a role in the participants' production of spurious *be*.

2.3 Thai

Thai is known as a language that makes no use of tense and number inflectional morphemes. Temporal information can be conveyed by means of time adverbs or the context where an expression is contained. Number agreement is expressed in a nominal structure in the form of a classifier, proper for a given numerical noun, but in a sentential structure where the lexical main verb is inflected for tense and *phi* in a language like English, in Thai such a means does not exist. Sentences (11) and (12) feature the verbs *cháj* and *maa*, the English counterparts of which, i.e., *use* and *come* co-occur with spurious *be* in this study.

(11) khǎw cháj thoorasàp mɨɨthɨ̌ɨ naj hôŋrian bòj
      he   use   phone   mobile   in   class   often
      "He often uses a mobile phone in class."

(12) sǎahéet làk   khɔ̌ɔŋ monláphít thaaŋ ʔaakàat maa càak manút
      cause main of   pollution   means air   come from humans
      "The main cause of air pollution comes from humans."

In (11) and (12), the verbs *cháj* and *maa* are not morphologically marked. Present tense and number agreement can be interpreted from the contexts. Like in the English counterparts (shown in the translations), there are no auxiliaries in either Thai sentence.

In terms of negative statements, as in (13), the negative element *mâj* precedes the (compound) verb *damnəən.kaan*, unlike in English, where the auxiliary *does* is obligatory in the presence of *not*.

(13) rátthabaan   mâj   damnəən.kaan
      government   not   process   action
      "The government does not take action."

Based on the three Thai sentences above, lexical main verbs neither overtly express tense and *phi*, nor does an auxiliary serve this purpose.

In terms of Thai copular sentences, there are two well-known copula terms in Thai, i.e., *pen* and *khɨɨ*. Kuno and Wongkomtong (1981 as cited in Wongwattana 2015), attribute *pen* to predicational and *khii* to identificational and specificational clause types (cf. Higgins' paradigm (Higgins, 1979 as cited in Wongwattana, 2015). The one relevant to the current study is *khɨɨ*, corresponding to the English main verb *be*, as shown in (14).

(14) sǎahéet làk   khɔ̌ɔŋ monláphít thaaŋ ʔaakàat khɨɨ prittikam khɔ̌ɔŋ manút
      cause main of   pollution   means air   be behavior of   humans
      "The main cause of air pollution is humans' behavior"

In (14), *khɨɨ* serves as a linking element (i.e. copula) between the subject and its identification (or specification). In terms of auxiliary *be*, Thai has a preverbal auxiliary *kamlaŋ,* marking the progressive aspect (PROG), as (15) illustrates.

(15) phûak   khǎw   kamlaŋ   rɔɔ   rótfaj
      group   they   PROG   wait   train
      "They are waiting for the train."

Based on (14) and (15), the copula *khɯɯ* appears when the complement identifies the subject, like English *be*; however, Thai does not employ an auxiliary as part of the progressive aspect marking.

In light of the Thai-English corresponding structures, Thai is similar to English in that both use either a copula element or a lexical verb in a declarative construction. In negative and progressive constructions, Thai does not use an auxiliary while English does. The fact that many learners in this study inserted spurious *be* in lexical main verb declaratives and negatives is unlikely to result from their L1.

2.4 Research Questions

As mentioned in the introduction, I will establish that spurious *be* serves as the element for the expression of tense and *phi*, similar to *do* in lexical main verb negatives. I will also extend this *be* insertion to lexical main verb declaratives. Before the claim can be established, a number of theoretical questions revolving around the participants' knowledge of the English tense and agreement system should be addressed as follows.

Firstly, in terms of *be*, given that spurious *be* behaves like *do*, is auxiliary and copula *be* (involving V-to-I) available and grammatical? If auxiliary and copula *be* tokens are inflected on target, this suggests that the two *be*'s occur in different contexts and that spurious *be* is used specifically for lexical main verb declaratives and negatives.

Secondly, as Do-support is not an option, it is likely that the learners have not acquired the auxiliary *do*. In this respect, is *do* unavailable (or virtually non-existent) in the data? Another point involves *have*, given that *have* involves V-to-I like *be*, but is not an option for tense and *phi* expression. A question as to how often *have* occurs, compared to *be* in the data, is relevant. In other words, does *have* occur less frequently than *be*?

Thirdly, given that bare forms of verbs are required in modal and infinitival clauses, were the verbs following modals the infinitive marker *to* bare in the data? If so, this suggests that the learners are aware of the non-explicit tense and *phi* in constructions where tense and *phi* are suppressed, which is part of their knowledge of L2 tense and agreement.

Fourthly, as Do-support applies to English lexical main verb interrogatives (involving CP), are there instances of spurious *be* in interrogatives, or are they restricted to IP? The answer to this question will contribute to identifying properties of spurious *be*, in comparison with *do*.

Finally, based on the assumption that spurious *be* is a means to express tense and *phi* in lexical main verb declaratives for these learners, when tense and *phi* are obscure to them, spurious *be* should be absent and the lexical main verbs should turn up in bare forms. In this respect, a question is whether or not lexical main verbs (in the absence of spurious *be*) are bare in the data.

Answers to these questions will enable supporting evidence that is consequential to the argument for spurious *be* as the locus for tense and agreement in L2 development. I will answer the above questions in respective order in section 4. Prior to section 4, I will present data that were used to answer the above questions next.

3. Data

This section includes information about participants (3.1), Lexical Main Verb data (3.2), IP-related data (3.3), and CP data (3.4). Note that IP-related data are the elements involving the expression and suppression of tense and *phi* in the IP, and CP data are concerned with the learners' interrogative expressions.

3.1 Participants

Data were obtained from exam papers of 32 Thai students attending a fundamental English course in the first semester of 2020 at a public university in Thailand. They came from three classes (11, 15, and 6, per class, respectively) and were those who made *be* errors.[9] The students ranged in age from 18-20, and were studying in different faculties, e.g. Science and Technology, Public Health, Law, and Political Science. Eleven students from the Science and Technology and Public Health faculties received an average score of 54.5 on the English part of the Ordinary National Education Test (O-NET); the 15 Law and 6 Political Science students achieved an average of 69 on the O-NET English test.

With respect to proficiency, the students were placed by the university in the lower and upper intermediate levels and enrolled in the same fundamental English course. In terms of incorrect *be* usage, from the total number of 57, 25 and 32 were made by the lower intermediate and the two upper intermediate groups, respectively.

The data were drawn from their midterm and final examination papers. There was a section in the midterm exam paper asking test takers to form open-ended questions with *wh*-phrases (to be discussed in 3.4). In the final exam paper, they were asked to write a paragraph of approximately 150 words on either "The Effects of Allowing Mobile Devices in Class" or "The Causes of Air Pollution." They were encouraged to use the grammatical patterns learned in class and the organization of cause and effect. On average, a paragraph consisted of 8-10 sentences. Overall, there were 32 instances of CP, and 333 IP-related and lexical main verb elements. Each participant produced an average of 10 IP-related and lexical main verb elements.[10]

3.2 Lexical Main Verb Data

The first set of data came from declarative sentences containing lexical main verbs. As spurious *be* is likely to be a placeholder for tense and *phi*, one would expect lexical main verbs in sentences where spurious *be* is absent to display in bare forms. There is a potential toward this trend. I will show data by item and by subject in Tables 1 and 2.

In terms of data by item, as shown in Table 1, there were 78 lexical main verb elements, of which 52 (67%) were bare. These forms included both accurate and inaccurate tokens, i.e., those with plural and singular subjects, respectively. The remaining 33% varied among inflected items for third person -*s*, past tense -*ed*, present perfect *have -en*, and those with inaccurate inflectional morphemes -*ing* and -*en*.

TABLE 1
Distributions of Lexical Main Verbs by Item

Form		Examples	Frequency	Percent
Bare	Correct	(17) People tend to use a lot of cars. (18) We use electricity as fuels.	29	67
	Incorrect	(19) *Air pollution damage to breathing. (20) *The student think that he can google.	23	
Inflected	-s	(21) The factory releases the contaminated smoke.	15	29
	-ed	(22) When human started to construct building	7	
	have -en	(23) The air pollution has come from many causes.	1	
Other	-ing	(24) *When we going out	2	4
	-en	(25) *once all over the world done it	1	
Total			**78**	**100**

Consider the data by subjects, divided into three proficiency groups, in Table 2. The Lower Intermediate Group and Upper Intermediate Groups 1 and 2 correspond to the three classes, noted in section 3.1.

TABLE 2
Distributions of Lexical Main Verbs by Subject

Group	Bare		Inflected			Other		Total
	Correct	Incorrect	-s	-ed	have -en	-ing	-en	
Lower Intermediate	6	11	1	2	-	2	-	22
Upper Intermediate Group 1	13	5	9	2	1	-	-	30
Upper Intermediate Group 2	10	7	5	3	-	-	1	26
Total	**29**	**23**	**15**	**7**	**1**	**2**	**1**	**78**

As shown in Table 2, the lower intermediate group committed the incorrect bare forms more than the upper intermediate groups (11 vs. 5 & 7 tokens). The lower intermediate group rarely used the -s inflected form while the upper intermediate groups employed it at a greater rate (1 vs. 9 & 5 tokens). Two notes are in order, given the by-item and by-subject data. Firstly, overall the learners used the (correct and incorrect) bare tokens as main verbs considerably more than the inflected main verbs, and accuracy increased with greater proficiency. Secondly, the -s inflected form was virtually non-existent in the less proficient group's data; its use grew with greater proficiency, like the use of correct bare forms. These data will be further discussed in relation to the fifth research question in section 4.5.

3.3 IP-related Data

IP-related data refer to elements that occupy the I head. These IP instances are relevant to research questions 1-3. In this section, the data will be presented from general to specific. Section 3.3.1 provides an overview by showing distributions of instances of *do*, *be*, *have*, modal, and infinitival clauses in the data. Section 3.3.2 reports different types of incorrect forms of *be*. Finally, section 3.3.3 specifies the two classes of spurious *be*, along with the data.

3.3.1 Overview of IP-related Data

Instances of IP data can be divided into six categories, including Copula and Aux *Be*, Incorrect *Be*, To-Infinitive, Modal, Main Verb *Have* and *Do & Don't*. These categories are virtually consistent with the elements involved in the expression and suppression of tense and *phi* discussed in 2.1. It is noteworthy that the Incorrect *Be* formed an independent category from Copula and Aux *Be*, because the grammatical versions of the Incorrect *Be* instances required *do* in negative sentences or inflected lexical main verbs, while the copula and auxiliary *be* data were all grammatical. In addition, except for

Incorrect *Be*, the data in the five categories included a very small number of ungrammatical instances, which mostly involved number agreement. To closely estimate the number of actual occurrences, these errors were not excluded.

Table 3 displays example sentences per pattern. The six categories were arranged in a manner corresponding to the order of high-low frequency counts (shown in Table 4).

TABLE 3
Categories of IP-related Elements and Examples

| Patterns	Sentence Examples
Copula & Aux *Be*	(26) a. The main cause of air pollution **is** behavior of human. b. When they **are** waiting the train c. The work **is** copied from website.
Incorrect *Be*	(27) a. *These particles and gases **are come** from car, large exhaust of trucks, factories, and volcanoes. b. *Government **are not take** action.
To-Infinitive	(28) **To prevent** this problem we should use clean energy.
Modal	(29) In your free time, you **can talk** with your friend more.
Main Verb *Have*	(30) *The use of mobile devices in class **have** many effects in a bad way.
Do & Don't	(31) a. All they **do** is spraying water in the air. b. The students **don't study** and play mobile devices in class.

In (26) *is* serves a linking role, connecting the subject to its identifying phrase; *are* is the auxiliary part of the progressive form; *is* forms a passive with the participle *copied*. Next, *are* precedes a VP, in which *come* is uninflected (27a); it also precedes *not*, which in turn, precedes a VP (27b). Sentences of the type in (27) are the focus of this study and will be discussed in detail in the next two sections. In (28), *prevent*, following *to,* is uninflected, on target. In (29), *talk* is uninflected, like *prevent* in (28), on target. In (30), despite its incorrect morphology, *have* could be thought of as the main verb. (31) illustrates the correct use of main verb *do* (31a) and its contraction form (31b).

Table 4 and Figure 1 below display percent distributions of all six categories in descending order. Overall, all forms of *be* accounted for almost half of the data (48%); infinitive and modal elements, which were on target, came second in frequency (39%). The main verb *have* and *do & don't* elements occurred at a much smaller rate (6 and 7%, respectively).

TABLE 4
Frequency Counts of IP-related Elements and Percent Distributions

Patterns	Frequency Counts	Percentages
Copula & Aux *Be*	64	25
Incorrect *Be* (*be)	57	22
Modal	54	21
To-infinitive	45	18
MV *Have*	19	8
Do & Don't	16	6
Total	**255**	**100**

FIGURE 1
IP-related Elements in Percent Distributions
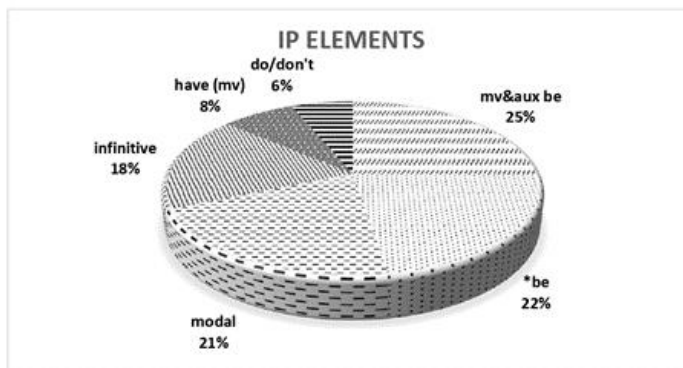
In terms of accuracy, with the exception of Incorrect *be*, approximately 90% of the instances were accurately inflected. The fact that most instances of these IP elements were on target suggests that the Thai learners have acquired IP. Specifically, they were able to produce main verb *have* and copula and auxiliary *be* (via V-to-I), and were aware of the non-overt tense and *phi* in the verbs following modals and the infinitive marker *to*. The problem area involved the minimal use of *do,* which corroborated with the overuse of *be* in contexts where *do* or an inflected verb was obligatory.

3.3.2 Incorrect *Be*

Table 5 displays four sentence patterns where incorrect *be* tokens were found, including *Be + Vbase, *Be + V3, *Be + Not, and *It's/Its.

TABLE 5
Subtypes of Incorrect Be Data

Patterns	Sentence Examples	Frequency
*Be + Vbase	(32) *So many schools **are allow** mobile devices in class.	30
*Be + V3	(33) *You **are concentrated** in your mobile device. (34) *The air pollution **has been occurred**.	12
*Be + Not	(35) *Government **are not take** action.	9
*It's/Its	(36) ***It's** can make many advantages for students. (37) ***Its** make effects in different ways.	6

Spurious *be* data are restricted to those of the *Be + Vbase and *Be + Not patterns. The ungrammaticality of sentence (33) is not related to unaccusative. The learner who produced (33) might have incorrectly chosen Be + V3 that appears in unaccusative forms such as *be concerned*. The ungrammaticality of sentence (34) involves ungrammatical unaccusatives. Neither (33) nor (34) fall within the scope of this study.

As for the *It's/Its* class in (36) and (37), the two forms, the apostrophe *it's* and the determiner *its*, were used interchangeably.[11] I classified them as Incorrect *Be* on the basis that if they show up as fully inflected *be*, ungrammaticality will arise. As *be* is not fully inflected, unlike *is/are*, the status of *It's* and *Its* is to be determined in future research, and they are not taken into account in this study.

As for (32) and (35), the *are* tokens incur ungrammaticality. In (32), the main verb *allow* is available for the expression of *phi* and the present tense. In (35), the auxiliary *does* is required, not *are*. The former resembled the latter data with regard to the presence of some form of *be*, inflected for tense and *phi*. These two sets of data are termed spurious *be* (as mentioned in section 1) and will be analyzed in this study. In the next section, I will illustrate more spurious *be* data, grouped by high to low frequency counts of the spurious *be* tokens.

3.3.3 Spurious Be

This section is concerned with spurious *be*, i.e., *Be + Not and *Be + Vbase in (32) and (35), the point under investigation. There were 38 spurious *be* tokens, including 8 Be+Not and 30 Be+Vbase instances. Fourteen and twenty-four errors were made by the lower intermediate and the upper intermediate group, respectively.[12] Next, I will discuss the two sub-categories of spurious *be* in detail.

3.3.3.1 *Be + Not

There were 9 *Be + Not tokens; *do* is required in the *be* position, exemplified in (38)-(42). In (42), the past participle form is required. Omitting (42), the spurious *be* data included the 8 tokens in this Be+Not subcategory, in the first set.

(38) *Government **are not** take action.

(39) *They **are not help** anything better.

(40) *The public transport **is not spread** in every city.

(41) *The problems of air pollution **are not disappear**.

(42) **People aren't concern about it.

3.3.3.2 *Be + Vbase

The second set of spurious *be* data involved *Be + Vbase. Table 6 displays the relevant data, grouped by high to low frequency counts of the individual co-occurring verbs.

TABLE 6
Subtypes of *Be + Vbase Data

Verb	Example	Frequency
come	(43) *These particles and gases **are come** from car, large exhaust of trucks, factories, and volcanoes.	5
have	(44) *Allowing mobile devices **is have** many good effects.	3
use, destroy, annoy, make	(45) * They **are use** technology. (46) *When human breath the air it **will be destroy** our lung. (47) *The sound of mobile devices like ringtone, text sound and other sounds **can be annoy** other students. (48) *Do you know that car **is make** a serious pollution?	8 (2 each)
allow, play, get, give, study, spend, pop up, emit, increase, happen, provide, reduce, remain, explain	(49) *Car usage **is increase** everyday. (50) *A huge amount of gas **is emit** in everyday. (51) *Building destruction is one of the problems that **are happen** in Thailand right now. (52) *People around the world **are also reduce** the use of cars.	14 (1 each)

For data clarification, two notes are in order. One is concerned with *are come* in (43); the other *will/can be* + Vbase in (46) and (47). I will establish that the *come* instances are the bare form, not the past participle and *will/can be* are modal instances of spurious *be*; as a result, they were included in spurious *be* data.

The five *be come* instances were produced by five different participants. Four of them did not exhibit passive or unaccusative forms. The fact that passives and unaccusatives hardly occurred in this group of learners suggests that all the *come* instances are the bare, uninflected form.

The *will/can be* + Vbase instances were restricted to two learners. One never made spurious *be* errors but produced four *will/can be*+Vbase instances. The other participant made one *will be destroy*, in addition to two spurious *be* errors. Given that both learners were able to produce modal (with bare verbs) and main verb *be* sentences on target, it is likely that these instances are modal forms of spurious *be*. Presumably, the learners employed *will/can be* to convey *phi* and mood (future and ability) explicitly in this particular case.

As shown in Table 6, verbs that co-occur with spurious *be* include both those taking one argument (study, (existential) have, pop up, increase, happen, emit) and two arguments (*come*, (main verb) *have, use, destroy, annoy, make, allow, play, get, give, spend, provide, reduce, remain, explain*). Given that there are numerous tokens encompassing one-place and two-place predicates, predicting the tendency of spurious *be* occurrence by means of verb types is not a viable approach.[13] In the upcoming analysis, I opt for syntactic means, i.e., insertion of *be* parallel to *do* into I, as an alternative. Since *do* that occupies I in lexical main verb declaratives raises into C in interrogatives, it is worthwhile examining the available CP data next.

3.4 CP Data

This section furnishes data in response to research question 4, involving CP. A small set of data was available in the midterm examination writing paper.[14] In the midterm writing paper, the test takers were asked to form two *wh*-questions, based on the answers given, as shown in analogous examples (53) and (54).

(53) a. Western Civilization class starts in one hour.
       b. When does Western Civilization class start?

(54) a. I'd like to cook pasta for dinner.
       b. What would you like to cook for dinner?

Sentences (53b) and (54b) were target *wh*-questions, corresponding to the information given in (53a) and (54a). Although both (53b) and (54b) involved CP, I restricted the CP data to the *when* responses, as the target response contained *does*, the issue under investigation. If some form of *be* occurred instead of *does*, this would suggest that spurious *be* applied to both IP and CP. If some form of *do* occupied C or no inversion between the subject and an auxiliary was evident, this would suggest spurious *be* was exclusively a property of IP. All 32 student participants contributed responses. These data will be discussed along with the results with regard to the other research questions in 4.4.

4. Results

Results in response to the five questions posed in 2.4 are reported in sections 4.1, 4.2, 4.3, 4.4, and 4.5. Section 4.6 presents a summary of the five answers and suggests how the current state of the learners' L2 grammars operate, based on the results.

4.1 Main Verb/Auxiliary *Be* and Spurious *Be* are Different

As shown earlier in Table 4, there were 64 tokens of main verb and auxiliary *be*, accounting for 25% in the IP data. Table 7 displays subtypes of the main verb and auxiliary *be*.

TABLE 7
Subtypes of Main Verb and Auxiliary *Be*

Subtype	Frequency Count	Percentage
NP	35	55.7
Progressive and Passive	11	17.2
Predicative Adjective (AP) & Prepositional Phrase (PP)	8	12.5
It's and That's	7	10.9
Expletive *there*	3	4.6
Total	**64**	**100**

Within this dataset, 35 tokens were *be* preceding a noun phrase; 11 tokens were part of progressives and passives; 8 tokens appeared before predicative adjectives or prepositional phrases; 3 followed expletive *there*, and the remaining 7 tokens involved *it's* and *that's*, e.g., *it's the best way* and *that's all*.

A notable point about spurious *be* is that it specifically occurred in lexical main verb negatives and lexical main verb declaratives. The data above clearly suggest that spurious *be* and the *be* in the above structures are different elements. Viewed structurally, the raising of *be* into I is available; it is correctly

inflected and the sentences are grammatical. Interestingly, the majority of the raised *be* sentences (56%) involved NP which identifies the sentential subject (termed identificational in Higgins' taxonomy (Higgins, 1979 as cited in Wongwattana, 2015)). The frequency of identificational sentences could be attributed to a grammatical pattern presented in the cause and effect writing lesson, e.g., *CO2 is one of the causes of pollution*. From a semantic view, the learners presumably attempted to preserve the raised *be* in the context of defining an entity, which is quite different from spurious *be* which has specific grammatical information.

While it was possible for all the learners to exhibit both identificational and spurious *be* sentences as pointed out above, most of them did not produce progressive and passive *be*. There were four learners from the upper intermediate proficiency group who had progressive and passive *be*; none of them exhibited lexical main verb negatives, which suggests that they employed different *be*'s; one for main verb and auxiliary and the other for lexical main verb sentences, similar to the majority of the learners.

4.2 *Do* and *Have* are Infrequent

With respect to aux *do*, there were 16 instances of *do* and *don't* elements, accounting for 6% of the IP data, as seen in Table 4. There were three tokens of main verb *do*, and only two full forms of *do* preceding *not*; the remaining eight were contracted *don't* and *didn't* and three *doesn't*. The subjects of these sentences were mostly pronouns, e.g., *if we don't start today, they didn't understand*. Omitting the main verb *do*, the auxiliary *do* occurred in contraction with *not*, with pronominal subjects. Although there were instances of auxiliary *do*, they were very small in number and restricted to some common pronominal subjects. Given this infrequent and restricted usage, it can be said that *do* was unfamiliar to the learners in the context of IP, and Do-support had not been fully acquired.

In terms of *Be + Not Vbase, most of the participants who committed it did not employ aux *do*.[15] One student clearly distinguished between main verb *do* and spurious *be*, i.e., *all they do is spraying water into the air vs. government are not take action*. This indicates that the learners employed *be* as an auxiliary in lexical main verb negatives.

With respect to main verb *have*, as shown in Table 4, there were 19 instances, accounting for 8% of the IP data. There were only two instances of *has*; *have* was used with plural and singular subjects, and 9 instances had *you, they, we, it*, while lexical nouns appeared in the remaining ones, e.g., *mobile device, students*, and *air pollution*. As normally assumed, *have* also raises to I to carry tense and *phi*. In the current data, it can be seen that *have* occurred in small number, and there were no instances of perfective negative, e.g., Have/Has not V3, which could, otherwise, be evidence supporting the raising of *have*. While

whether or not the learners have acquired the raised *have* remains unclear, the small number of *have* tokens explains why *have* was not an option for tense and *phi* agreement, unlike *be*.[16]

4.3 Bare Verbs in Modal and Infinitive Constructions

Based on the data reported in Table 4, there were 99 tokens of modal and To-infinitive, accounting for 39% of the IP data. In 98 tokens (99% of the data), the verbs following the modals and *to* were bare, uninflected, on target, as illustrated in (28) and (29), in Table 3. Generally, the infinitive marker either showed up at the beginning of a sentence or followed a lexical verb or modal.[17] In terms of modals, the modal terms included *can/can't, could, may, must, shall, should*, and *will*. This part of the data suggests that the learners were able to suppress overt tense and *phi* in the contexts where mood and hypothetical tense (marked by *to*) were at play. As spurious *be* involved tense and *phi*, the absence of spurious *be* in this set of data is conceivable.

4.4 Spurious *Be* is Restricted to IP

As noted in 3.4, there were 32 responses (1 answer per participant) of *wh*-questions introduced by *when*. The target sentence is reproduced below.

(53b) When does Western Civilization class start?

In Table 8 below, *when* appeared at the beginning as expected in all responses; fifteen answers had inversion of *does* and the subject, on target. When *did* and *will* are included, evidently, about 53% of the answers had C filled by an auxiliary. In addition to the three auxiliaries, the presence of *is in three tokens of (57) might lead one to assume the raising of spurious *be* from I to C. Despite such a potential, the raising of spurious *be*, which accounted for 9% of the data, was not a common strategy. A more common strategy, in line with L2 studies (Lee, 2008; Pozzan, 2011), was non-inversion, accounting for 37% of the data, as shown in (58) to (60). Interestingly, given that most of these learners had spurious *be* in declaratives (involving I), this *be* virtually did not show up in C.

TABLE 8
Distributions of Inversion and Non-inversion in 32 When-Questions

Subtype	Auxiliary	Example	Frequency
Inversion	does	(55) When does Western Civilization class begin?	15 (47%)
	did/will	(56) When did/will Western Civilization class begin?	2 (6%)
	*be	(57) *When is Western Civilization class begin?	3 (9%)
Non-inversion	Inflected verb	(58) *When Western Civilization class begins/begin?	6 (19%)
	*be Vbase	(59) *When Western Civilization class is begin/is begins?	3 (9%)
	will	(60) *When Western Civilization class will begin?	3 (9%)

A question may arise as to the status of *does*, as about half of the learners produced spurious *be* in declaratives while they correctly supplied *does* in *wh*-questions. Tentatively, *do* is peculiar to interrogatives, and *be* declaratives. In this case, a follow-up question is how *do* can be marked for tense and *phi* (e.g., *does*), the features unavailable in C. One alternative is insertion of *does* in C which requires an account of how *does* acquires Q feature. Another is insertion of *do* in I to acquire tense and *phi*, and the inflected *do* raises into C, which requires an account of why *be* is inserted in I as well. I leave this question open for future research.

4.5 Bare Lexical Main Verbs and Spurious *Be*
In 3.2, the lexical main verb data revealed that bare VPs were commonly used across the three groups of learners, and inflected verb forms increased slightly with greater proficiency. The prevalence of bare VPs should now be compared to the instances of spurious *be* since they both appeared in lexical main verb declaratives.

The less proficient learners' data were interesting in this respect because they seemed to show that bare forms were used when they were not certain about tense and *phi*, and the spurious *be* was used vice versa. The more proficient learners' data seemed to show optionality between the two structures, although bare forms were a more preferred option. If spurious *be* is a strategy to show tense and *phi* overtly while bare VPs do so when tense and *phi* are obscured, one expects that once the learner has decided to express the tense and *phi* by spurious *be*, he/she should not use inflection on the main verb as another option. This is the reason bare VPs are expected to co-occur with spurious *be*.

Quite uniformly across the individual participants' data, 25 subjects committing spurious *be* did not make use of inflected lexical main verbs. Examples supporting this assumption, from eight lower intermediate learners, are shown in Table 9.

TABLE 9
Spurious *be* vs. Lexical Main Verb Instances of Eight Lower
Intermediate Participants

Subject	Spurious *be*	Lexical Main Verb
1	(61) You **are use** a mobile phone in class.	none
2	(62) Allowing mobile devices in class **is have** many good or bad effects.	(73) students **use** mb devices in different ways. (74) If student **use** mb devices for support to learn…
3	(63) Government **are not take** action. (64) They **are not help** anything better.	(75) In every day life when we **going** out we **met** a lot of pollution such as water pollution when we **transported** by boat.
4	(65) The public transportation **is not spread** in every city.	(76) We going to help together such as the paper factory **use** heat
5	(66) the main cause of air pollution **is almost come** from humans	(77) bad air that we **emit** to the sphere
6	(67) So many schools **are allow** mobile devices in class (68) they **are play** game, listen music on mobile devices	none
7	(69) although human cause air pollution but **they're get** impact from air pollution too	(78) Air pollution **make** many problem to human body (79) For example, people **use** motorcycle and car for travel (80) Burning a trash from people also **impact** to environment and cause air pollution (81) If human **stop** their behavior that cause air pollution
8	(70) the vehicles **are give** a carbondioxide to the air (71) that **is make** an air pollution (72) they **are use** technology too much	(82) it **cause** the unstable of environment

As shown in Table 9, for an individual participant, when inflected *be* and V-base forms were used, either lexical main verbs were uninflected, or there were no lexical main verb data available for comparison. The forms *is* and *are* are accurate; they agree in tense and *phi* with those in I of the sentential subjects.[18]

That spurious *be* plays a tense and agreement marking role is reminiscent of Radford's (1998) remark on *do* as an aspectual marker in Early Modern English and present-day varieties of English. Examples from Shakespeare's *Macbeth* and Caribbean creoles in Harris' (1986) and Rickford's (1986) studies are reproduced from Radford's (iii) and (iv), pp. 128-129, in (83) and (84), respectively.

(83) Macbeth doth come.   (Shakespeare's *Macbeth*)

(84) He does be sick.   (Caribbean creoles)

Sentence (83), according to Radford (1998), suggests two possibilities; if *do* is a dummy element and, *doth*, meaning 'does', was inserted as a last resort, *doth* would be meaningless; what it did was maintain a four-syllable rhyming pattern, e.g., A drum! A drum! Another possibility, according to him, was to assume *doth* as an element conveying some semantic content, such as *aspect*. He brought up (84), which can be interpreted as 'He is usually sick' to elaborate this status. As *does* marks a habitual aspect, *do*, according to him, may not be entirely contentless.

Aside from the status of *do*, when one compares inflected *do* with spurious *be*, one finds that the presence of an inflected auxiliary (preceding a verbal stem) denoting tense and number agreement or aspect is conceivable. Since tense and number instantiation is associated with nominative Case checking, all of which convey finiteness (or "finite Infl" in term of Hornstein et al., 2005), finiteness can be regarded as a trigger for spurious *be*. I will use the notation [+fin] as part of the derivation to the output in the analysis.

To summarize this section, it is evident that bare forms, which obscure tense and *phi,* occurred predominantly in the presence of spurious *be*, suggesting that the learners, particularly the lower proficiency ones, spared spurious *be* for overt expression of tense and *phi*. Evidence for the target language option, i.e., inflecting lexical main verbs, exists in the higher proficiency group.

4.6 Summary and Suggestions on L2 Grammars

Prior to the analysis of spurious *be*, I will summarize the answers to the five research questions and make suggestions on the current state of these learners' L2 grammars. Firstly, main verb *be* occurred in typical copular sentences on target, while spurious *be* appeared in lexical main verb declaratives and negatives. Secondly, there was a very small number of the auxiliary *do*, and

most of the instances appeared in contraction forms. Thirdly, bare verbal instances followed modals and the infinitival marker *to* accurately, suggesting that I can host modality and *to*, in addition to tense and *phi*. Fourthly, in the lexical main verb *wh*-question data, more than half contained some form of *do*, while spurious *be* occurred minimally. Finally, lexical main verb instances (in sentences where spurious *be* was absent) were predominantly bare, regardless of number agreement.

These answers suggest that IP and CP operate in the learners' L2 grammars. In terms of the morphosyntax of I, modality and the hypothetical mood are settled; tense and *phi* remain unsettled. When main verb or auxiliary *be* is available, the learners are able to inflect it on target. In lexical main verb declaratives, the learners employ spurious *be* when they are certain of tense and *phi*, and opt for bare VPs when tense and *phi* are obscure to them. In lexical main verb negatives, as *do* is sporadic, spurious *be* is clearly the option. Spurious *be* is restricted to I; it does not raise into C.

5. Analysis

This section presents the analyses of spurious *be* in lexical main verb negative clauses (5.1) and lexical main verb declaratives (5.2).

5.1 Spurious *Be* in Lexical Main Verb Negatives

Based on the above evidence, for this group of learners, spurious *be* is the option for lexical main verb negatives while *do* was limited. In this respect, spurious *be* can be considered an element like *do*, that is inserted into I, to be associated with tense and *phi*. As a consequence, the following lexical main verb is bare, uninflected. Thus, the presence of *not* triggers *be insertion, parallel to the presence of *not* in English negatives. In addition, there is feature checking of strong N feature of I, i.e. Case, which attracts sentential subject from VP to the Spec,IP domain. Schematically, the derivation of **the public transport is not spread in every city* is shown in (85).

(85) a. VP is merged with Neg; NegP projects
→ [$_{NegP}$ not [$_{VP}$ the public transport spread in every city]]
b. NegP is merged with I; IP projects
→ [$_{IP}$ [$_{NegP}$ not [$_{VP}$ the public transport spread in every city]]]

c. *be* is introduced and inserted into I, which carries tense, *phi*, and Case
→ [$_{IP}$ be+I [$_{NegP}$ not [$_{VP}$ the public transport spread in every city]]]
d. *the public transport* raises into Spec,IP; Case is checked in the Spec,IP domain
→[$_{IP}$ the public transport$_i$ is [$_{NegP}$ not [$_{VP}$ t$_i$ spread in every city]]]

I extend *be* insertion to instances containing *will be* and *can be* in **will be destroy* and *can be annoy*, in (46) and (47). On the grounds that *be* is used to express tense and *phi* for these learners*, will be* and *can be* should be able to express future and ability mood as well. *Will be* or *can be* is a chunk which is inserted in I, yielding, e.g., **the air will be destroy our lungs*.

Notably, spurious *be* is restricted to IP; it most probably does not raise into C, as *do* is still available for question formation. While in L1 English, lexical main verbs are assumed to be fully inflected when they enter the numeration, the L2 learners in this study treat them as bare forms, and opt for *be* as the element to instantiate tense and agreement morphology.

5.2 Spurious *be* in Lexical Main Verb Declaratives

With respect to spurious *be* in lexical main verbs, I have established that it is a means through which tense and *phi* are overtly expressed. Insertion of *be* into I is a viable approach, given that bare verbal stems are obscure for finite clauses and the presence of *be* can be parallel to *do* in Early Modern English and Caribbean creoles. In terms of bare verbal forms, I consider them as items with indistinguishable number (singularity/plurality) and tense. A consequence of this is the entire clause may remain in VP, while in spurious *be* sentences, the sentential subject raises into Spec,IP. The derivation of **you are use a mobile phone in class* is shown in (86) below.

(86)   a. VP is merged with I
→ [$_{IP}$ [$_{VP}$ you use a mobile phone in class]]
b. [+fin] is instigated; *be* inserted into I, which carries tense, *phi*, and Case
→ [$_{IP}$ be+I$_i$ [$_{VP}$ you t$_i$ use a mobile phone in class]]
c. *You* raises into Spec,IP; Case is checked in Spec,IP
→ [$_{IP}$ you$_j$ are$_i$ [[$_{VP}$ t$_j$ t$_i$ use a mobile phone in class]]

Note that the instigation of [+fin] as a trigger for spurious *be*, indicated in (86b), is motivated on grounds of the absence of overt inflection in the lexical main verb tokens. In the derivation of lexical main verb negatives, finiteness is also present, but the notation [+fin] is omitted for the focus on *not* in (85).

6. Discussion

The findings can be discussed in three respects. Firstly, IP and CP are available for intermediate L2 learners and there is no L1 transfer, quite in line with Epstein et al.'s (1996) Direct Access to UG approach. Secondly, spurious *be* is correctly inflected for tense and *phi* while lexical main verbs are bare, confirming the ease vs. difficulty pattern of *be* vs. third person -s, found in

Bailey et al. (1974), and Larsen-Freeman (1975). Viewed from minimalism, the way lexical main verbs are fully inflected when entering the numeration in L1 English may not apply to L2 grammar. Evidently, a specific rule, e.g., the insertion of *be* into I to carry (finite) tense and *phi*, applies in the course of development. In addition, the finite I suggests the raising of sentential subjects from VP into Spec,IP (for nominative Case checking). Thirdly, spurious *be* is restricted to I, while *do* can occupy C. If insertion of an inflected *do* in C can be established, this suggests that L1 and L2 grammars are different; while functional categories are available, L1 and L2 learners do not entirely employ the same operations. In terms of **be* in L2 grammar, it can serve as a marker for incapable subjects in unaccusatives; in addition, evidence from this study suggests that it can convey finiteness in lieu of *do* and lexical main verbs.

7. Recommendations and Conclusion
7.1 Recommendations

This study employed naturally occurring data. Frequencies of elements used to support the analysis were uncontrolled. Some helpful data to strengthen the presence of sentential subjects such as expletive *it* and existential *there* appeared only a few times. Further studies are recommended to create controlled numbers of stimuli across IP and CP to confirm the presence of finite tense and *phi,* after spurious *be* has been found. In terms of English proficiency, data from a lower level and a higher level of proficiency can potentially reveal more insight into when spurious *be* begins and whether or not a decline in *be* and an increase in *do* and inflected lexical main verbs appear at a later state. In addition, given that *be* has been chosen as the auxiliary in lieu of *do*, an investigation into the frequencies of *be* and *do* in L1 English corpora is desirable. If *be* occurs more frequently than *do* in English corpora, there is the suggestion of the important role of input in L2 acquisition.

Pedagogically, the use of *be* where a lexical main verb is required deserves attention from EFL practitioners. On the one hand, instructions of *do* in main verb negatives should be delivered in relation to lexical main verb interrogatives. In parallel fashion, it is recommended that instructions of *be* in auxiliary and copular along with negative sentences be given in relation to interrogative sentences. By means of exposure to explicit, systematic lessons, it is hoped that L2 learners of English can distinguish between *do* and *be* and use them accurately.

7.2 Conclusion

This study investigates spurious *be,* the form of *be* preceding bare VP and NegP complements (termed spurious *be*), in L2 written expressions. *Be* has been studied in relation to other functional morphemes, or it has been a tool to determine what constitutes L2 initial states. It has also been examined

intensively in overpassivized unaccusatives, but this particular instance of *be* has not received much attention. The current study specifically fills this missing part in L2 research, with scrutiny on data and analysis. Its main aim is to identify the status of spurious *be* and how it comes about. Data from intermediate Thai EFL students, including IP-related elements, lexical main verbs and C-related tokens, have been analyzed. The following findings are revealed. Firstly, spurious *be* is inflected for (finite) tense and *phi* and is restricted to IP. Secondly, it occurs in lexical main verb and negative constructions, in conjunction with bare lexical main verb and (inflected) copular sentences, and interrogatives introduced by *do*. Based on the findings, spurious *be* is analyzed as an element similar to *do*, and is inserted into I to carry tense and *phi*.

Evidence in this study supports the availability of functional categories provided by UG and the absence of L1 transfer, partially in line with Epstein et al.'s (1996) Direct Access to UG, and the presence of an L2 specific rule, which differentiates the L2 from L1 system. It also provides insights into the roles of **be* in L2 acquisition. In addition to serving as an indicator for the incapable subject in overpassivized unaccusatives, **be* plays a functional role of expressing finiteness that would otherwise be carried by *do* and lexical main verbs. Given that the data were uncontrolled, it is recommended that further studies conduct controlled experiments to strengthen evidence utilized in this study, along with a corpus analysis to ensure the frequency of *be* in authentic English.

Notes

1. In one of the predecessors of Minimalism (Chomsky, 1991/1995a), lexical main verbs acquire tense and agreement suffixes by means of Affix-Hopping, e.g., affixes lower to lexical main verbs.

2. For instance, in French, a UG rule of raising of verbs to I (so-called V-to-I) applies, while in English such a UG rule does not apply, and Do-support rescues the derivation from crashing. I reproduce Pollock's (1989) in (i), which suggests V-to-I, below. Consider (i) in comparison with (ii), which is ungrammatical in English.

    (i) Pierre   ne   mange   pas.
       Pierre   ne   eats   not
    "Pierre does not eat."
    (ii) *Pierre eats not.

3. The auxiliary *have* in perfective aspect is another element of this type.

4. Clustering is a model of language acquisition predicted by Universal Grammar. It was originally stated to be related to the pro-drop parameter (Chomsky, 1981). This idea has generated substantial research in SLA, and has been tested with other parameters. In Singhapreecha (2000), the V-to-I parameter has been examined with respect to clustering. Given that English progressive and perfective negative sentences commonly involve the raising of aux (*be* or *have*) from V-to-I, she predicts that L2 learners who have acquired V-to-I should be able to produce both sentence types accurately.

5. According to Hornstein et al. (2005), traditionally unergative and unaccusative predicates are differentiated by the position the only argument is base-generated. In the underlying structure of unergatives, the only argument occupies Spec,VP, while in the underlying structure of unaccusatives the only argument occupies the complement of V.

6. Kanlapana and Singhapreecha (2014) examined Change of Location (CL) and Change of State (CS) unaccusatives with 70 Thai EFL learners. Based on Kim's (2007) findings, they hypothesized that Thai learners would detect CS sentences (e.g., *the building was collapsed) more accurately than CL sentences (e.g., *the envelope was arrived). Their individual verb analysis indicated that *was arrived* was easier to detect than *was collapsed*. Similar to Simargool (2006), they conducted interviews with participants and attribute the ease vs. difficulty pattern to Kellerman's (1983) Reasonable Entity Principle.

7. According to Kellerman's (1983 as cited in Oshita, 2000) Reasonable Entity Principle, learners have a tendency to treat an L2 as if it were a reasonable entity and in the absence of specific knowledge about the L2, they will maximize the systematic, the explicit, and the logical in their interlanguage.

8. In the current study, there were 5 grammatical passives, and 12 overpassivized unaccusative tokens. The overpassivized items accounted for 4.7% of the IP-related elements (12 from 255 tokens). As spurious *be* is the concern of this study, the data to be analyzed are restricted to those in main verb declaratives and negatives. The overpassivized unaccusative such as *the pollution has been occurred* in this study could be thought of as an overgeneralization, an instance of movement-marker, or an effect of incapability of the subject as an initiator, addressed in the earlier studies.

9. There were 37, 43, and 42 students in class 1, 2, and 3, respectively. The student participants who made spurious *be* errors account for 26% of the total number of students from the three classes.

10. In terms of research ethics, after the final examination, I emailed the student participants to request the use of their data in the interest of the current study and received their permission.

11. The other three tokens are in (i) to (iii).

   (i) *It's can make other people annoyed. (ii) *It's not affect only humans but also environment too. (iii)* It's might be die if you breathe it too much. In both (i) and (ii), the *-s* could be an agreement clitic. It is unclear what role *it's* in (iii) plays; if *be* can be omitted, *it's* might be a filler for the sentential subject, on a par with expletive *it*.

12. The average lower and higher proficiency learners produced 1.27 and 1.14 errors, respectively. Given that a similar frequency occurred across the two groups, proficiency may not play a role in predicting how often one produces spurious *be*.

13. One approach that is worth considering in future research involves treating spurious *be* as emphatic *do*. Baker (1995) proposes a rule which enables a tensed *do* to take a bare-stem verb phrase as a complement to account for emphatic *do* sentences. According to him, a sentence with a tensed *do* is a special-purpose finite clause, with extra stress on the finite *do*. On a par with Baker's approach to emphatic *do*, if spurious *be* is stressed in oral expressions, emphasis could be part of the account for the *Be+Vbase sentences.

14. Before the midterm examination, the instructor integrated (open-ended) *wh*-question formation into a lesson about Thai culture. In this lesson, a poster featuring the period when Thai people were encouraged to wear western clothing was used to elicit questions with *wh*-words, i.e., *who*, *what*, *when*, *where*, and *how*. The students were taught how to form (close-ended) *yes-no* and *wh*-questions with *do* and *be*.

15. There was only one instance of *don't* in an imperative, i.e., *don't even think about how all of pollution effects to the world*.

16. There were a few auxiliary *have* instances in ungrammatical unaccusatives as shown in Table 5. Due to the ungrammaticality and its association with unaccusatives, these instances were not taken into consideration.

17. Examples of *to* following a matrix verb or modal are shown in (i) and (ii). (i) They **decide to use** a steaming machine. (ii) everyone **need to keep** your mobile devices in the desk.

18. There were seven upper intermediate learners who utilized both spurious *be* and inflected lexical main verbs; this group was probably in transition to elimination of spurious *be*.

References

Bailey, N., Madden, C., & Krashen, S. D. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning, 24*(2), 235-243.

Baker, C. L. (1995). *English syntax* (2nd ed.). MIT Press.

Burt, M. K., Dulay, H. C., & Hernández-Chávez, E. (1973). *Bilingual Syntax Measure* (Restricted Edition). Harcourt Brace Jovanovich.

Burzio, L. (1986). *Italian syntax: A government-binding approach*. Reidel.

Chomsky, N. (1981). *Lectures on Government and Binding.* Foris.

Chomsky, N. (1995a). Some notes on economy of derivation and representation. In *The minimalist program* (pp. 129-166). MIT Press. (Reprinted from *Principles and parameters in comparative grammar,* pp. 417-454, by R. Freidin, Ed., 1991, MIT Press)

Chomsky, N. (1995b). *The minimalist program*. MIT Press.

Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences, 19*(4), 677-714.

Hirakawa, M. (1995). L2 acquisition of English unaccusative constructions. In D. MacLauglin & S. McEwan (Eds.), *Proceedings of the Annual Boston University Conference on Language Development 19* (pp. 291-302). Cascadilla Press.

Harris, J. W. (1986). Expanding the superstrate: Habitual aspect markers in Atlantic Englishes. *English World-Wide, 7*(2), 171-199.

Hornstein, N., Nunes, J., & Grohmann, K. K. (2005). *Understanding Minimalism*. Cambridge University Press.

Kanlapana, N., & Singhapreecha, P. (2014, May 31). *Arrive easier than collapse: Evidence from Thai EFL learners* [Paper presentation]. The 14th Annual Conference of the Japan Second Language Association (J-SLA 2014), Kwansei Gakuin University, Nishinomiya Uegahara Campus, Osaka, Japan.

Kim, J. T. (2007). Semantic aspects in L2 English overpassivization. *The linguistic Association of Korea Journal, 15*(4), 55-76.

Koopman, H., & Sportiche, D. (1991). The position of subjects. *Lingua, 85*, 211-258.

Larsen-Freeman, D. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly, 9*(4), 409-419.

Lee, S.-Y. (2008). Argument-adjunct asymmetry in the acquisition of inversion in *wh* questions by Korean learners of English. *Language Learning, 58*(3), 625–663.

Oshita, H. (2000). What is happened may not be what appears to be happening: A corpus study of 'passive' unaccusatives in L2 English. *Second*

Language Research, 16(4), 293-324.
https://doi.org/10.1177%2F026765830001600401

Ouhalla, J. (1994). *Transformational grammar: From rules to Principles and Parameters*. St. Martin's Press.

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

Pozzan, L. (2011). *Asking questions in learner English: First and second language acquisition of main and embedded interrogative structures* [Unpublished doctoral dissertation]. City University of New York.

Radford, A. (1998). Syntax: A minimalist introduction. Cambridge University Press.

Rickford, J. R. (1986). Social contact and linguistic diffusion: Hiberno-English and New World Black English, *Language, 62*(2), 245-289.

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

Simargool, N. (2006). What is happening to what was happened? *PASAA*, *39*, 49-69.

Singhapreecha, P. (2000). *The acquisition of case, tense, and agreement features: A study of Thai learners of English* [Unpublished doctoral dissertation]. City University of New York.

Trask, R. L. (1993). *A dictionary of grammatical terms in linguistics.* Routledge.

Vainikka, A, & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Turkish and Korean adults learning German. In T. Hoekstra & B. D. Schwartz. (Eds.), *Language acquisition studies in generative grammar* (pp. 265-316). John Benjamins.

Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research, 12*(1)*, 7-39.

Wongwattana, U. S. (2015). Complexities of Thai copular constructions. *Journal of the Southeast Asian Linguistics Society, 8*, 97-120.
http://hdl.handle.net/1885/95200