The overgeneration of *be* + verb constructions in the writing of L1-Malay ESL learners in Malaysia

Abdul Aziz Roslina and Zuraidah Mohd Don
University of Malaya / Malaysia

**Abstract** – This study investigates a syntactic problem in the writing of ESL learners whose first language is Bahasa Melayu or Malay. It focuses specifically on *is*, *are*, *was*, and *were* overgenerated with inflected and uninflected lexical verbs to form two primary constructions, namely *be + V* and *be + Ved* (or *Ven* in the case of strong verbs). This study aims to examine the patterns of *be* overgeneration constructions produced by the learners and determine if these are the outcome of tense and agreement marking, as postulated by Ionin and Wexler (2001, 2002). The data for the study were obtained from the Malaysian Corpus of Learner English (MACLE), a learner corpus developed by the University of Malaya. The findings reveal that uninflected verbs occur more frequently than inflected verbs in the position after *be*, which translates into higher occurrences of the *be + bare V* construction in comparison to the *be + Ved* construction. Both constructions are also found to occur more frequently with transitive verbs. The findings suggest that (i) the overgeneration of *be + bare V* is the result of agreement marking, while (ii) *be + Ved* is the outcome of assigning the tense feature. These findings suggest that the overgeneration of *be* constructions produced by L1-Malay ESL learners could be the product of a developmental aspect of language acquisition. This traces back to the system underlying the patterns of overgeneration, which is clearly made up of non-random constructions governed by very specific interlanguage grammar.

**Keywords** – *be*, overgeneration, learner, ESL, corpus

1. **INTRODUCTION**

Earlier studies highlight the inclination of ESL/EFL to use *be* in various ways not conforming to its standard forms and functions, such as the use of the *be + V* construction to mark tense and agreement (see Ionin and Wexler 2001, 2002) and *be + Ven* for expressing assertion or emphasis by overpassivisation of unaccusatives2 (see Yip 1995; Balcom 1997; Ju 2000; Oshita 2000). The same overgeneration construction has also been attested in the language data of L1-Malay ESL learners in Malaysia. Error analysis research conducted by Wee (2009), and Wee et al. (2010) on the writing of L1-Malay ESL learners in Malaysia found the *be* overgeneration construction to be one of the major types of errors

---

1 This study was funded by the Ministry of Higher Education Malaysia. Some of the findings of this study were presented at the Asia Pacific Corpus Linguistics Conference, Auckland, New Zealand, February 2012. We would like to thank two anonymous reviewers for their careful reading and valuable comments. All responsibility for errors and shortcomings is of course entirely our own.

2 Burzio (1986) divided intransitive verbs into two distinct groups: unergative and unaccusative verbs. Semantically, unaccusative verbs include verbs with non-volitional acts like *burn*, *melt*, *fall*, and *happen*, while unergative verbs entail volitional acts like *dance*, *walk*, and *work*. Both unergative and unaccusative verbs only take one argument but differ in their d-structure: unaccusative verbs take an internal argument while unergative verbs take an external argument (Park and Lakshamanan 2007).
committed by these learners in addition to errors in agreement and tense. Even more intriguing is the fact that the non-target *be* insertion was also found in the language data of ESL learners from various other L1 backgrounds, including Russian (see Ionin and Wexler 2001), Chinese (see Yip 1995; Balcom 1997; Ju 2000; Lee and Huang 2004), Spanish (see Oshita 2000; Fleta 2003), Japanese (see Oshita 2000), and Korean (see Oshita 2000; Park and Lakshmanan 2007).

This study aims to investigate this interesting syntactic problem in the writing of ESL learners whose first language is Bahasa Melayu. It focuses specifically on *is*, *are*, *was*, and were overgenerated with inflected and uninflected lexical verbs to form two primary constructions, namely *be* + *V* and *be* + *Ved* (or *Ven* in the case of strong verbs), as exemplified in (1). The main objectives are to examine the patterns of *be* overgeneration constructions produced by L1-Malay ESL learners and determine if these constructions are the outcome of tense and agreement marking, as postulated by Ionin and Wexler (2001, 2002).

(1) a. the lion is go down
b. they are help people when people in trouble. (Ionin and Wexler 2001: 84)
c. What was happened yesterday?
d. The leaves were fallen down. (Yip 1995: 136)

2. LITERATURE REVIEW

In previous studies, the overgeneration of *be* was often associated with a special category of intransitive verbs, known as unaccusative verbs. Yip (1995), in her investigation on L1-Chinese interlanguage (IL) grammar, observed that learners often overgenerate *be* in the passivisation of unaccusative verbs, which takes the form of *be* + *Ven*. She noted that the unaccusative verbs that were normally passivised belong to a particular semantic class, usually defined by change of state or location and lack of volitional control. L1-Chinese learners in an experimental study conducted by Balcom (1997) also exhibited the same tendency. When given a grammaticality judgement task, all inappropriate passive morphology involving unaccusative verbs was regarded as grammatical by the learners. They were also observed to use the construction in their spontaneous writings.

Oshita (2000) investigated a corpus of 3,362 essays of Italian, Spanish, Japanese, and Korean learners of English from the Longman Learners’ Corpus and found that the most common passivisation errors occurred with unaccusatives. They were realised in the *be* + *Ven* structure, as in *they were happened a few days ago* (see Oshita 2000: 312). The researcher provided this instance as overt evidence of the learners’ attempts to mark NP movement. In English, the passive is often marked by the movement of the argument NP from the object position to the subject position (see Oshita 2000: 316–317). The researcher argued that L2 learners of English overgeneralised this rule and extended it to unaccusative verbs in their interlanguage.

In an experimental study, Ju (2000) was also interested in inappropriate passivisation involving unaccusative verbs. Rather than applying syntactic analysis to the errors, she took into consideration the cognitive factors involved in the constructions of such errors. Ju hypothesised that the choice to passivise unaccusatives depended mostly on whether or not the learners were able to conceptualise the agents in the discourse. When an agent or cause was a part of the learners’ mental representation (externally caused events), the unaccusative verb would be passivised, while it would not be passivised when the agent was not clear (internally caused events). The findings from her forced-choice task revealed that an average of 40 to 41 percent of the overpassivisation errors were unaccusatives with externally caused events, compared to only 18 to 19 percent of unaccusatives with internally caused events, suggesting that overpassivisation of unaccusatives was influenced by the learners’ ability to conceptualise the agent. This researcher concluded that overpassivisation of unaccusative verbs was not determined entirely by the lexico-semantics of the verb and suggested that “cognitive factors play a key role in the acquisition of unaccusatives” (Ju 2000: 105).

Inappropriate insertion of *be* before a main verb in what seems to be overpassivisation has also been reported by several studies examining language data of ESL learners in Malaysia (see ee 2009; Arshad and Hawanum 2010). Unlike the overpassivisation reported in the studies reviewed earlier (Yip 1995; Balcom 1997; Ju 2000; Oshita 2000), the insertion of *be* (which is termed ‘*be* overgeneration’ in this study) by the ESL learners in Malaysia was observed to involve not only unaccusative verbs but also transitive verbs. Wee (2009), in her analysis of errors in the written language of L1-Malay ESL learners, found instances of *be* overgeneration before transitive verbs, as in *The nurse was bandaged her leg*, alongside overgeneration involving unaccusative verbs, as in *The accident was happened at Jalan Raja Laut* (Wee 2009: 355). Overgeneration of *be* involving transitive verbs was also reported in Arshad and Hawanum

---

3 Bahasa Melayu, the mother tongue of the Malay learners involved in this study, is the national language of Malaysia.
4 The formation of passives involves abstract movement transformation, through which a grammatical element occupying a given position in a phrase marker is moved to a different position, as shown in the following examples:
(i) *The dog attacked Henry.*
(ii) *Henry was attacked by the dog.*
Most passive constructions are formed with the auxiliary *be* and the –*ed* participle (Biber et al. 1999: 166).
(2010) in their corpus-based investigation of errors produced by Malaysian ESL learners with regard to the use of auxiliary be. Learners inserted be before transitive verbs, such as buy and enjoy, as in My father was bought a dog for me and I was very enjoyed the roller coaster ride (Arshad and Hawanum 2010: 169).

In addition, L1-Malay ESL learners also demonstrated the tendency to produce be + bare V, as in My family and I was go to Pulau Tioman (see Wee 2009: 355), in which the lexical verb is not inflected. This construction is similar to those cases of overgeneration reported on by Ionin and Wexler (2001) in their analysis of spontaneous speech of child L2 learners in Russia. The L1-Russian learners were found to produce the be + bare V construction, which was thought to be impaired by progressives. After a detailed examination of all the instances of overgeneration utterances, these researchers concluded that most of these utterances were not intended as progressives; rather, they had generic and stative meaning, as well as past and future meaning, which explained the absence of the present participle –ing form in the construction. The learners in actual fact did not misuse –ing; instead they inserted be to mark tense and/or agreement on the verb (see Ionin and Wexler 2001). The researchers concluded that these constructions instantiated clear attribution to poor morphological mapping, whereby learners were unable to access the appropriate affixal inflection and resorted to the use of defaults, which were the suppletive inflections.

According to Wee (2009), the insertion of be before main verbs, in particular the past tense forms was/were, by ESL learners in Malaysia could be the result of the overgeneralisation of English grammar. Learners could have interpreted was/were as a marker for past tense, assuming that if is is used to mark present tense in the present progressive, is + Ving, and was is used to mark the past progressive, was + Ving, then it would be logical to use the was/were + Ved construction to express the past tense too (Wee 2009). Arshad and Hawanum (2010) provided another explanation to this inappropriate be insertion: it could be the outcome of the learners’ confusion between copulative and auxiliary be. The learners could have treated the be verb as the main verb and used the past tense forms was/were to indicate past time.

The review of previous studies has revealed two distinct patterns of be insertion: be + Ved (or –en) and be + V constructions. The former is termed overpassivisation, which according to the literature is commonly constructed with inflected unaccusative verbs forming a structure similar to the English passive. Researchers explained that overpassivisation is the product of either the learners’ confusion about the English passive (see Oshita 2000) or their inability to conceptualise the agent of unaccusative verbs (see Yip 1995; Ju 2000). On the other hand, the latter construction, which is termed overgeneralisation of be by Ionin and Wexler (2001, 2002), interestingly occurs with unfingted or stem verbs and is believed to be the outcome of tense and/or agreement marking. Malaysian ESL learners have the tendency to produce both syntactic variants. The aim of this study is to investigate these constructions further in order to uncover the patterns in which they occur and determine what influences such constructions. In particular, this study seeks to answer the following research questions: (i) what are the patterns of be overgeneration constructions produced by L1-Malay ESL learners?, and (ii) to what extent be overgeneration constructions produced by L1-Malay ESL learners are the result of tense and agreement marking?

3. METHODOLOGY

This study adopts corpus methodology, which requires access to a digital corpus database and the use of corpus computational tools for annotating and analysing purposes. This section describes the corpus methodology employed for this study, which includes the description and explanation of the corpus data, data coding, and analysis.

3.1. Corpus data

The data for this study were obtained from the L1-Malay ESL learners corpus, a subcorpus of the Malaysian Corpus of Learner English (MACLE). MACLE was developed under the supervision of a team of researchers from the University of Malaya, Kuala Lumpur (see Knowles et al. 2006) and it consists of approximately 800,000 words accumulated from argumentative essays written by second to fourth year undergraduates. In the formation of the L1-Malay ESL learners subcorpus, all the scripts written by L1-Malay learners, which amounted to 366 scripts, were extracted from MACLE. The subcorpus consisted of approximately 198,200 words at the time the study was conducted. The size of the L1-Malay ESL learners subcorpus is deemed suitable for the purpose of investigating the use of be given that it is a high-frequency verb. According to Granger (1998: 10), learner corpora are not expected “to reach the gigantic sizes of native corpora”; first, because it is not easy to gather learner data and, second, because the optimum size of the corpus depends mostly on the linguistic investigation to be carried out (see De Haan 1992). In the case of be, this verb is used very frequently in the written register; thus, gathering sufficient occurrences of overgenerated be from the subcorpus would not be a problem.
In general, the L1-Malay learners involved in the study have had between 11 to 12 years of formal English language lessons at school and at university. Based on the average score they obtained in the Malaysian University English Test (MUET) Band 3, the learners can be considered as ‘modest users’ of English, who are described by the MUET Band 3 scoring as having “modest command of the language, modestly expressive and fluent, can use language appropriately with noticeable inaccuracy, have modest understanding of language and context and able to function modestly in the language” (Band 3 Description, Examination Council, Ministry of Education Malaysia 2012). Table 1 below summarises the learner profile.

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male: 79 (21.6%)</td>
</tr>
<tr>
<td></td>
<td>Female: 287 (78.4%)</td>
</tr>
<tr>
<td>Native language</td>
<td>Bahasa Melayu (Malay)</td>
</tr>
<tr>
<td>Semester of study</td>
<td>2nd – 4th semester</td>
</tr>
<tr>
<td>Average age</td>
<td>24 years old</td>
</tr>
<tr>
<td>Average exposure to English</td>
<td>Before university: 11 years⁵</td>
</tr>
<tr>
<td></td>
<td>During university study: 3 semesters</td>
</tr>
<tr>
<td>Average Band for MUET</td>
<td>Band 3 (Modest users)</td>
</tr>
</tbody>
</table>

Table 1. Learner profile

3.2. Data coding

Prior to the coding process, the analysis parameters were identified and set. The first step was to set the parameters for the forms of be, which include four finite forms: is, are, was, and were, inserted or overgenerated before a lexical verb to produce the nontarget be + bare V and be + Ved constructions. The second step was to set the parameters for the pre-be and post-be constituents, which include the types of subjects, types of subject predicatives, and the forms and classes of the post-be verbs. Table 2 below summarises the analysis parameters for the study.

<table>
<thead>
<tr>
<th>Item</th>
<th>Be forms</th>
<th>Subjects</th>
<th>Subject predicatives</th>
<th>Verb form</th>
<th>Verb class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>is</td>
<td>nouns</td>
<td>noun phrase</td>
<td>base form</td>
<td>transitive</td>
</tr>
<tr>
<td></td>
<td>are</td>
<td></td>
<td>adjective phrase</td>
<td>3rd singular –s, –es, –ies</td>
<td>unergative</td>
</tr>
<tr>
<td></td>
<td>was</td>
<td></td>
<td>prepositional phrase</td>
<td>past participle –ing</td>
<td>unaccusative</td>
</tr>
<tr>
<td></td>
<td>were</td>
<td></td>
<td>infinitive to-clause</td>
<td>past participle –ed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pronouns</td>
<td>that-clause</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>wh-clause</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Analysis parameters of overgenerated be and the pre-be and post-be constituents

The third step was to develop the tagsets for the manual coding procedure. The tags were developed using SGML (Standard Generalised Mark-up Language) syntax, which involved the use of a pair of balanced pair of angle brackets (<…> was used for a start tag and </…> for the end tag; see McEnery and Wilson 2001). As an example, the word student would be tagged <NP><student><NP>, where the tag ‘NP’ is used for ‘noun phrase’. All the overgenerated be forms were coded using the tag <Ovg> denoting ‘overgeneration’. The subjects were categorised into nouns and pronouns with the tags <NP> and <PN>, respectively. As for the subject predicatives, they included noun phrases, adjective phrases, prepositional phrases, infinitive to-clauses, that-clauses, and wh-clauses. These predicatives were tagged <NP>, <AP>, <PP>, <InfC>, <ThtC>, and <WhC>, respectively. As for the forms and classes of the post-be verbs, they were coded using the tags shown in Table 3 below. It is important to note that all instances of past tense and possible past participle forms of regular verbs were tagged <V-ed> for the sake of simplicity. This decision was made mainly because it was very difficult to determine if learners were forming passives (be + Ven) when they produced constructions such as was happened, are increased, or are agreed.

---

⁵ The Malaysian University English Test is a standardised English language proficiency test which consists of four components; listening, speaking, reading, and writing. It is a pre-requisite in applying for admission to all public universities and colleges in Malaysia. The test awards six band scores, from Band 1, which represents the basic users of English, to Band 6, which represents very good users of English.

⁶ In Malaysia, English as a Second Language is taught as a subject for 6 years at primary school and 5 more years at secondary school.
Transitive Unaccusative Unergative

<table>
<thead>
<tr>
<th></th>
<th>VT</th>
<th>UAC</th>
<th>UER</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd person singular forms</td>
<td>VT-s</td>
<td>UAC-s</td>
<td>UER-s</td>
</tr>
<tr>
<td>Past tense forms</td>
<td>VT-ed</td>
<td>UAC-ed</td>
<td>UER-ed</td>
</tr>
<tr>
<td>Present participle forms</td>
<td>VT-ing</td>
<td>UAC-ing</td>
<td>UER-ing</td>
</tr>
</tbody>
</table>

Table 3. Tagsets for coding forms and classes of post-be lexical verbs

The fourth step involved manually coding of the data. With the aid of a word processor, be overgeneration and the pre-be and post-be constituents were coded using the tagsets developed for this purpose. The extracts in (2) show how be overgeneration instances were coded:

(2) a. <PN>It</PN> <Ovg>is</Ovg> <Uer>depend</Uer> <PP>on the student</PP>.
   b. As to conclude, <NP>all examples</NP> <Ovg>are</Ovg> <Vt-ed>showed</Vt-ed> <ThtC>that money is the root of evil</ThtC>.

3.3. Data analysis

WordSmith Version 5 was utilised to analyse the corpus. This software for lexical analysis was developed by Mike Scott (2008). The concordance lines for all be overgeneration constructions were simply obtained by typing the selected be overgeneration instance (eg. <Ovg>is</Ovg>) at the search terms. The outputs were stored in separate files and were manually analysed for frequencies and percentages.

4. RESULTS

This section opens with some general findings as regards the distribution of be overgeneration according to the finite forms is, are, was, and were. The following subsections report the occurrences of be overgeneration according to the pre-be and post-be constituents, which include the types of subjects, subject predicatives, and the classes and forms of the main verbs in the position after overgenerated be.

4.1. Distribution of be overgeneration according to forms

Table 4 provides a summary of the distribution of be overgeneration found in the corpus. The overgeneration constructions are classified in accordance to the four finite be forms; is, are, was, and were.

<table>
<thead>
<tr>
<th>Form</th>
<th>%</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>is</td>
<td>43.64</td>
<td>0.31</td>
</tr>
<tr>
<td>are</td>
<td>45.76</td>
<td>0.23</td>
</tr>
<tr>
<td>was</td>
<td>9.09</td>
<td>0.20</td>
</tr>
<tr>
<td>were</td>
<td>1.52</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 4. Results of be overgeneration according to forms

The present forms is and are record the highest percentages of overgeneration occurrences, approximately 44 and 46 percent, respectively, while the past forms was and were record about 9 and 2 percent of the overgeneration instances, respectively. The frequencies show that the present forms are found to be overgenerated significantly more frequently (89 percent) than the past forms (11 percent). This is highly expected, as the learners were required to write argumentative essays on general topics such as ‘Money is the root of all evils’ that would not have required much use of the past tense.

As for the rates of occurrence, the form is is overgenerated at the highest rate, followed by are and was. It is interesting to note that even though was records lower instances of overgeneration, its occurrence rate is comparatively high (0.20 percent). This finding seems to confirm the possibility of learners assigning the tense feature by inserting the past form was/were, as suggested by Wee (2009). Nevertheless, the limited number of be overgeneration occurring with the past forms was and were, ie. 6.55 and 1.03 percent respectively, makes it difficult to support such a conclusion. Extracts (3) below are samples of overgeneration constructions involving the forms is, are, and was:

Extracts (3)
(3) a. *Everyday we heard that crime is always happens.* (B0034-05)
b. *… doing the bad things to get money, this is means use money to get more money.* (B0007)
c. *They are only take care about themselves, ignore other people because don’t want to threat their safety.* (B0043-05)
d. *There are one way communication so that lecturer do not know the students are understand or not.* (B0038-05)
e. *When Gulf War was begin, the world economic also become down.* (B0016-05)

4.2. Pre-be and post-be constituents in the be overgeneration constructions

4.2.1. Types of subjects

Subjects are divided into two main categories; noun subjects (NP) and pronoun subjects (PN). Table 5 below summarises the distribution of overgenerated be according to the types of subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Noun Subject (NP) %</th>
<th>Pronoun Subject (PN) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>is</td>
<td>20.91</td>
<td>22.73</td>
</tr>
<tr>
<td>are</td>
<td>19.09</td>
<td>26.67</td>
</tr>
<tr>
<td>was</td>
<td>4.55</td>
<td>4.55</td>
</tr>
<tr>
<td>were</td>
<td>1.21</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>45.76</td>
<td>54.24</td>
</tr>
</tbody>
</table>

Table 5. Distribution of types of subjects in the be overgeneration constructions

There appears to be a higher proportion of PN subjects (54 percent) preceding the overgenerated be in comparison to NP subjects (46 percent). As for individual be forms, for both is and are, PN subjects are more common than NP subjects. The same pattern, however, is not observed with overgeneration involving the past forms was and were. With overgenerated was there appears to be an equal percentage of occurrences of PN subjects and NP subjects. As for were, the occurrences of NP subjects exceed those of PN subjects. In general, be is found to be overgenerated with both PN and NP subjects, but PN subjects are found to be used more frequently as the subjects of overgenerated be in comparison to NP subjects.

Several studies on ESL learners’ variability in the use of be, especially involving the omission of be, associate covert be to the types of subject NP (see, for example, Tode 2003, 2007; Wilson 2003; Herat 2005; Pine et al. 2008; Roslina and Mohd Don). Roslina and Mohd Don (2013), for instance, concluded that omission of be was determined by the types of subject NP. In their investigation of the patterns of be omission by L1-Malay ESL learners, these researchers found a higher tendency for be omission after NP subjects than after PN subjects, due to the way be verbs were acquired. Following Wilson (2003) and Pine et al. (2008), they explained that be verbs were acquired much in the same way learners acquired lexis, that is, in chunks of subject-be combinations as in he’s, you’re, and I’m, which made omission after PN subjects less frequent than after NP subjects.

Be could also be automatically supplied with the PN subjects in the overgeneration constructions. However, this explanation would not justify overgeneration involving NP subjects, which records just about 8 percent fewer examples than overgeneration with PN subjects. It can thus be concluded that overgeneration in the L1-Malay data is attested with both PN and NP subjects. Examples of be overgeneration occurring with NP subjects were given above in (3a), (3d), and (3e), while (3b) and (3c) illustrated overgeneration involving PN subjects.

4.2.2. Types of subject predicatives

The next constituent analysed is the subject predicatives proceeding overgenerated be. We have identified 6 types of subject predicatives to complement the be + V overgeneration constructions in the L1-Malay data: noun phrase (NP), adjective phrase (AP), prepositional phrase (PP), infinitive to-clause (InfC), that clause (ThtC), and wh-clause (WhC). Table 6 below summarises the percentages of the occurrences of these subject predicatives in the position after overgenerated be.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Noun Phrase</th>
<th>Adjective Phrase</th>
<th>Prepositional Phrase</th>
<th>Infinitive to-clause</th>
<th>That-clause</th>
<th>Wh-clause</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP %</td>
<td>AP %</td>
<td>PP %</td>
<td>InfC %</td>
<td>ThtC %</td>
<td>WhC %</td>
</tr>
<tr>
<td>is</td>
<td>11.21</td>
<td>1.52</td>
<td>16.97</td>
<td>0.91</td>
<td>1.21</td>
<td>1.21</td>
</tr>
<tr>
<td>are</td>
<td>18.79</td>
<td>2.42</td>
<td>10.61</td>
<td>3.33</td>
<td>0.61</td>
<td>0.00</td>
</tr>
<tr>
<td>was</td>
<td>2.42</td>
<td>0.00</td>
<td>3.03</td>
<td>0.30</td>
<td>0.00</td>
<td>0.30</td>
</tr>
<tr>
<td>were</td>
<td>1.52</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>33.94</td>
<td>3.94</td>
<td>30.61</td>
<td>4.55</td>
<td>1.82</td>
<td>1.52</td>
</tr>
</tbody>
</table>

Table 6. Distribution of subject predicatives in the be overgeneration construction
Table 6 shows that subject predicative NPs and PPs record the highest percentages of occurrence after overgenerated *be*, approximately 34 and 31 percent, respectively. As expected, they occur mostly after the present forms *is* and *are*. The percentages of occurrence for the other subject predicative types are rather low, in all cases below 5 percent. It can be concluded that *be* overgeneration with subject predicatives in the L1-Malay learner data most commonly occurs in *be + V/Ved + NP* and *be + V/Ved + PP* constructions. Example (4a) below illustrates *be* overgeneration before a subject predicative NP, while (4b) shows and instance of overgeneration preceding a PP.

(4) a. *… if the wrongdoer or the offender is committed high level crime such as gang robbery, trafficking drugs, rape etc… (L0017a)
   b. *… the theory part is just taught about the various type of method, history and etc. (B0037-05)

4.2.3. Forms of post-*be* verbs

In light of the findings in previous studies, *be* overgeneration occurs with either bare verbs, as in *is go*, or with inflected verbs, as in *was happened*. The analysis of *be* overgeneration in the L1-Malay ESL learner data also took into account the forms of the post-*be* verbs. This section presents the results of the analysis of the forms of post-*be* verbs across the four finite *be* inflections *is*, *are*, *was*, and *were*. Table 7 below summarises the results.

<table>
<thead>
<tr>
<th></th>
<th>Uninflected</th>
<th>Inflected</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>is</em></td>
<td>29.09</td>
<td>14.55</td>
</tr>
<tr>
<td><em>are</em></td>
<td>28.18</td>
<td>17.58</td>
</tr>
<tr>
<td><em>was</em></td>
<td>3.64</td>
<td>5.45</td>
</tr>
<tr>
<td><em>were</em></td>
<td>1.52</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>62.42</td>
<td>37.58</td>
</tr>
</tbody>
</table>

Table 7. Forms of post-*be* verbs in *be* overgeneration

The results suggest higher occurrences of *be + bare V* constructions, since approximately 62 percent of the lexical verbs are found to be uninflected in comparison to only about 38 percent of inflected main verbs. Of the *be + bare V* constructions, 29 percent occur with *is* and 28 percent with *are*, thus suggesting that overgeneration involving bare verbs occurs with both singular and plural subjects.

As for *be + Ved* overgeneration, approximately 18 percent of the examples occur with *are* and about 15 percent with *is*. The comparatively fewer occurrences of *was/were* in the whole corpus account for the small percentages of overgeneration involving these forms. As shown in Table 7, only about 5 percent of the overgeneration dataset with inflected lexical verbs involve *was* and no occurrence is recorded for *were*, while only about 4 percent of uninflected verbs occur with *was* and 2 percent occur with *were*. In conclusion, *be* overgeneration constructions produced by the L1-Malay ESL learners are observed to occur more frequently with uninflected verbs than with inflected verbs.

Further analysis of the concordances of *be + V* overgeneration reveals that the pattern of *be* insertion resembles those found in the studies conducted by Ionin and Wexler (2001, 2002), which show that learners appear to insert *be* before a lexical verb in their attempts to mark agreement, as shown in (5):

(5) a. *The medias are always show the luxurious and exclusive lifestyle and this can make someone who doesn’t have money to acquire something luxurious to dream about it. (I0009)
   b. *... so for me they are just go to joined in “Khidmat Negara” for to enjoy theirself. (F0090)
   c. *Knowledge about the job is come from the theoretical that was teaching at the university. (B0089)

There appears to be almost no variability in the subject and *be* agreement, suggesting that learners are marking agreement with the use of suppletive inflection (see Lardiere 1998; Ionin and Wexler 2001, 2002). Ionin and Wexler (2001, 2002) also suggest that overgeneration such as this signals that the learners are treating the suppletive inflections as a default to mark agreement. This seems to be the case with the instances of *be + bare V* overgeneration in the L1-Malay ESL learners data.

Further analysis was conducted on *be* overgeneration with inflected lexical verbs. This analysis was carried out to determine the forms that the lexical verbs take in the construction. Lexical verbs are found to be inflected with the third person inflection –*s*, past tense or past participle –*ed*, and present participle –*ing*. Table 8 below summarises the distribution of *be* overgeneration with lexical verbs inflected with the morphemes –*s*, –*ed*, and –*ing*. 

---

---
As shown in Table 8, 66 percent of be overgeneration instances involving inflected lexical verbs occur with verbs inflected with the morpheme –ed, 31 percent with –s, and only 11 percent with –ing. The higher percentage of be being overgenerated with –ed raises the possibility that they could be the result of inappropriate applications of the English passive rule. This is in keeping with previous studies (Yip 1995; Balcom 1997; Oshita 2000; Ju 2000). As can be seen in (6) below, L1-Malay learners also tend to produce be overgeneration constructions that are similar to overpassivisation. However, unlike the overpassivisation construction found in previous studies (Yip 1995; Balcom 1997; Oshita 2000; Ju 2000), these constructions do not occur exclusively with unaccusative verbs. Except for (6b), which involves the unaccusative verb happen, (6a) and (6c) involve the use of the transitive verbs show and require. However, in light of the affixal inflection to the main verbs, the overgeneration exemplified in (6) could be impaired passive.

(6) a. *As a conclusion, it is clearly showed that money is the root of evil. (FP0132)
    b. *... that the lyric is trying to convey the meaning of the evil is always happened because of money. (FP0056)
    c. *Job world is required not the typical one but person who can generate the profit for the firms/governments in facing competitive world nowadays. (E0040)

Nevertheless, the affixal inflection can also be used by learners to check the tense feature. The extracts in (7) below provide examples of be + Ved overgeneration that appear to convey past time: (7a) could be intended for 'The problem is that most of graduates failed during the interview session’ and (7d) could mean ‘Most of them graduated with degrees’. The affixal inflection –ed is thus believed to be used for checking the tense feature, while the suppletive inflection functions as agreement marker.

(7) a. *The problem is that, most of graduates are failed during the interview session. (LB0052B)
    b. *Most of them are graduated in degrees. (E0046)

The figures in Table 8 also point to another interesting be overgeneration construction, namely be + Vs. Learners’ use of the third person morphemes –s, –es, and –ies, as shown in (8), suggests two possibilities: first, that learners are double marking agreement by using both suppletive and affixal inflections; and, second, the suppletive inflection is inserted to mark agreement, while affixal inflection performs the task of checking the tense feature in a way similar the overgeneration examples in (7) above. The second possibility is more likely, as all the extracts in (8) appear to convey present time and, thus, they use the inflectional morpheme –s. These be + Vs constructions provide further support to the idea that the main verbs in be overgeneration constructions are inflected as a mechanism to check the tense feature.

(8) a. *It is depends on oneself to make him/her have the skills. (B0090)
    b. *Learning process in university is begins with theoretical to introduce students to the roots of their field of study. (K0080)
    c. *It is seems that one person will get advantages to that exchange. (F0053)

4.2.4. Class of post-be lexical verb

The analysis of post-be constituents also includes the class of the main verb following the overgenerated be. The main verbs are classified into transitive, unergative, and unaccusative. In addition to that, information on whether the verbs are inflected or uninflected is also included. Table 9 below summarises the results.

<table>
<thead>
<tr>
<th> </th>
<th>Transitive (Vt)</th>
<th>Unergative (Uner)</th>
<th>Unaccusative (Unac)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninflected</td>
<td>30.09%</td>
<td>15.76%</td>
<td>7.58%</td>
</tr>
<tr>
<td>Inflected</td>
<td>18.80%</td>
<td>11.82%</td>
<td>7.58%</td>
</tr>
<tr>
<td>Total</td>
<td>48.89%</td>
<td>27.58%</td>
<td>15.16%</td>
</tr>
</tbody>
</table>

Table 9. Distribution of post-be verbs according to verb classes

As can be seen in Table 9, there is a higher percentage (49 percent) of be overgeneration occurring with transitive verbs in comparison to unergative and unaccusative verbs. Moreover, about 30 percent of the transitive verbs are uninflected in comparison to only about 19 percent that are inflected. As for the intransitive verbs, overgeneration involving
unergative verbs is more common than overgeneration involving unaccusative verbs. Nearly 28 percent of the overgeneration constructions occur with unergative verbs, while only about 15 percent of the examples occur with unaccusative verbs. Almost 16 percent of the unergative verbs are uninflected, while about 12 percent are inflected. The unaccusative verbs record the same percentage for inflected and uninflected forms (7.58 percent).

The figures clearly indicate that be overgeneration in the data of L1-Malay ESL learners in this study occurs more often with transitive verbs. This finding contrasts with those of earlier studies (Yip 1995; Balcom 1997; Oshita 2000; Ju 2000). Such a difference may be due to the purpose of the overgeneration. L2 learners in the studies conducted by Yip (1995), Balcom (1997), Oshita (2000), and Ju (2000) overgenerated be to passivise the unaccusative clause, in which case the unaccusative verb would be inflected with the morpheme –ed/–en. By contrast, L1-Malay learners’ be overgeneration derives from the learners’ attempts to mark agreement feature with the suppletive inflection, which is evinced by the higher percentage (62 percent) of the overgeneration constructions occurring with uninflected verbs.

5. DISCUSSION AND CONCLUSIONS

The analysis of the constituents before and after overgenerated be has revealed the patterns of be overgeneration in the L1-Malay data. The analyses show that be is often overgenerated after PN and NP subjects and before subject predicative NPs and PPs. As regards the class of post-be verbs, the majority of the examples with overgenerated be are preceded by uninflected transitive verbs. In light of these findings, be overgeneration in the L1-Malay ESL learner data is more likely to occur in the following patterns:

\[
\begin{array}{ll}
\text{PN} + \text{be} + \text{Vt} + \text{NP} \\
\text{PN} + \text{be} + \text{Vt} + \text{PP} \\
\text{NP} + \text{be} + \text{Vt} + \text{NP} \\
\text{NP} + \text{be} + \text{Vt} + \text{PP}
\end{array}
\]

The findings reveal that there are systematic patterns of be overgeneration in the data of L1-Malay ESL learners. There are two distinct patterns identified in the overuse of be with lexical verbs. First is the be + bare V construction, where the lexical verb takes its base form. This type of overgeneration construction occurs very frequently (62 percent) in the corpus data. The verb be in this construction is believed to be used by the learners to mark agreement. The learners’ tendency to resort to suppletive inflection as an agreement marker has already been documented in Zobl and Liceras (1994), Lardiere (1998), and Ionin and Wexler (2001). Ionin and Wexler (2001) stress that perhaps at this stage, the learners seem to master tense, with only partial or incomplete knowledge of agreement. They also add that the learners could be treating be as a default for marking agreement, since they might be having difficulties accessing the affixal agreement. This seems to be the case with the be + bare V overgeneration produced by the L1-Malay learners in this study.

The second be overgeneration pattern found in the corpus data is be + Ved, in which the lexical verbs are inflected with the past tense participle –ed. Unlike the be + Ven construction found in Yip (1995), Balcom (1997), Oshita (2000), and Ju (2000), be overgeneration involving inflected main verbs in the L1-Malay data does not occur exclusively with unaccusative verbs. Instead, it occurs most frequently with transitive verbs. Although the constructions manifest themselves as instances of impaired passives, the data suggest that the affixal inflections in the constructions are used to check the tense feature. This is supported by another interesting be overgeneration construction, namely be + Vs, in which the main verbs are inflected with the third person morpheme –s. The learners appear to be using the affixal inflection to check the tense feature and at the same time to mark agreement with the use of the suppletive inflection. At this stage, the learners seem to master tense, with only partial or incomplete knowledge of agreement. They tend to generalise the agreement rule and apply it unnecessarily to main verbs, which already have their agreement feature marked by affixal inflection.

The findings of this study suggest that be overgeneration in the corpus data of L1-Malay ESL learners could be the outcome of a developmental aspect of language acquisition. This traces back to the system underlying the patterns of overgeneration, which are made up of non-random constructions governed by very specific interlanguage grammar. These constructions constitute pieces of evidence of the learners’ projection of the IP system in English. The findings thus provide insights into how far and how successful the L1-Malay learners have acquired the English functional categories of tense and agreement after approximately twelve years of learning English as a second language. With these findings, ESL researchers and practitioners are also furnished with an understanding of the difficulties and confusions that these learners encounter in the course of mastering the English functional categories and may assist them in developing ways to help learners overcome such specific difficulties.


Lee, Nancy and Yue Yuan Huang. 2004. To be or not to be – the variable use of the verb BE in the interlanguage of Hong Kong Chinese children. *RELC* 35/2: 211–228.


