

Overgenerated *be* from Topic Marker to Verbal Inflection

Kitaek Kim

Seoul National University

1. Introduction

L2 morpheme order studies found that the ‘copula *be*’ is the first morpheme acquired by L2 learners (Bailey et al., 1974; Dulay & Burt, 1973, 1974; Ionin & Wexler, 2002; Zobl & Liceras, 1994). Interestingly, it has been widely observed among L2 learners of English and Romance languages that L2 learners use *be* with thematic verbs (e.g. *she is go home*) (Ahn, 2003; Bernini, 2003; Haberzettl, 2003; Hahn, 2000; Huebner, 1983, 1989; Ionin and Wexler, 2002; Lee, 2002; Shin, 2000, 2001; Starren, 2006; Yang, 2001, 2002, 2006). The researchers, Ionin and Wexler (2002), term this *be* the overgenerated *be*, and the syntactic properties of the overgenerated *be* has been controversial. A group of researchers argues that it is a topic marker (Ahn, 2003; Hahn, 2000; Huebner, 1983; Sasaki, 1990; Shin, 2000); the other group of researchers maintains that it is an early morphological appearance of a functional category, such as tense, agreement or aspect (Bernini, 2003; Ionin & Welxer, 2002; Starren, 2006; Yang, 2001, 2002, 2006). I call the former approach the "topic marker view", and the latter "functional category view".

This study does not argue that the two views are antithetical, but rather that the property of the overgenerated *be* they examine is different. I argue that the stages of interlanguage have different characteristics from one another. Specifically, I propose that the studies of the topic marker view have analyzed the characteristics of L2 interlanguages in the topic-prominent stage (TopP stage), and that those of the functional category view have investigated the properties of L2 interlanguage in the subject-prominent stage (SP stage). Many studies found that L1 topic or subject features can be transferrable in the L2 acquisition (Jin, 1994; Jung, 2004; Sasaki, 1990; Yuan, 1995). It is interesting to note that the mother tongues of L2 participants in the studies of the topic marker view, Laotian in Huebner (1983), Japanese in Sasaki (1990), and Korean in Ahn (2003), Hahn (2000), and Shin (2000), all of which belong to languages that have topic prominent features, according to Li and Thompson (1976), and that the topic marker is one of characteristics in the topic prominent languages.

The present study hypothesizes that the overgenerated *be* produced by topic prominent L1-speaking English learners initially serves as a topic marker because of L1 transfer. Then the topic-marking property gradually vanishes, as the overgenerated *be* develops into a target-like verbal inflection. Here, the shift of function from a topic marker to a verbal inflection in the overgenerated *be* reflects the interlanguage development that the sentence-initial NP redresses its property from a topic into a target-like subject. The production data that support this claim come from the L2 acquisition of English by native speakers of Korean. Before turning to L2 data, let me briefly review the conceptual distinction between the topic and the subject, and examine prevailing arguments regarding the properties of the overgenerated *be*.

2. Literature review

2.1. Interlanguage development from TopP stage to SP stage

Li and Thomson (1976) discuss the distinctive properties of the topic and the subject in the grammar. According to them, the subject is normally determined by the verb and is selectionally related to the verb; and the subject often obligatorily controls verb agreement. In contrast, the topic is syntactically independent from the rest of the sentence. It is discourse-dependent, which serves as the center of attention of the sentence and must thereby be definite.

Based on this distinction between the topic and the subject, Li and Thomson (1976) propose a typological difference between topic-prominent (TopP) languages and subject-prominent (SP) languages. Languages in the same typological type share a number of syntactic properties. For example, TopP languages have a topic marker, double subjects, and null subjects, whereas SP languages lack them. The SP languages, but not the TopP languages, have a dummy subject. In the generative approach to language acquisition, Baker (2001) proposes that the typological differences between TopP languages and SP languages can be viewed in terms of a parameter, called, 'topic-prominent parameter'. The topic-prominent parameter is considered to be on when "a sentence may be made up of an initial noun phrase (the topic) and a complete clause that is understood as a comment on that topic", and off when "no topic phrase distinct from the clause is allowed" (Baker, 2001, p. 182).

Studies on interlanguage development have focused on the different syntactic properties present in TopP and SP languages to determine the relative degree of topic-prominence or subject-prominence (Fuller & Gundel, 1987; Sasaki, 1990; Jin, 1994; Jung, 2004; Yuan, 1995). To verify an interlanguage developmental trend, the characteristics of TopP languages and SP languages have received particular attention, including, for example, double subject, null subject and dummy subject. That is, the L2 productions of double subject and null subject demonstrate the TopP properties of interlanguage, while the productions of dummy subject indicate the SP properties of interlanguage.¹ The role of topic/subject prominence typology in L2 acquisition has been discussed as following: Fuller and Gundel (1987) proposed that the earlier L2 learners' stage is the TopP stage regardless of L1, whereas Jin (1994) substantiated that SP stage can be L2 learners' earlier stage. Sasaki (1990) showed the interlanguage development from TopP stage to SP stage with L1 Japanese learners acquiring L2 English, whereas Jung (2004) demonstrated the interlanguage development from SP stage to TopP stage with L1 English-speaking L2 Korean learners.

2.2. Overgenerated *be*: A topic marker or a functional category?

L2 learners frequently use forms of *be* in utterances that contain a thematic verb in place of a progressive participle (e.g., *She is like banana*). Perhaps the overgenerated *be* is one of the most controversial phenomena in the 2000s. Accounts of this phenomenon fall into two broad categories. One view attributes the phenomenon to the topic marker. This topic marker view has been espoused by a number of L2 researchers (e.g., Ahn, 2003; Hahn, 2000; Huebner, 1983; Sasaki, 1990; Shin, 2000;). Yet there are other researchers (e.g., Bernini, 2003; Ionin & Welxer, 2002; Starren, 2006; Yang, 2001, 2002, 2006) who argue that the overgenerated *be* is an early morphological appearance of a functional category, such as tense, agreement, or aspect.

Those who share the topic marker view point to a common L1 typological feature: Lao, Japanese, and Korean belong to languages that have topic-prominent features. Huebner (1983) found that *Ge*, his Lao learner of English, used *is* or *isa* to distinguish presupposition from assertion, or topic from comment, and he concluded that the *is(a)* is a topic-comment boundary marker. Below is an example of *is(a)* produced by *Ge*.

Hua aen song, isa bodii sik
 "As for Hua and Song, their bodies were sick."

Ge used *is(a)* very frequently at first, and then gradually decreased the use of it. His responses were marked with *is(a)* 65 percent of the time in Tape One, 45 percent of the time in Tape Two, 4 percent of the time in Tapes Three and Four, and 14 percent of the time in Tapes Five and Six. Sasaki (1990) also indicates that the beginning English learners whose L1 is Japanese used the overgenerated *be* as a topic marker (e.g. *shoes is tiger give*). Sasaki argues that the overgenerated *be* is used to signal a boundary between a topic and a comment. Ahn (2003), Hahn (2000), and Shin (2000, 2001) suggest that the overgenerated *be* produced by Korean beginners of English is also a topic marker. For example, "*She is*

¹ Sasaki (1990) investigated the interlanguage development from L1 topic-prominence to L2 subject-prominence in the following way: The study examined an existential sentence with a locative topic and found that topic-stage learners tend to produce sentences with topicalized locative at the sentential-initial position (e.g. *on the table two books exist*), whereas subject-stage learners tend to produce more target-like sentences with locatives at the sentence final position (e.g. *there are two books on the table*).

like music” means “*As for her, she likes music*” (Shin, 2001:2) where *is* becomes a transfer of *-un-* (or *-nun-*), which is a topic marker in Korean.

In contrast to the topic marker view, the functional category view suggests that the overgenerated *be* is an early morphological marker of a functional category, such as tense, agreement, or aspect. Ionin and Wexler (2002) analyzed the spoken productions of Russian ESL beginners. They found that the overgenerated *be* was produced very frequently by these Russian learners: Of the 28 transcripts, 18 transcripts showed at least one instance of the overgenerated *be*, and a total of 108 utterances of the overgenerated *be* accounted for nine percent of all inflected utterances across all the transcripts. Ionin and Wexler focused on the finding that most of the overgenerated *be* were followed by uninflected stem forms. Here, they claimed that (some of) the L2 learners’ use of *be* forms was a ‘substitute’ for an affixal inflection in a non-progressive clause.

Yang (2001, 2002, 2006) points out that the overgenerated *be* is an underdeveloped functional category that is positioned at the head of FP (underdeveloped IP). He found that the overgenerated *be* produced by Korean EFL beginners has properties of agreement and tense: the overgenerated *be* showed its form correctly in *am/are/is* depending on the subjects, and it was used in finite clauses where the tense was marked.

Bernini (2003) and Starren (2006) analyzed the overgenerated *be* produced by L2 learners of Romance languages. Bernini studied an Eritrean learner of Italian longitudinally, and argues that the overgenerated *be* is an “explicit link to finiteness with lexical elements, but which cannot incorporate finiteness as inflected verbs in the target language do” (Bernini, 2003. p. 175); Starren (2006) examined Turkish and Moroccan learners of Dutch from a longitudinal perspective and proposes that the overgenerated *be* is produced to mark tense and aspect.

Interestingly, the two views seem to be divided by L1 typological type. Note that all the studies on the topic marker view took participants whose mother tongues have properties in the topic-prominent languages: Laotian in Huebner (1983), Japanese in Sasaki (1990), and Korean in Hahn (2000), Shin (2000), and Ahn (2003). According to Li and Thompson (1976), Laotian is a topic-prominent language; Japanese and Korean are languages that are both topic-prominent and subject-prominent. However, all but Yang’s (2001, 2002, 2006) studies in the functional category view examined participants whose mother tongues do not have the properties of the topic-prominent languages: Russian in Ionin and Wexler (2002), Eritrean in Bernini (2003), and Turkish and Moroccan in Starren (2006)².

Table 1. Views on the overgenerated *be*

Views on the overgenerated <i>be</i>	Studies	Mother Tongues of the Participants	Topic-Prominence in the Mother tongue
Those holding the topic marker view	Huebner (1982)	Laotian	O
	Sasaki (1990)	Japanese	O
	Hahn (2000), Shin (2000), Ahn (2003)	Korean	O
Those holding the functional category view	Ionin & Welxer (2002)	Russian	X
	Bernini (2003)	Eritrean	X
	Starren (2006)	Turkish, Moroccan	X
	Yang (2001, 2002, 2006)	Korean	O

² The mother tongues of the participants in the functional category view seem to belong to subject prominent languages. Li and Thompson (1976) discuss that subject-prominent languages are as following: Indo-European, Semitic, Niger-Congo, Finno-Ugric, Dyrbal, Indonesian, and Malagasy.

As shown in Table 1, the studies in the topic marker view present one feature in common, which is that the mother tongues of the participants have the properties of topic prominent languages. Here, it may be argued that the L1 transfer is reflected in the overgenerated *be*. However, Yang (2001, 2002, 2006), examining the L2 productions of the topic prominent L1 Korean-speaking L2 English learners, stands on the functional category view, pointing out that the overgenerated *be* has a property of the functional category. If we assume that the interlanguage data that Yang (2001, 2002, 2006) focused on are in slightly more advanced level than those of topic marker view, the assumption lead us to consider the following possibility: the gradual development of the overgenerated *be* from a topic marker to a functional category (an inflection) when interlanguage develops from L1 TopP languages to L2 SP languages.

3. Methodology of the current study

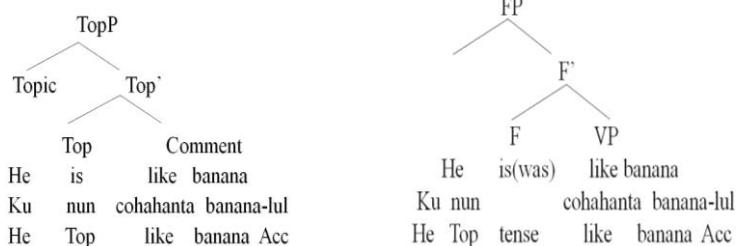
3.1. The research hypotheses

This study hypothesizes that the morpho-syntactic property of Korean-speaking English learners' overgenerated *be* develops from a topic marker to a functional category, more specifically, to a verbal inflection. Korean-speaking English learners construct a topic-comment structure in the initial state of L2 English production. The overgenerated *be* produced therein is used as a topic marker because of the L1 transfer of the Korean topic marker *-un/-nun-*. Then, the learners produce a target-like subject-predicate structure in their gradual interlanguage development of the overgenerated *be* equipped with an inflectional property. This study suggests three specific hypotheses based on the distinction between topic and subject provided by Li and Thomson (1976), and at the same time, proposes the syntactic trees for both of the TopP stage and the SP stage based on Rizzi (1997) and Yang (2001, 2002, 2006). These are presented in Table 2.

Table 2. Hypotheses and their syntactic trees

	TopP Stage	SP Stage
Hypothesis #1: Subject	Double subject or null subject is observed for L1 transfer.	Double subject or null subject is hardly ever observed.
Hypothesis #2: Subject-verb agreement	Subject-verb agreement is rarely found since a topic is independent from verb in a sentence.	Subject agrees with verb since the subject is syntactically licensed by inflection.
Hypothesis#3: Over-generated <i>be</i>	As a topic marker, the overgenerated <i>be</i> does not carry inflectional features of the verb, such as a tense feature.	As an inflection, the overgenerated <i>be</i> can be inflected for tense.

The syntactic structure for overgenerated *be*



3.2. Study participants

23 Korean first-year middle school students participated in this study. All the students were 13 years old and enrolled in S Middle School (a pseudonym) in Seoul. Of the group, 8 were female, and 15 were male. They had studied English as a foreign language for less than one year in their middle school and for four years in their elementary school. No student had any experiences of living abroad for more than three months. The average amount of the accumulated time that the students were exposed to

English instruction in their schools for more than four years was about 200 hours. The 23 students in this study all took an English writing class once a week as an extra-curriculum program at their middle school.

To investigate the interlanguage development in the use of the overgenerated *be*, this study divided the students into three groups based on their English proficiency: the level of proficiency goes from G1 being the lowest to G3 being the highest. The proficiency-based grouping of the 23 students was at first conducted according to the judgments made by three lecturers who taught them in the writing class. The lecturers were asked to evaluate all of the 23 students according to the ACTFL (American Council on the Teaching of Foreign Languages) writing proficiency guideline. Based on the results of the evaluation, the 23 students were divided into three groups: 5 students in G1, 7 students in G2, and 11 students in G3. According to the three lecturers, all the G1 students were correctly placed in either novice-high or intermediate-low proficiency and all the G2 students belonged either to the intermediate-low or the intermediate-mid proficiency of the ACTFL guideline. Additionally, all of the G3 students were labeled as intermediate-mid proficiency. Then, the average scores of the students' mid-term and final English exams at school were additionally used as a general English proficiency measurement for the study. The average scores of the three groups on their mid-term and final English exams at school were 75.4 for G1, 89.5 for G2, and 96.1 for G3³. This breakdown indicates that G1 was the lowest group in terms of English proficiency, followed by G2 at a medium level, with G3 being the highest level of proficiency. The Scheffé test results also confirmed that the groups were indeed different from one another in terms of English proficiency.

This study assumes that the interlanguage development of Korean-speaking English learners, i.e. topic-prominent grammar due to L1 transfer to subject-prominent grammar compatible with L2 structures, can be observed by comparing the three groups. That is, G1 reflects the earlier stage, G3 the advanced stage, with G2 placed in between the two groups.

3.3. Data collection

The 23 students all took an English writing class once a week in an extra-curriculum program for 16 weeks. They were given a total of 20 diverse topics on which they were asked to produce an English writing sample in 15 minutes on average. The lecturers provided the students with some lexical words to facilitate their writing, but a special care was taken to ensure that the lecturers never offer any grammatical morphemes. At the end of the program, a total of 377 writing samples were collected for analysis.

3.4. Data analysis

In Section 3.1, this study proposed the three specific hypotheses with respect to the interlanguage development from TopP stage to SP stage: 1) the double subject or null subject is only observed in the TopP stage, 2) subject-verb agreement is rarely found in the TopP stage, and 3) the overgenerated *be* functions as a topic marker in the TopP stage, but as a verbal inflection in the SP stage.

The first hypothesis was tested by computing the number of the double subject and null subject. Since both are properties that indicate topic-prominence, a lower level group was expected to use more double subjects and null subjects in their L2 productions.

³The scores of each student in the three groups are as follows.

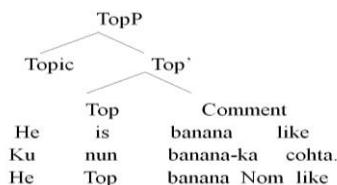
G3	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Avg.
Score	95.5	98.5	95.5	97.0	91.5	100	98.5	95.0	92.0	96.0	97.5	96.1
G2	S12	S13	S14	S15	S16	S17	S18			Avg.		
Score	88.0	95.0	88.0	89.5	89.5	No data	87.0			89.5		
G1	S19	S20	S21	S22	S23					Avg.		
score	77.5	73.5	83.0	67.5	75.5					75.4		

To test the second hypothesis, the suppliance rate of 3rd person singular *-s* in obligatory contexts was examined. Since topic is syntactically independent from verb, the suppliance rate of 3rd person singular *-s* in the TopP stage was expected to be lower than the suppliance rate of 3rd person singular *-s* in the SP stage.

Finally, this study suggests three ways to examine the morpho-syntactic property of the overgenerated *be*. First, this study calculated the average production rate of the overgenerated *be*. The overgenerated *be* was expected to be productively used as a topic marker in the TopP stage as shown in Sasaki (1990) or Hahn (2000). Then, it was predicted to be used as an underdeveloped verbal inflection in the early SP stage, but to be vanished finally in the SP stage. Based on the assumption of Ionin and Wexler (2002) that the overgenerated *be* can be a ‘substitute’ for an affixal inflection in a non-progressive clause, it is predictable that the overgenerated *be* disappears as learners become better at marking affixal inflections in the SP stage.

Second, this study examined past-tense marking on the overgenerated *be*. It was expected that the overgenerated *be* produced in the TopP stage would not carry past-tense. For example, a sentence like “he was like banana” is not expected to be observed in the TopP stage interlanguage.

Third, this study analyzed the overgenerated *be* shown in the double subject construction. When the overgenerated *be* is used as a topic marker, the overgenerated *be* will be followed by topic rather than subject. For example, to produce a double subject construction for the meaning, “as for him, he likes a banana”, learners may prefer to produce “He is banana like or He is like banana” rather than “He banana is like.” See the following syntactic tree for more details of this discussion.



In the syntactic tree, the topic of the sentence is *he*, and the subject is *banana*. If the overgenerated *be* is a topic marker, it will be mostly followed by topic rather than subject. Table 3 shows the expected results for these hypothesis tests.

Table 3. Expected results for the tests of the three hypotheses

	TopP Stage	SP Stage
Subject forms	· Presence of double subject & null subject	· Lack of double subject & null subject
Subject-verb agreement	· Lower suppliance rate of 3 rd person <i>-s</i>	· Higher suppliance rate of 3 rd person <i>-s</i>
Over-generated <i>be</i>	<ul style="list-style-type: none"> · Productive use as a topic marker because of L1 transfer · Lack of past-tense marking · Followed by topic rather than subject in the double subject construction 	<ul style="list-style-type: none"> · Use as an verbal inflection, then gradually decrease in use · Presence of past- tense marking

4. Results

4.1. Subject forms: Double subject and null subject

A total of 27 double subject constructions was found only in G1 and G2 levels. This study found two types of the double subject construction. First was a type consisting of topic, *overt* thematic subject and thematic verb. The followings are examples of these double subject constructions.

Type 1 of Double Subject Construction: Topic + overt thematic subject + thematic verb

(1a) English is one teacher come my home (*Kimye, G1*)

“As for English, one teacher come to my home.”

(1b) Chusok is all family got together. (*Kimye, G1*)

“In Chusok, all family got together.”

In (1a), the topic is *English*; the overt thematic subject is *one teacher*; and the thematic verb is *come*. In (1b), the topic is *Chusok*, the overt thematic subject is *all family*, and the thematic verb is *got*.

The second type consists of topic, *covert* thematic subject and thematic verb⁴. See the following examples.

Type 2 of Double subject construction: Topic + covert thematic subject + thematic verb

(2a) Vacation go to grandfather and grandmother house. (*Jo, G2*)

“Last vacation, I went to my grandfather and grandmother’s house.”

(2b) Our class *was* introduce myself. (*Jung, G1*)

“In our class, I introduced myself.”

In (2a), the topic is *last vacation*; the thematic subject is covert; the thematic verb is *go*. In (2b), the topic is *our class*; the thematic subject is covert; the thematic verb is *introduce*. This second type is best viewed as the double subject construction, as a sentential topic, distinguishable from a thematic subject, was used. Thus, this study regards the second type as a form of the double subject construction.

Of the 27 double subject constructions, G1 produced 15, and G2 produced 12. In terms of the proportion of the double subject constructions out of total sentences⁵, 5.8% (15/259) were in G1, while only 1.3% (12/912) were in G2. The chi-square test found the difference in the proportions between G1 and G2 as statistically significant at the level of .01 (chi-square=16.73, $p=.0001$). This analysis suggests that the students in G1 relied on the L1 property of the topic prominence in the productions of L2 English.

Likewise, the null subjects were only found in G1 with tokens of six. The null subject is another characteristic of topic prominent languages, distinguished from the subject prominent languages. Table 4 shows the proportion of such double subject and null subject constructions in all of the sentences with thematic verbs in each group.

Table 4. Proportion of double subject and null subject constructions

	G1 (lowest proficiency)	G2 (medium proficiency)	G3 (highest proficiency)
①# of sentences with double subject	15	12	0
②# of sentences with null subject	6	0	0
③# of total sentences	259	912	1852
Rate [= (①+②)/③]	8.1%	1.3%	0%

This table shows that double subject and null subject constructions, which are the distinctive characteristics of topic prominent languages, were present only in G1 and G2. Between the two groups, the constructions were more frequently used by G1 than they were by G2 (8.1% in G1 and 1.3% in G2). The chi-square test found that the difference was statistically significant at the level of .01 (chi-square=31.00, $p=.0001$). That is, the L1 property of topic-prominence was more prominently shown in the English productions by the G1 group than those by the G2 group. Thus, the first hypothesis is supported. The data provide evidence that G1 belongs to the TopP stage, followed by G2, with G3 being close to the SP stage.

4.2. Subject-verb agreement

Table 5 presents the average rates of the three groups on the suppliance of 3rd person singular agreement markings -s in obligatory contexts.

⁴ In fact, there were another types of the double subject constructions, such as “*topic + overt subject + copula be + noun or adjective*” (e.g. *he is test is good*). These constructions were excluded in this study. That is because this study focuses on the overgenerated *be* in a sentence where thematic verbs were used as main verbs.

⁵ This study only deals with sentences where the main verb is a thematic verb. Therefore, the ‘total sentences’ in this study excluded sentences where the main verb was the copula *be*.

Table 5. The Suppliance Rate for 3SG -s in Obligatory Contexts

Group	G1 (lowest proficiency)	G2 (medium proficiency)	G3 (highest proficiency)
Rate	18.8% (12/64)	37.3% (62/166)	67.6% (148/219)

G1 marked the third person singular *-s* in the obligatory contexts the most poorly (18.8%), followed by G2 (37.3%), with G3 marking it the most accurately (67.6%). This difference among the groups indicates that the division of the three groups reflects inflectional development levels, i.e., G1 is the poorest, followed by G2, with G3 being the best. As discussed, the agreement with the verb is the property of the subject, not that of the topic. Hence, this result supports that G1, the lowest English proficiency group, is in the TopP stage while G3, the highest English proficiency group, is in the SP stage, with G2 placed in between the two stages. Here, the second hypothesis is supported, namely that subject and verb agreement is rarely found in the TopP stage.

4.3. The overgenerated *be*

4.3.1. The production rate of the overgenerated *be*

Since the overgenerated *be* appears only with main verbs, it is possible to calculate the average production rate of the overgenerated *be*: i.e., the total number of clauses with thematic main verbs divided by the total number of the overgenerated *be*⁶. The overgenerated *be* was hardly shown among the writings of the G3 students, whereas both G1 and G2 students produced the overgenerated *be* considerably more. The average production rates for the overgenerated *be* in each group are presented in Table 6.

Table 6. Average Production Rate of the Overgenerated *be*

	G1 (lowest proficiency)	G2 (medium proficiency)	G3 (highest proficiency)
# of overgenerated <i>be</i>	85	172	18
# of clauses with thematic main verbs	259	912	1852
Average production rate	32.8%	18.9%	0.9%

As Table 6 shows, the overgenerated *be* was productively used only in the lower two groups, G1 and G2. Between the lower two groups, as shown by the average production rate of the overgenerated *be* in each group, the students in G1 used the overgenerated *be* more prominently (32.8%) than did those in G2 (18.9%). The difference was statistically significant at the level of .01 (chi-square=13.84) according to the result for the chi-square test. This result shows a trend, namely that the overgenerated *be* gradually decreases as interlanguage develops into the SP stage.

4.3.2. Tense marking in the overgenerated *be*

This study found that the overgenerated *be* was frequently used as a tense marker in the G2 rather than in the G1. In these two groups, three ways to mark the past-tense were observed: 1) correct past-tense marking on thematic verbs (e.g. *she went home*), 2) incorrect past-tense marking on the overgenerated *be* (e.g. *she was go home*), 3) incorrect double past-tense markings both on thematic verb and on the overgenerated *be* (e.g. *she was went home*). Table 7 shows the types of the past-tense markings in both of G1 and G2.

⁶ Here, the “total number of clauses with thematic main verbs” excluded the number of sentences whose main verb was inflected for *-ing* since *be* with the verb inflected for *-ing* is the grammatically correct form of the progressive. In the writings of G1 and G2 students, there were a total of 27 sentences whose main verb was inflected for *-ing*. In addition, the passive types, *to*-infinitives, *ing*-infinitives, interrogatives, imperatives were also excluded.

Table 7. Past-tense marking in obligatory contexts

Past tense marking types	G1 (lowest proficiency)		G2 (medium proficiency)	
	Token	Rate	Token	Rate
	Correct past-tense marking for thematic verb (Ex. She went home.)	54	56.3%	166
Past-tense marking for the overgenerated <i>be</i> (Ex. She was go home.)	4	4.2%	<u>33</u>	<u>12.6%</u>
Double past-tense marking (Ex. She was went home.)	1	1.0%	5	1.9%
Lack of past-tense marking (Ex. She go home.)	37	38.5%	58	22.1%
Obligatory contexts	96	100%	262	100%

The overgenerated *be* was more frequently used to mark past tense in G2 than in G1. In G2, 14.5% marked past-tense for the overgenerated *be* (12.6% of the overgenerated *be* past-tense marking + 1.9% of double past-tense marking) whereas in G1, 5.2% marked past-tense for the overgenerated *be* (4.2% of the overgenerated *be* past-tense marking + 1.0% of double past-tense marking).

The chi-square test found that the difference was statistically significant at the level of .05 (chi-square=4.70, p=.03). That is, the overgenerated *be* was more prominently used as a tense marker in G2 than it was in G1. The overgenerated *be* to bear tense features indicate that it functions as a verbal inflection rather than a topic marker.

4.3.3. Location of the overgenerated *be* in the double subject construction

This study found that when the overgenerated *be* is shown in the double subject construction, it is always followed by topic as seen in following examples.

- (3a) English is one teacher come my home (*Kimye, G1*)
 “As for English, one teacher come to my home.”
 (3b) Chusok is all family got together. (*Kimye, G1*)
 “In Chusok, all family got together.”

In (3a) above, the topic of the sentence is not *one teacher* but *English*. Note that the overgenerated *be* follows not *one teacher*, but *English*, the topic. Likewise, in (3b), the overgenerated *be* follows the topic, *Chusok*, although another subject, *all family*, does not accompany the overgenerated *be*.

This study also observed that the overgenerated *be* was remarkably seen in the double subject constructions. Of all the double subject constructions, 70.4% (19/27) involved the overgenerated *be*. Note that the average production rate for the overgenerated *be* was 32.8% (85/259) in G1 and 18.9% (172/912) in G2 as previously demonstrated in Table 6. In fact, this study found that the double subject constructions were largely observed in G1. This finding also supports the claim that the overgenerated *be* serves as a topic marker in the lowest proficiency group, namely in the TopP stage.

5. Discussion

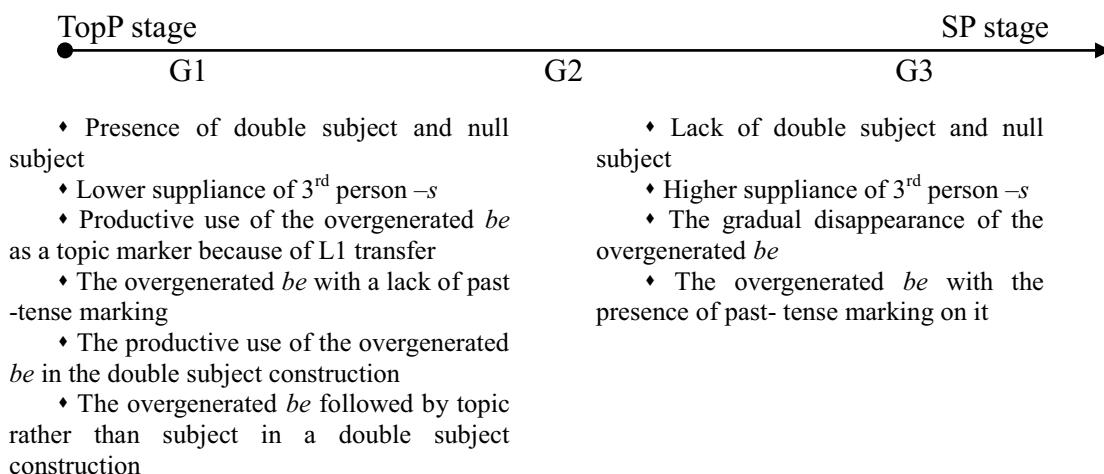
It has been widely observed that the copular *be*, the first grammatical morpheme acquired by L2 learners of English, is overgenerated (e.g. *she is go home*). The present study claims that the overgenerated *be* produced by Korean-speaking English learners initially functions as a topic marker, but then it develops into a verbal inflection. In English, the function of the copular *be* is to connect the subject with non-verbal predicate. However, it is highly probable that Korean beginners of English use the copular *be* to connect topic with comment because the topic-comment is a basic construction in Korean. Here, the copular *be* can be seen to be analogous to the topic-comment boundary marker, *-un/-nun-* in Korean.

The results of the present study solidify the claim. To begin, this study found that the lower proficiency group belongs to the TopP stage, and the higher proficiency group reaches the SP stage.

The lower proficiency group showed a tendency to use the topic-comment construction: When compared with both G2 and G3, G1 used the double subject and null subject constructions more frequently (8.1% in G1, 1.3% in G2, and 0% in G3). In addition, G1 was the poorest with subject-verb agreement, followed by G2, with G3 being the best in such aspect.

This study found that the overgenerated *be* is produced more frequently in the TopP stage. The production rates of the overgenerated *be* were 32.8% in G1, 18.9% in G2, and 0.9% in G3. The overgenerated *be* in the TopP stage serves as a topic marker as seen in the case where the overgenerated *be* is followed by topic rather than subject in double subject constructions. Then the overgenerated *be* develops into a verbal inflection in the SP stage. The G2 students marked past tense on the overgenerated *be* more frequently than did the G1 students.

Finally, the overgenerated *be* vanishes, as learners fully acquire the rule of correct tense marking, such as the past tense *-ed* or third person *-s*. Note that G3 in the SP stage hardly produced the overgenerated *be* and showed the most accurate use of third person *-s*. The following continuum summarizes the results of the present study.



However, we have to note that the L2 learners have a quite correct agreement between subject and *be* from the earlier stage as Yang (2001, 2002, 2006) points out. This study also observed that the students in this study showed a quite correct agreement between subject and *be*: For example, a total of 257 overgenerated *be* included only seven agreement errors between subject and *be*. The seven agreement errors appeared with open class and plural subjects as following: “Some students is”, “Guests is”, “Child and mother is”, “Child and mother is”, “Francisco’s family and friend is”, “guest and Mr. Bean is”, “Children is”. Six were collected from G1, and one was from G2.

Two accounts can be considered for this matter. One possibility is that the agreement produced by the learners in earlier stage is not a syntactic agreement between subject and *be* but a chunk. Wilson (2003) proposes that the agreement between subject and *be* can be a chunk: his analysis of child L1 acquisition data demonstrated that *be* was missing a lot when it accompanies infrequently used open class subjects. However, *be* correctly appeared when it followed closed class subjects (e.g. *he, you*) or frequently used open subjects (e.g. *Tom, pony*). Wilson, thereby, argues that constructions such as “I am”, “you are”, “he is” can be placed as a chunk in the learner’s mental lexicon due to its high input frequency. In the current study, the 257 subject NPs consisted of 140 open class nouns and 117 pronouns. Most of the 140 open class subjects looked frequently used nouns (e.g. student, teacher) or friendly names in a given story (e.g. Mr. Bean). However, regardless of whether overgenerated *be* is accompanied by or not, most of the subject NPs in the lower level groups were closed class or frequently used nouns. That is, it is difficult to attribute the occurrences of the overgenerated *be* to the matter of input frequency since most of the words produced from the lower proficient learners’ lexicon belong to high-frequency words. That is, the “chunk” account is not feasible enough to address the correct agreement between subject and *be*. A more likely possibility is that the overgenerated *be* as a topic marker contains some agreement features heading the projection TopP, a type of CP. This account is based on Yang (2006). Yang argues that *be* heads either CP or IP in the interlanguage of Korean-

speaking English learners. He proposes a configuration for the double subject constructions where *be* is marked both with topic (first NP) and subject (second NP) (e.g. *He is friend is many*) as following: [_{CP} [_C *be*] [_{IP} [_{I'} [_I *be*] [_{VP}]]]]. Note that *be* can be the head of CP as well as the head of IP. The correct agreement between subject and topic marker *be*, may be due to the possibility that agreement features reside in C in the TopP stage. If so, it may be proposed that the agreement features could reside in both C and I in the TopP stage, but that the agreement features would be located only in I with tense features in the SP stage. For the elaboration of this account, the questions of what roles C plays in the interlanguage grammar and why agreement features initially can reside in C in the TopP stage should be delved into in subsequent studies.

6. Conclusion

This study argued that the overgenerated *be* produced by Korean-speaking English learners initially functions as a topic marker, and then develops into a verbal inflection. The findings of this study can be extended in the following ways. First, a longitudinal study could be performed to explore the developmental patterns of the inflectional functional categories. Second, further research could be conducted with participants other than Korean-speaking English learners. The different patterns, if any, might be able to illustrate the acquisition of L2 functional categories more clearly.

References

- Ahn, Sung-Ho. (2003). A note on the topic-comment stage in Korean EFL syntactic development. *Studies in Generative Grammar*, 13(2), 369-382.
- Bailey, Nathalie, Madden, Carolyn, & Krashen, Stephen D. (1974). Is there a 'natural sequence' in adult second language acquisition? *Language Learning*, 24(2), 235-243.
- Baker, Mark C. (2001). *The atoms of language*. New York: Basic Books.
- Bernini, Giuliano. (2003). The copula in learner Italian: Finiteness and verbal inflection. In Christine Dimroth & Marianne Starren (Eds.), *Information structure and the dynamics of language acquisition* (pp. 159-186). Amsterdam: John Benjamin Publishing Company.
- Dulay, Heidi C. & Burt, Marina K. (1973). Should we teach children syntax? *Language Learning*, 23, 245-258.
- Dulay, Heidi C. & Burt, Marina K. (1974). Natural sequences in child second language acquisition. *Language Learning*, 24, 37-53.
- Fuller, Judith, W. & Gundel, Jeanette. K. (1987). Topic-prominence in interlanguage. *Language Learning*, 37(1), 1-17.
- Haberzettl, Stefanie. (2003). "Tinkering" with chunks: Form-oriented strategies and idiosyncratic utterance patterns without functional implications in the IL of Turkish speaking children learning German. In Christine Dimroth & Marianne Starren (Eds.), *Information structure and the dynamics of language acquisition* (pp. 45-63). Amsterdam: John Benjamin Publishing Company.
- Hahn, Hye-ryeong. (2000). *UG Availability to Korean EFL learners: A longitudinal study of different age groups*. Unpublished doctoral dissertation. Seoul National University.
- Huebner, Thom. (1983). *A longitudinal Analysis of the acquisition of English*. Ann Arbor, MI: Karoma Publishers.
- Huebner, Thom. (1989). Establishing point of view: the development of coding mechanism in a second language for the expression of cognitive and perceptual organization. *Linguistics*, 27, 111-143.
- Ionin, Tania. & Wexler, Kenneth. (2002). Why is 'is' easier than '-s'? acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research* 18(2), 95-136.
- Jin, Hong Gang. (1994). Topic-prominence and subject-prominence in L2 acquisition: Evidence of English-to-Chinese typological transfer. *Language Learning*, 44(1), 101-122.
- Jung, Euen Hyuk. (2004). Topic and subject prominence in interlanguage development. *Language Learning*, 54(4), 713-738
- Li, Charles N. & Thompson, Sandra A. (1976). Subject and topic: a new typology of languages. In Charles N. Li (Ed.) *Subject and topic* (pp. 458-489). New York: Academic Press.
- Rizzi, Luigi. (1997). On the fine structure of the left periphery. In Liliane Haegeman (Ed.). *Elements of grammar* (pp. 281-337). Dordrecht: Kluwer.
- Sasaki, Miyuki. (1990). Topic prominence in Japanese EFL students' existential constructions. *Language Learning*, 40(3), 337-368.
- Shin, Jung-Sun. (2000). *Functional category acquisition by Korean EFL learners: the role of UG in foreign language learning*. Unpublished doctoral dissertation. Seoul National University.
- Shin, Jung-Sun. (2001). L1 influence in foreign language learning: Topic-prominence in Korean EFL Learners' Interlanguage Grammar. *Foreign Language Education*, 8(1), 1-21.

- Starren, Marianne. (2006). Temporal adverbials and early tense and aspect markers in the acquisition of Dutch. In Veerle van Geenhoven (Ed.) *Semantics in Acquisition* (pp. 219-244). Dordrecht: Springer.
- Yang, Hyun-Kwon. (2001). Categorical Properties of Korean EFL Learner's "Be". *Korean Journal of English Language and Linguistics* 1(1), 59-79.
- Yang, Hyun-Kwon. (2002). Korean EFL learner's acquisition of English inflectional features. *Korean journal of English Language and Linguistics* 2(2), 227-248.
- Yang, Hyun-Kwon. (2006). Functional projection in Korean ESL learner's interlanguage syntax. *The SNU journal of Education Research*, 15, 117-128.
- Yuan, Boping. (1995). Acquisition of base-generated topics by English-speaking learners of Chinese. *Language Learning*, 45(4), 567-603.
- Wilson, Stephen. (2003). Lexically specific construction in the acquisition of inflection in English. *Journal of Child Language*, 30, 75-115.
- Zobl, Helmut. and Liceras, Juana. (1994). Functional categories and acquisition orders. *Language Learning*, 44(1), 159-180.

Selected Proceedings of the 2009 Second Language Research Forum: Diverse Contributions to SLA

edited by Luke Plonsky
and Maren Schierloh

Cascadilla Proceedings Project Somerville, MA 2011

Copyright information

Selected Proceedings of the 2009 Second Language Research Forum:
Diverse Contributions to SLA
© 2011 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-444-7 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Kim, Kitaek. 2011. Overgenerated *be* from Topic Marker to Verbal Inflection. In *Selected Proceedings of the 2009 Second Language Research Forum*, ed. Luke Plonsky and Maren Schierloh, 70-81. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2525.