

Research Article

Min-Chang Sung, Kitaek Kim* and Bora Nam

Influence of topic-prominent L1s on the use of L2 English copula *be*: a corpus-based study

<https://doi.org/10.1515/iral-2024-0062>

Received February 25, 2024; accepted October 29, 2024; published online November 19, 2024

Abstract: This study explores the influence of first language (L1) on the acquisition of second language (L2) English, specifically examining the use of copula *be* as a topic marker. Although prior research suggests that L1 transfer may impact L2 acquisition of copula *be*, the evidence remains inconclusive. To address this issue, the study analyzes the use of *be* in L2 English spoken by learners with topic-prominent L1s (viz., Mandarin Chinese, Japanese, Korean) and those with non-topic-prominent L1s (viz., Filipino). Using the ICNALE corpus, the study investigates whether L1 topic-prominence affects the use of *be* in L2 English. The findings reveal that L2-English learners with topic-prominent L1s use copula *be* more often with definite NPs than those with non-topic-prominent L1s, indicating the influence of L1 topic-prominence properties on the use of copula *be*. Additionally, learners from topic-prominent L1s more frequently produce nontargetlike topic-comment structures using *be* verb (e.g., *restaurant is many people eat their foods*). Based on these findings, the study discusses the implication of L1 transfer effects in L2 acquisition and emphasizes the importance of a more global and refined understanding of language variation in L2 teaching and learning.

Keywords: English *be* verb; second language acquisition; L1 transfer; topic markers; corpus analysis

***Corresponding author: Kitaek Kim**, Department of English Language Education, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea, E-mail: kitaek@snu.ac.kr. <https://orcid.org/0000-0002-1242-034X>

Min-Chang Sung and Bora Nam, Department of English Language Education, Gyeongin National University of Education, Incheon, Republic of Korea. <https://orcid.org/0000-0001-9561-6374> (M.-C. Sung). <https://orcid.org/0000-0001-8872-8941> (B. Nam)

1 Introduction

The influence of first language (L1) has received considerable attention in the field of second language (L2) acquisition but still remains elusive. Many studies have claimed for L1 transfer effect based on performance similarities between L1 and learners' interlanguage (see Foley and Flynn 2013, for a review). The interpretation of such interlanguage performance, or errors, however, is actually ambiguous, as similar errors have also been observed in L1 acquisition or among other groups of L2 learners who do not share relevant L1 properties (Brown 1973; Dulay and Burt 1974). In this regard, Jarvis (2000) discussed that 'performance similarities between L1 and interlanguage' do not suffice to conclude L1 transfer but require additional data such as 'intra-L1-group similarities' (i.e., similarities between people with the same L1) and 'inter-L1 group differences' (i.e., differences between groups with different L1s) (see Schwartz and Sprouse 2000 for a similar discussion). Considering the three-way methodological requirements (Jarvis 2000), the current study addresses an unresolved L1 transfer issue – L2 learners' use of *be* as a topic marker.

The English verb *be* is one of the earliest morphemes acquired by L2-English learners (Zobl and Liceras 1994) and has received significant attention in L2 research. Notably, nontargetlike *be*-overuse (e.g., *John is love Mary*) has been reported with L2-English learners from various L1 backgrounds, and many researchers have attempted to identify the source of this error. Many studies have claimed that such an error is due to L1 transfer (Hahn 2000; Kim 2011; Nam 2020, among others), specifically as a morphological realization of topic-coding (i.e., linguistic methods or strategies used to identify or mark the topic of a sentence) that have transferred from their L1s, which are topic-prominent languages organizing sentences primarily around the topic-comment structure. Note that these languages are distinct from subject-prominent languages, which organize sentences based on the subject-predicate structure (Li and Thompson 1976). Many of these studies, however, have methodological limitations since their claims are solely based on performance similarities between L1 and interlanguage within a single L1 group and do not meet the methodological requirements outlined by Jarvis (2000).

Building on the studies of L1 influence, particularly regarding topic-coding as a feature observed across different languages, we hypothesize a more pronounced association between the use of *be* and topic-coding in L2-English learners from L1 backgrounds that exhibit similar topic-prominence characteristics. This relationship is expected to be more evident among L2 learners whose L1s significantly differ, specifically between those from topic-prominent and non-topic-prominent language backgrounds. Therefore, the current study adopts 'intra-L1-group similarities' and 'inter-L1-group differences' (Jarvis 2000) to explore 'intra-topic-prominent-L1-group similarities' (i.e., similarities between L1 groups whose

languages share similar topic-prominence) and ‘inter-topic-prominence-contrastive-L1-group differences’ (i.e., differences between L1 groups whose languages differ in their topic-prominence).

To delve into this collective topic-prominent L1 effect, the current study explores the use of *be* in L2 spoken English corpora from the International Corpus Network of Asian Learners of English (ICNALE: Ishikawa 2019). The analysis is based on the hypothesis that *be* functions as a topic marker in interlanguage grammar, drawing on prior studies discussed in Section 2. Specifically, we examine whether L2 English learners with different topic-prominent L1s (Mandarin Chinese [hereafter, Chinese], Japanese, Korean) display similar usage patterns of *be* and compare their usage to that of learners from a non-topic-prominent L1 (Filipino). This contrastive analysis broadens the applicability of our findings, offering valuable insights for L2 teachers to develop tailored teaching strategies grounded in an understanding of interlanguage (Ortega 2012).

2 Literature review

This section reviews the use of the English copula *be*, highlighting its potential topic-marking properties. It then discusses topic-comment structures in languages such as Chinese, Japanese, and Korean. Additionally, it examines previous findings on L2 use of the English *be* and identifies the research gaps that the present study aims to address.

2.1 Copula *be* and topic marking

Be can be divided into auxiliary and copula.¹ They differ from each other in formal and functional aspects. Auxiliary *be* denotes aspect and/or voice of the verb phrase when used with a present participle (progressive; e.g., *he is running*) or a past participle (passive; e.g., *it is cancelled*). Copula *be* represents a relation between a subject and a predicate (e.g., *I am a doctor*), serving as a ‘linking’ verb (Halliday 2014; Stowell 1981) with little content meaning (Partee 1986).

While there are a number of proposals on the notion of ‘topic’ (among many others, Reinhart 1981; Gundel 1988), ‘givenness’ and ‘aboutness’ encompass all these proposals (Roberts 2019, p. 406): (a) A topic must be something that is either familiar (or given) itself or is an identifiable member of some familiar set of entities

¹ There is another type, existential *be*, which expresses the presence of something, mostly in the *there*-constructions (e.g., *there was a car accident*). The current study put aside this existential *be* from discussion for its irrelevance to the focus of this study.

(givenness); (b) a topic is what the utterance in which it occurs is about (aboutness). Notably, there is a clause type whereby copula *be* should follow a topic, as acknowledged in Mikkelsen (2005, p. 161): “almost all researchers agree that the subject of a specificational clause is topic.”² See (1a-b) for examples.

- (1) a. *A doctor is John.
b. The doctor is John. (Mikkelsen 2005, p. 154)

Mikkelsen claims that the subject of a specificational copula clause should be topic based on the observation that the subject NP requires a definite NP subject (*the* + NP; as in 1b), but not an indefinite NP subject (*a* + NP; as in 1a). This observation is particularly interesting for this study. Exploring which article – *the* or *a* – occurs with a subject NP may indicate the topicality of the subject NP in part, given that (a) topic can be defined in terms of givenness, which is related to the familiarity to the hearer (e.g., hearer-old, hearer-new; Prince 1981, 1992), and (b) that the hearer-old information is expressed with *the* and the hearer-new information is with *a* or zero-articles (except generic expressions) (Bickerton 1981; Hawkins 2015; Huebner 1983).

Given that ‘comment’ is what is being said about the topic (Gundel 1974), copula clauses containing ‘topic’ subjects may reflect a topic-comment structure. Den Dikken (2005) supports this idea, claiming that standard English, along with Belfast English which shows a strong tendency to use the topic-comment structure (Finlay 1988, p. 691), exploits the topic-comment structure for some copula constructions (see O’Neill 2015; Partee 2010 for a similar account). He then provides the syntax for this topic-comment structure in which the first constituent is “the *topic*, occupying the specifier position of a TopP” whereby the head takes the second constituent, “the *comment*, as its complement” (p. 708).

The topic-comment structure is in fact a prominent feature in East Asian languages that are categorized as topic-prominent languages, such as Korean, Japanese, and Chinese (Li and Thompson 1976). In these languages, there are topic markers widely used in various types of topic-comment structures. For example, in Korean, the topic marker *-(n)un* is used between a topic and a comment, as in (2), where the topic marker is placed between the topic (teacher) and the comment (very busy).

- (2) Sensayngnim-un maywu pappu-ta.
teacher-TOP very busy-DECL³
‘The teacher is very busy.’

² Mikkelsen (2005) distinguishes ‘specificational copula clause’ from ‘predicational copula clauses’ (e.g., *John is the teacher*), claiming that the subject NP of ‘specificational copula clauses’ is a topic.

³ Abbreviations: ACC – accusative; DECL – declarative; NOM – nominative; PRS – present; PST – past; TOP – topic.

Notably, the topic marker shares some superficial similarities with the English copula *be*. It typically follows the topic NP; it does not carry specific content meaning, typically not affecting the propositional meaning of the sentence. Moreover, since topic is universally grammaticized as subject most strongly (Talmy 1983), topic markers – placed between topics (usually the subject) and comments – appear parallel to copulas, which are placed between the subject and predicate. These similarities are also found with the Japanese topic marker *-wa*, as in (3), where the topic marker follows the topic NP *sensei* ‘teacher’.

- (3) Sensei-wa totemo isogashi-desu.
 teacher-TOP very busy-PRS.DECL
 ‘The teacher is very busy.’

In Chinese, another topic-prominent language that employs a topic-comment structure, topics are generally marked by initial placement. Notably, the copula *shì* can also function similarly to Korean *-n(un)* and Japanese *-wa* in certain circumstances, as it is typically placed after the topic (Chan 2004; Hsieh 2009) and it should follow the topic in specificational constructions (Tham 2008), as shown in (4a).⁴ Unlike Korean *-(n)un* and Japanese *-wa*, however, the Chinese *shì* cannot take an Adjective Phrase (AP) as its complement (Hsieh 2009). For example, (4b) is ungrammatical as *shì* cannot take the AP *hěn máng* ‘very busy’ as its complement.

- (4) a. Lǎoshī shì Sanmao.
 teacher COP Sanmao
 ‘The teacher is Sanmao. / *A teacher is Sanmao (Tham 2008, p. 73)
 b. Lǎoshī (*shì) hěn máng.
 teacher very busy
 ‘The teacher is very busy.’

While we discussed some similarities among English *be*, Korean *-(n)un*, Japanese *-wa*, and Chinese *shì* with regards to topic marking, English has noteworthy differences from the other topic-prominent languages. One such difference is regarding so-called double subject constructions (Li and Thompson 1976). The double subject construction is considered to have two subjects: one is a topical subject, and the other is the subject of a clause-level comment, as demonstrated in the Korean sentence (5), where the topic marker *-nun* is placed between the topic *khokkili* ‘elephant’ and a

4 In the literature on information structure, studies have examined the use of the Chinese word *shì* and the Korean marker *-(n)un* as focus or contrastive focus markers. For example, Zhan (2012) discussed how phrases following *shì* can receive focus; similarly, *-(n)un* in Korean can serve as a contrastive marker when a specific prosody is employed (Lee 2006). It would be worthwhile to explore in a future study whether the English copula *be* can also possess a focus marking property that is transferred from L1s, although this is beyond the scope of the present study.

clause-level comment consisting of a subject *kho* ‘nose’ and a predicate *kil-ta* ‘long’. Notably, the English translation equivalent in (5) does not allow copula *be* between the topic and the clausal comment.⁵ Instead, when the comment is a clause, a separate topic marking expression, such as *as for*, should be used.

- (5) Khokkili-nun kho-ka kil-ta.
 elephant-TOP nose-NOM long-DECL
 ‘As for elephants, their nose is long’.

Second, while the topic markers in Korean and Japanese and *shì* in Chinese can be followed by a verbal predicate, as in (6a-c), respectively, English copula *be* does not allow a verbal predicate as its complement, as in (6d).

- (6) a. Na-nun ece sensayngnim-ul manna-ss-ta.
 I-TOP yesterday teacher-ACC meet-PAST-DECL
 b. Watashi-wa kino sensei-o ai-mashita.
 I-TOP yesterday teacher-ACC meet-PAST-DECL
 c. Wǒ shì zuótiān kàndào wǒ de lǎoshī.⁶
 I TOP yesterday meet my teacher
 d. *I am met my teacher yesterday.

If certain features are expressed through overt morphology in L1, but with zero morphology in L2, L2 learners may assume the target language also has overt morphology to express those features (Lardiere 2008; Montrul 2001). In other words, L2 learners may look for substitute L2-specific lexical items to host these features. In this regard, previous research has focused on *be* as a substitute for L1 topic markers. If L2-English learners with topic-prominent L1s indeed use copula *be* as a substitute for the topic marker, the linguistic analysis thus far yields the following predictions:

- (1) L2-English learners with topic-prominent L1s will use copula *be* productively in the course of representing topic-comment structure; in turn, hearer-old NPs will likely appear in the subject position of copula *be* more frequently than hearer-new NPs. One way to test this is, as discussed with Mikkelsen’s (2005)

5 Unlike Korean *-(n)un* and Japanese *-wa*, we acknowledge that Chinese *shì* does not occur between the topic and the comment in a double subject construction, as seen in (ii).

(ii) Dàxiàng bízǐ cháng
 Elephant nose long
 ‘As for elephants, their nose is long.’

One reason *shì* cannot be used after ‘elephant’ is that *shì* cannot take a descriptive clause as its complement.

6 (6c) is a the medial bare *shì* construction assigning focus on any constituent to the right of *shì*, accordingly making the subject the topic (Paul and Whitman 2008).

examples (1a-b), by analyzing whether the definite article *the* (as opposed to the indefinite *a(n)*) appears with the subject NP in a sentence containing copula *be*.

- (2) When L1-Chinese learners use copula *be* as a substitute for Chinese *shì*, the use may be restricted to NP complements, compared to learners of other L1-topic prominent languages.

2.2 L2 usage of *be* verb

Numerous studies have explored how a common feature shared across different L1s similarly influences L2 acquisition (Cadierno 2010; Stam 2006). It is widely acknowledged that L2 learners are similarly influenced by the same feature across different L1s, with this impact being pronounced when the L1 and L2 differ significantly in those features (Jeon and Sung 2023). These collective L1 effects have also been observed in L2 usage of English verb *be*. A myriad of studies, mostly based on the analysis of L2 learner corpus, have highlighted how L2 usage varies depending on the topic-prominence of the learners' L1s, and investigated the nontargetlike use of *be* (e.g., John *is* want pizza) by L2 learners whose L1 belongs to topic-prominent languages (e.g., Chinese, Japanese, Korean, Lao). They suggest that such nontargetlike *be* serves as a topic-marker that creates a link between a topic and a comment, as in L1 (Hahn 2000, among many others).⁷ For example, Shin (2000) speculated that Korean learners' erroneous production of *she is like music* reflects *be* as being used as a topic marker, meaning *as for her, she likes music*. Similarly, Huebner (1983) found that his Lao learner of English used *is* or *isa*, as in (7), to distinguish presupposition from assertion, or topic from comment, and concluded that *is(a)* is a topic-comment boundary marker.

- (7) Hua aen song, isa bodii sik.
 “As for Hua and Song, their bodies were sick.” (Huebner 1983, p. 79, (3.38))

Given that the topic-prominence of L1 affects nontargetlike *be* in L2 English, similar effects may exist for targetlike *be* in L2 English. In particular, there are notable similarities among topic markers and copula *be* placing between a topic and a comment, as discussed in Section 2.1. It has been suggested that copula *be* may have topic-marking properties transferred from topic-prominent L1s (Hahn 2000; Nam 2020; Shin 2000). For example, Nam (2020) conducted an elicited production task in

⁷ It was also claimed that the nontargetlike *be* is an early morphological marker of a functional category, such as tense, agreement, or aspect (among many others, Ionin and Wexler 2002), which we do not investigate further because this study focuses on L1-transfer. Nevertheless, it is worthwhile to note that the functional category view mostly examined participants with non-topic-prominent L1s (Kim 2011; Nam 2020).

which participants were asked to complete a sentence following a prompt word, as exemplified in (8). The task was administered for three groups (native speakers of English, L1-Russian learners, and L1-Korean learners), and it was found that only the Korean learners frequently used copula sentences, as in (9), comprising 12 % of their responses.

- (8) Q: When do Sara's brothers play tennis and ping pong?
A: Tennis _____, ping pong _____.
- (9) Tennis *is every Sunday* and ping pong *is every Friday*.
- (10) Theynisu-nun ilyoil-ey chi-ko thakkwu-nun kumyoil-ey
tennis-TOP Sunday-at play-and ping pong-TOP Friday-at
chin-ta.
play-DECL
'As for tennis, he plays on Sundays, and as for ping pong, he plays on
Fridays.'

In (8), the question introduces two topical nouns *tennis* and *ping pong*. In (9), the L1-Korean participant puts *is* between each topical noun and its comment. This topic-comment structure appears quite similar with the topic-comment structure in the Korean translation equivalent in (10), in which the topic marker *-nun* is placed between a topic and a comment. Based on the finding, Nam (2020) speculated that *tennis is every Sunday* might not be a typical copula sentence, but rather a topic-comment structure.

Lee and Huang (2004) report more subtle effects of L1 topic-prominence features as they examine the use of *be* by L1-Chinese 4th graders ($n = 270$) in a primary school in Hong Kong. They collected production data from a story-writing task and found that the children performed better on the use of copula *be* with a noun (e.g., *I am a king*) than with an adjective (e.g., *I am good*). Assuming that the Chinese *shì* corresponds to copula *be* in their interlanguage, the gap in the use of copula *be* before adjectives and nouns may reflect the transfer of the L1 configurational regulation, which disallows *shì* before adjectival comments but allows it before nominal comments (for a similar finding, see Hsieh 2009).

Although these previous studies have suggested the effects of topic prominent L1s on the use of *be* in interlanguage, several gaps still remain. Firstly, previous research mainly examines nontargetlike *be* (e.g., *she is like music*) to investigate the transfer of topic prominence. Only a few studies have discussed the possibility that both copula *be* and nontargetlike *be* have the same function, a topic-marker property in the interlanguage grammar of L1-topic-prominent languages. If *be* serves as a topic marker in interlanguage due to L1 transfer, then copula *be* and the nontargetlike *be*

may not be essentially different for these learners. Secondly, most of the previous studies claiming an L1-effect (e.g., use of *be* as an L1-topic-marker equivalent, frequent use of NP complements by L1-Chinese learners) are solely based on L1-interlanguage similarities. Such ‘L1-interlanguage similarities’, however, are not sufficient to conclude L1 transfer (Jarvis 2000) rather than universal L2 strategies if there are no ‘inter-L1 group differences’. Finally, previous studies exploring *be* in L2 speaking tend to examine a small number of L2 learners, which may limit the generalizability of findings. Therefore, the investigations on use of *be* in a large collection of spoken data (e.g., a learner corpus) are needed to overcome this limitation.

To address these limitations, the current study explores a topic-marker property in both copula *be* and nontargetlike *be*. To understand comprehensively how the topic-prominence of the L1 affects the use of these two types of *be* in interlanguage, learner corpora from three topic-prominent L1 (Chinese, Japanese, Korean) learner groups and a non-topic-prominent L1 (Filipino) learner group are compared.

Based on the linguistic similarities of copula *be* and topic markers in topic-prominent L1s, as discussed in Sections 2.1 and 2.2, the research questions of this study are formulated as follows.

- 1) Do the learners from topic-prominent L1s use copula *be* more frequently than those from non-topic-prominent L1s to express topic-comment relations?
- 2) Does the use of *be* and its variants by the learners from topic-prominent L1s show topic-marking features?
- 3) Does the topic-marking variation among the topic-prominent L1s affect the use of copular *be*?

With the hypothesis that learners from topic-prominent L1s are likely to associate *be* verbs with topic-coding due to L1 influence, the expected results are as follows. First, learners from topic-prominent L1s are anticipated to show a higher proportion of copula *be* relative to auxiliary *be*, as the latter primarily encodes aspect or voice rather than topic-comment relations. Second, the proportion of *the*-NP subjects versus *a(n)*-NP subjects is expected to be higher in topic-prominent L1 groups, along with more nontargetlike uses of *be*, such as double-subject constructions or *be*-insertion before finite thematic verbs due to the transfer of L1 topic-comment structures. Finally, within the topic-prominent L1 groups, L1-Chinese learners are expected to show a higher frequency of NP complements due to L1 transfer effects on the complementizer type used with the copula.

3 Methods

3.1 Corpora

The present study analyzed the International Corpus Network of Asian Learners of English (ICNALE: Ishikawa 2019). This corpus contains a variety of written and spoken production data such as essays, monologues, and dialogues. Among them, we specifically selected the native monologue corpus and four learner monologue corpora – viz., Chinese, Japanese, Korean, and Filipino – to collectively examine the impact of L1-topic-prominence on spontaneous production of *be* verbs. Other learner corpora in the ICNALE were not considered in this study due to data-related constraints such as their small size (e.g., Hong Kong corpus with only 50 participants), the inclusion of participants with multiple L1s (e.g., L1-Urdu and L1-Punjabi participants in the Pakistan corpus), or regional variations of the Chinese language (e.g., Singapore and Taiwan corpora), which would compromise the consistency and comparability required for this analysis.

The speech data of the five target corpora were collected using the same procedure. Participants gave a self-introduction and made four 60-s argumentative speeches: (1) the first speech about the first topic, (2) the second speech about the first topic, (3) the first speech about the second topic, and (4) the second speech about the second topic. The repetition of the same topics was based on the observation that a single recording does not provide enough data to represent one's speaking performance (Ishikawa 2014).

The two speech topics were college students' part-time jobs and a ban on smoking in restaurants. These topics were presented orally using the following prompts: (1) *It is important for college students to have a part-time job*, and (2) *Smoking should be completely banned at all the restaurants in the country*. The preparation time was 20 s for the first speeches and 10 s for the second speeches. The audio-recorded speeches were transcribed and tagged with parts of speech (hereafter, POS) on the Sketch Engine System, which uses the grammar rule of the Penn Treebank Project (Taylor et al. 2003).

The five target corpora consisted of 2,600 recordings produced by 650 participants. The native corpus included 600 recordings by 150 native English adult speakers, totalling 91,967 words, while the learner corpora had 2,000 recordings produced by 500 English learners, totalling 185,761 words.

The learner participants were classified by the proficiency levels of Common European framework of reference for languages, namely the CEFR (Council of Europe 2020): the proficiency classification was based on an L2 vocabulary size test (Nation and Beglar 2007) and their scores in standardized proficiency tests such as

Table 1: Number of recordings (words) for B1 level in the learner corpora.

Filipino	Chinese	Japanese	Korean
352 (43,554)	504 (55,443)	360 (24,132)	232 (15,468)

TOEFL and TOEIC (Ishikawa 2019). The CEFR describes six reference levels for L2 proficiency, from A1 to C2, which can be grouped into three types of language users: basic (A1 and A2), independent (B1 and B2), and proficient user (C1 and C2). All the learners were as A2 (waystage), B1 (threshold), or B2 (vantage) level on the CEFR proficiency scale.

Since controlling for L2 proficiency is crucial when testing a particular between-group factor in L2 acquisition studies (Schwartz and Sprouse 2000), we focused primarily on analyzing data from the B1 group, the group with largest data (Table 1).

According the CEFR scale for overall oral production, B1-level learners can “fluently sustain a straightforward description of one of a variety of subjects within their field of interest” (Council of Europe 2020, p. 62).

3.2 Analysis

Every case of *be* and its variants was coded to one of the three major functions: auxiliary, copular, and existence. We extracted every case of *be* verbs in the native and the four learner corpora, using the six relevant POS tags of the Sketch Engine System (Taylor et al. 2003): VB (*be*), VBP (*am*, *are*), VBZ (*is*), VBD (*was*, *were*), VBG (*being*), and VBN (*been*). Then, we removed inaccurately tagged cases, such as mistaking the adverb *maybe* for the modal verb *may* and the verb *be*. This left us with a total of 14,733 instances of *be* verbs across all five corpora.

However, many instances were challenging to categorize into one of the three major functions. Most of them were incomplete *be* verb phrases as in self-correction (e.g., *he is*, *he has a ball*) or a *be* verb followed by its repetition (e.g., *he is*, *is tall*). These cases were coded as “incomplete”. We also found many cases of nontargetlike *be* verbs, a common error in L2 English (e.g., *he is make money*), so we used a separate code for these nontargetlike *be*.

The coding of *be* and its variants was based on the neighboring words and their POS tags. For example, a *be* verb that is followed by present participles (e.g., *is running*) or past participles (e.g., *be cancelled*) was coded as auxiliary *be*, while *be* and its variants that are followed by adjectives (e.g., *is happy*) or determiners (e.g., *am a*) were coded as copular *be*.

This automatic classification was complemented by manual inspection to guarantee high accuracy in verb coding. For example, the verb *is* in the sentence *it is designated area* should be coded as copula *be*, but it was automatically coded as auxiliary *be* because it was followed by a past participle. To correct this sort of coding error, a considerable amount of manual coding was conducted by looking into the entire sentences. This enabled us to correct the automatic coding errors in approximately 500 cases of *be* verbs.

The coding of subject articles preceding copula *be* was also based on the neighboring words and their POS tags. Firstly, singular nouns preceding copula *be* were identified, and then the *a* and *the* uses preceding up to four words away from them were coded. If articles were not adjacent to nouns, the intervening words were restricted to nouns (e.g., *part-time*), adjectives (e.g., *good*), adverbials (e.g., *frequently*), and present/past participles (e.g., *smoking*, *known*).

The coding of *be* and other elements was primarily carried out by the first author, with each coding instance reviewed by the third author. The coding process was iterative, and revisions continued until both authors reached full agreement. The data file can be found at https://osf.io/ru3tg/?view_only=36b63a6a7f594292b3da22b46551c092.

After the preliminary analysis of *be* usage in the four corpora, frequency-based analyses were conducted to determine whether and to what extent different topic-prominent L1s affect English learners' use of *be* verbs.

- 1) To investigate the frequency distribution of the *be* verb, the proportion of copula *be* as opposed to auxiliary *be* was compared across the groups.
- 2) To explore the topic-marking features of *be* and its variants, the proportions of “*the* + singular NP” subjects and “*a(n)* + NP” subjects before copula *be* were analyzed across groups. In addition, to identify distinctive uses of *be* variants presumably due to the transfer of topic-prominent L1, a total of 133 cases of errors with nontargetlike *be* were found and analyzed.
- 3) Finally, to answer whether the intra-topic-prominent-L1 variation may affect L2 use of copula *be*, the uses of NP and non-NP complements after copula *be* were compared in the three learner corpora with topic-prominent L1s (viz., Chinese, Japanese, and Korean).

4 Results

4.1 Frequency of copula *be*

Table 2 presents the frequency of *be* verbs across the native and four learner corpora, along with their standardized frequencies assuming each corpus had a word count of

Table 2: Frequencies of *be* verbs across the corpora.

Corpus	Native	Learner			
		Filipino-B1	Chinese-B1	Japanese-B1	Korean-B1
Corpus size (words)	91,967	43,554	55,443	24,132	15,468
Observed frequency	5,362	2,492	2,432	1,152	838
Standardized frequency	5,830	5,722	4,386	4,774	5,418

The standardized frequencies assumed that each corpus has one hundred thousand words.

100,000. The standardized frequency of *be* verbs was highest in the native corpus (=5,830), and lowest in the Chinese-B1 corpus (=4,386). The chi-square test based on the frequency of *be* verbs and the corpus size indicated a significant difference among the five corpora, $\chi^2 = 171.87$, $p < .001$.

Table 3 shows the frequency distributions of auxiliary *be* and copula *be* across the corpora. The proportion of copula *be* was higher than that of auxiliary *be* in every group. In particular, there were greater gaps observed among the topic-prominent L1 groups, where the frequency of copula *be* was over three times greater than that of auxiliary *be* (e.g., 1,560 versus 424 in the Chinese group). Meanwhile, the frequency ratio of copula *be* and auxiliary *be* was about two-to-one in the native and Filipino groups.

According to the chi-square test, there was a significant difference in the frequency proportions of copula *be* and auxiliary *be* among the five corpora ($\chi^2 = 244.32$, $p < .001$). Post-hoc analyses of pairwise comparisons with Bonferroni correction of $\alpha = .005$ ($=0.05/10$) revealed that the topic prominence-contrastive-L1 pairs (viz., a pair of a non-topic-prominent L1 and a topic-prominent L1; e.g., native speakers versus Chinese learners) had greater chi-square statistics than the intra-topic-prominent-L1 pairs (e.g., native speakers versus Filipino learners), as shown in Table 4.

Table 3: Frequency distributions of auxiliary *be* and copula *be* in the corpora.

Corpus	Copula <i>be</i>	Auxiliary <i>be</i>	Sum
Native speaker	3,256 (67.3 %)	1,582 (32.7 %)	4,838 (100 %)
Filipino	1,524 (67.0 %)	750 (33.0 %)	2,274 (100 %)
Chinese	1,560 (78.6 %)	424 (21.4 %)	1,984 (100 %)
Japanese	748 (83.3 %)	150 (16.7 %)	898 (100 %)
Korean	577 (85.9 %)	95 (14.1 %)	672 (100 %)

Table 4: Chi-square statistics of post-hoc comparisons for copula *be* and auxiliary *be* in the native and B1-learner corpora.

Post-hoc pair		Chi-Square value	<i>p</i>
Topic-prominence-contrastive-L1 pairs	Native-Chinese	86.99	***1.1E-20
	Native-Japanese	91.94	***8.9E-22
	Native-Korean	96.03	***1.1E-22
	Filipino-Chinese	71.53	***2.7E-17
	Filipino-Japanese	83.93	***5.1E-20
	Filipino-Korean	90.05	***2.3E-21
Intra-topic-prominent-L1 pairs	Native-Filipino	0.06	0.813
	Chinese-Japanese	8.44	*0.004
	Chinese-Korean	16.71	***4.4E-05
	Japanese-Korean	1.92	0.166

p* < .005, *p* < .001, ****p* < .0001.

All six topic-prominence-contrastive-L1 pairs showed significant differences in the frequency distribution of copula *be* and auxiliary *be* at $\alpha < .0001$ ($=0.001/10$). In contrast, only two of the four intra-topic-prominent L1 pairs showed significant differences, mostly due to the higher frequencies of copula *be* in the Japanese B1 and Korean B1 corpora. The mean chi-square value for the significant topic-prominence-contrastive-L1 pairs was 86.75, while that of the significant intra-topic-prominent L1 pairs was 12.58. These results indicate that the use of copula *be* was similar among groups whose L1 is topic-prominent, while the use of copula *be* varied among groups with non-topic-prominent L1s.

4.2 Topic-marking features of *be* and its variants

The relationship between *be* and topic-marking was examined by comparing the proportions of copula *be* and auxiliary *be* between topic-prominent L1 groups and the others. Nontargetlike uses of *be* were also identified and analyzed in each group.

4.2.1 Use of articles in subject NPs of copula *be*

The proportions and frequencies of copula *be* depending on the subjects are shown in Table 5. Note that after “*the* + singular NP” subjects, the proportion of copula *be* was much higher in the Chinese, Japanese, and Korean learner corpora than the native and Filipino corpora. Specifically, Chinese, Japanese, and Korean learners used *the*-NP subject more frequently than *a(n)*-NP subject before copula *be*, with a range of

Table 5: Frequency proportions of copula *be* after *the*-NP and *a(n)*-NP subjects.

Corpus	<i>The</i> -NP Subject	<i>A(n)</i> -NP Subject	Sum
Native	81 (60.4 %)	53 (39.6 %)	134 (100 %)
Filipino	38 (42.7 %)	51 (57.3 %)	89 (100 %)
Chinese	94 (77.0 %)	28 (23.0 %)	122 (100 %)
Japanese	63 (88.7 %)	8 (11.3 %)	72 (100 %)
Korean	52 (83.9 %)	10 (16.1 %)	62 (100 %)

77.0 % (Chinese) to 88.7 % (Japanese). In contrast, Filipino learners showed an opposite pattern, using *a(n)*-NP subject (57.3 %) more frequently than *the*-NP subject (42.7 %) before copula *be*.

The chi-square test found that the five corpora were significantly different in their subject article choice before copula *be* ($\chi^2 = 82.792, p < .001$). The post-hoc analyses of pairwise comparisons with $\alpha = .005$ ($=0.05/10$) reported greater chi-square statistics among the topic-prominence-contrastive-L1 pairs than the topic-prominent-L1 pairs (see Table 6). All the six topic-prominence-contrastive-L1 pairs showed significant differences in the frequency distribution of *the*-NP and *a(n)*-NP subjects before copula *be* (all $ps < .005$), while none of the topic-prominent-L1 pairs showed statistical significance at $\alpha < .005$. These results indicate that English learners with topic-prominent L1s used copula *be* with *the*-NP subject more frequently than those with non-topic-prominent L1s.

Table 6: Chi-square statistics of post-hoc comparisons for copula *be* and auxiliary *be*.

Post-hoc pair		Chi-Square value	<i>p</i>
Topic-prominence-contrastive-L1 pairs	Native-Chinese	8.14	0.004*
	Native-Japanese	17.76	2.5E-05***
	Native-Korean	10.66	0.0011*
	Filipino-Chinese	25.93	3.5E-07***
	Filipino-Japanese	35.96	2E-09***
	Filipino-Korean	25.73	3.9E-07***
Intra-topic-prominent-L1 pairs	Native-Filipino	6.77	0.009
	Chinese-Japanese	4.04	0.045
	Chinese-Korean	1.17	0.280
	Japanese-Korean	0.67	0.414

* $p < .005$, ** $p < .001$, *** $p < .0001$.

4.2.2 Nontargetlike *be*

From the B1 learner corpora, a total of 100 cases of nontargetlike *be* were identified. The frequencies of nontargetlike *be* were higher in the learner corpora with topic-prominent L1 than the corpus with non-topic-prominent L1: 37, 32, and 20 cases in the Chinese, Japanese, and Korean corpora, respectively, but only 11 cases in the Filipino corpus. The standardized frequency assuming a corpus size of one hundred thousand words was highest in the Japanese corpus (=132.6), followed by the Korean (=129.3), the Chinese (=66.7), and the Filipino corpus (=25.3). For qualitative analyses of collective topic-prominence L1 effects, the cases of nontargetlike *be* were categorized into four patterns:

– <i>be</i> + uninflected verb:	smoke is feel bad (Japanese)
– <i>be</i> + inflected verb:	it's also hurts (Chinese)
– <i>be</i> + modal verb:	The part-time job is must needing that (Korean)
– <i>be</i> + commentary sentence:	Because restaurant is many people eat their foods and talk about their affairs (Korean)

The majority of the cases (79 cases) were in the “*be* + uninflected verb” pattern, which was the most frequent across all learner corpus (for the frequency data of nontargetlike *be* in each corpus, see Appendix). The nontargetlike *be* in this pattern include both inflected forms (e.g., *am*, *are*, *is*) and bare form (viz., *be*), as in (11) and (12), respectively, and was often followed by one or two non-verb components such as *not*, *really*, and *sure*, as in (13). In the “*be* + uninflected verb” pattern, the nontargetlike *be* usually appeared to agree with the tense, person, and number (14).

- (11) a. I *am* quite agree with the first opinion (Chinese)
b. because working *is* help us for growing mental (Japanese)
- (12) a. so to *be* have good balance of body uh the students have to do a job (Korean)
b. you wouldn't *be* go to the shock (Filipino)
- (13) a. that's not really do any good for people (Filipino)
b. It will *be* more surely reduce the amount of people (Chinese)
- (14) still I *was* not work (Japanese)

The second most frequent pattern (13 cases) was “*be* + inflected verb,” which was also observed in all learner corpora. In four cases, the nontargetlike *be* and inflected verbs appeared to have the same tense/agreement information, such as 3rd person singular present (15a) or past (15b). However, in the remaining nine cases, the nontargetlike *be* and inflected verbs did not match either in tense or agreement. For

example, as shown in (16a-b), a bare-form nontargetlike *be* was followed by a past-tense-marked thematic verb.

- (15) a. It's also hurts (Chinese)
 b. my grandfather was died because of the smoking cancer (Japanese)
- (16) a. it can also be relied on the student itself (Filipino)
 b. because smoking is stressed my body (Korean)

While it is possible that the learners used the patterns “*be* + uninflected verb” and “*be* + inflected verb” to create a topic-comment structure, it is difficult to draw a definite conclusion since these patterns also appeared in the Filipino corpus as well. Additionally, some examples may have different reasons for using these patterns, such as overpassivization, as seen in (16a).⁸

While these two frequent patterns, “*be* + uninflected verb” and “*be* + inflected verb”, were found in all learner corpus, the other two patterns were produced by only those with topic-prominent L1. The “*be* + modal verb” pattern, with just three cases, was exclusive to topic-prominent L1 learners. In these instances, a non-targetlike *be*-form was inserted before modal verbs like *will*, *can*, and *must* as in (17). Notably, the subjects in all three examples were related to the speech topics: *part-time job* in (17a) and to *arubaito* (17b), meaning part-time job in Japanese, related to the first speech topic, while *it* in (17c) referred to smoking, the second speech topic.

- (17) a. the part-time job is must needing that (Korean)
 b. to arubaito is can do in other time (Japanese)
 c. it's will relax, a way of relaxation (Chinese)

Another nontargetlike use of *be* verbs supporting its relation to topic marking was found in the “*be* + commentary finite clause” pattern, with four instances in the Korean corpus and one in the Japanese corpus. In this pattern, *be* verbs are placed between topics and commentary finite clauses about the topics. For example, in (18a), the Japanese learner linked the topic *part-time job* and an advantage for university

⁸ We also found double *be* errors (“X + *be* + Y + *be*”) in Chinese and Japanese learner corpora and classified them into the “*be* + inflected verb”. Although this pattern seems to have two topics (viz., X and Y) if assumed that the *be* verbs mark topics, a close examination reveals that it actually involves a repetition of *be* with a single topic. For example, in (iiiia), the word in the Y position is an adjunct, which cannot be considered as a topic. In (iiiib), the word in the Y position is an embedded subject of a *wh*-phrase. In a typical *wh*-question, *be* should precede the subject (e.g., what is society?), while in an indirect speech as in (iiiib), *be* should follow the subject, which may have confused the participant to produce *be* twice in a single sentence.

- (iii) a. the society is still is very important (Chinese)
 b. they can know what is society is (Japanese)

students with *is*. Similarly, using *is*, the Korean learner linked the topic *restaurant* and a clause supporting the smoking ban in restaurants in (18b).

- (18) a. part-time job is they can, they can get money (Japanese)
b. restaurant is many people eat their foods (Korean)

All in all, learners of English with topic-prominent L1 produced nontargetlike *be* errors more frequently than those with non-topic-prominent L1. In addition, they produce peculiar types of nontargetlike *be* to mark sentential topics (e.g., *be* + modal verb and *be* + commentary finite clause), showing effects of topic-prominent L1 on the nontargetlike uses of *be* verbs.

4.3 Intra-topic-prominent L1 group variations in complements of copular *be*

To examine the effect of the intra-topic-prominent-L1 variation concerning the complements copular *be* can take, the use of NP and non-NP complements by learners from three topic-prominent L1s was analyzed, as shown in Table 7.

L1-Chinese learners used an NP complement for copula *be* more frequently than L1-Japanese and L1-Korean learners. The chi-square test revealed that the frequency distributions of NP and non-NP complements for copula *be* were significantly different between Chinese and Korean ($\chi^2 = 10.13$, $p = .0015 < .01$) and between Chinese and Japanese corpora ($\chi^2 = 18.79$, $p < .001$), but not between Korean and Japanese corpora ($\chi^2 = 0.49$, $p = .48$). In other words, Chinese learners used NP complements after copula *be* significantly more frequently than Japanese and Korean learners.

For additional evidence, we took a look at the proportions of NP versus non-NP complements after copula *be* at a lower level, viz., A2. In the Chinese learner data, the proportion of NP complements decreased as their proficiency increased, from 31.7 % at A2 to 28.3 % at B1. In contrast, the other two learner corpora showed the opposite trend, with higher-level learners using NP complement more frequently. These results suggest that the Chinese learners' heavier reliance on NP complement after

Table 7: Frequencies of NP and non-NP complements for copula *be*.

Corpus	NP	Non-NP	Total
Chinese-B1	442 (28.3 %)	1,118 (71.7 %)	1,560
Japanese-B1	149 (19.9 %)	599 (80.1 %)	748
Korean-B1	124 (21.5 %)	453 (78.5 %)	577

copula *be*, compared to the Japanese and Korean learners, may be attributable to L1 transfer; however, this pattern can gradually disappear as they become more proficient in L2 English.

5 Discussion

The current study explored whether and to what extent L1-topic-prominence affects the use of English *be* verb in interlanguage, yielding three main findings. First, we found significant difference in the proportions of auxiliary *be* and copula *be* between the Native English speakers and Filipino-B1 corpora, on the one hand, and the Chinese-B1, Japanese-B1, and Korean-B1 corpora, on the other hand. The L2-English learners with topic-prominent L1s used copula *be* in a significantly higher proportion than the native speakers and the learners with non-topic-prominent L1. It is also found that the use of copula *be* was similar within the topic-prominent L1 groups. Given that copula *be* often serves as a topic marker for its position and function (Mikkelsen 2005), these findings suggest the topic prominence in different L1s indeed invariably play a role in the use of copula *be*. That is, the learners with topic-prominent L1s appear to have a greater reliance on the use of copula *be*.

Second, we found that learners from topic-prominent L1s exhibit more topic-marking features in their use of *be* verb compared to those from non-topic-prominent L1s. The former group more frequently used copula *be* with the-NP subjects rather than *a(n)*-NP subjects and produced more nontargetlike *be* verbs. We hypothesized that if copula *be* links topics and comments, the subject NP would be more likely to be hearer-old, leading to more frequent use of *the*-NP subjects. Indeed, Chinese, Japanese, and Korean learners showed this pattern, with *the*-NP subjects appearing 77.0 % (Chinese: 94 vs. 28 tokens) to 88.7 % (Japanese: 63 vs. 8 tokens) of the time before copula *be*. In sharp contrast, Filipino learners reversed this pattern, using *a(n)*-NP subjects more frequently (57.3 %: 51 vs. 38 tokens). This supports the idea that learners with topic-prominent L1s use copula *be* with greater topic-marking properties, as *the*-NP subjects typically convey hearer-old information.

Additionally, learners from topic-prominent L1s produced more nontargetlike *be* structures, such as *be*-insertion, with frequencies at least twice those of non-topic-prominent L1 learners. These learners also exclusively used three nontargetlike patterns where *be* verbs appear to function as topic markers. For example, in the “*be* + modal verb” pattern, *be* sets the speech topic, followed by the speaker’s epistemological stance conveyed through the modal verb. The higher frequency of these nontargetlike *be* structures among learners from topic-prominent L1s supports the claim that L1 topic-prominence features can lead to more nontargetlike uses of *be* verb.

Finally, the study found intra-topic-prominent L1 variation within the topic-prominent L1 groups. L1-Chinese learners of English used copula *be* more frequently before NP-complements than the other topic-prominent L1 learner groups. We suggest that Chinese learners' heavy reliance on NP complement in the use of copula *be* may be ascribed to L1 transfer. That is, due to the transfer of L1-Chinese in which AP and PP complements are not allowed for the Chinese copula *shì*, L1-Chinese learners who may use copula *be* as a substitute for *shì* could use NP complements more frequently after copula *be*.

The three main findings contribute methodologically and theoretically to the study of linguistic transfer in L2 acquisition. Previous studies that claimed the use of *be* as a topic marker relied heavily on 'performance similarities between L1 and interlanguage'. However, such performance similarities do not suffice to conclude L1 transfer but require additional data, viz., 'inter-L1-group differences' and 'intra-L1-group similarities' (Jarvis 2000). In addition, it is important to control for L2 proficiency when learners are compared to investigate the effect of a particular factor on L2 acquisition (Schwartz and Sprouse 2000). In this regard, the current study made a methodological advancement, focusing on 'intra-topic-prominent-L1-group similarities' and 'inter-topic-prominence-contrastive-L1-group differences' to provide more convincing evidence for collective topic-prominence L1 effects on the use of *be*.

6 Conclusions and pedagogical implications

This study confirms the consistent impact of topic prominence from multiple L1s on L2 English *be* usage, demonstrating that learners from topic-prominent L1s more frequently use *be* as a topic marker. While the study offers valuable insights, it has some limitations, such as a focus on only four L1 groups, including only one non-topic-prominent L1 group, and primarily B1 level learners, highlighting the need for broader investigations. Moreover, the analysis was conducted at the corpus level rather than at the individual text level, which limits our ability to provide detailed variance data. Additionally, the analysis of (in)definite subject noun phrases was limited to the contrast between *a(n)* and *the* preceding singular nouns, with the definiteness of plural nouns underrepresented. Despite these constraints, this study lays the groundwork for future research on linguistic transfer, emphasizing the importance of considering linguistic similarities and differences in a broader context.

Pedagogically, it is essential for L2 teachers to understand the variations between learners' L1 and the target L2 and address predictable challenges via curriculum design and constructive scaffolding. Specifically, the overuse and/or nontargetlike use of *be* verb among L2 English learners with topic-prominent L1s

may be avoided via guided practice, such as describing various people’s actions with the auxiliary *be* (e.g., *A boy is singing*), and corrective feedback, such as providing more targetlike sentences for learners’ nontargetlike *be*-overuse constructions.

Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: The authors state no conflict of interest.
Research funding: None declared.
Data availability: The raw data can be obtained on request from the corresponding author.

Appendix: Frequencies of four types of nontarget-like *be* in the four learner corpora

Types of nontargetlike <i>be</i>	Filipino	Chinese	Japanese	Korean	Sum
<i>be</i> + uninflected verb	10	29	27	13	79
<i>be</i> + inflected verb	1	7	3	2	13
<i>be</i> + modal verb		1	1	1	3
<i>be</i> + commentary clause			1	4	5
Total	11	37	32	20	100

References

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor, MI: Karoma Publishers.

Brown, Roger. 1973. *A first language*. Cambridge, MA: Harvard University Press.

Cadierno, Teresa. 2010. Motion in Danish as a second language: Does the learner’s L1 make difference? In ZhaoHong Han & Teresa Cadierno (eds.), *Linguistic relativity in second language acquisition: Thinking for speaking*, 1–33. Bristol, UK: Multilingual Matters.

Chan, Alice Y. W. 2004. Syntactic transfer: Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal* 88(1). 56–74.

Council of Europe. 2020. *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Strasbourg: Council of Europe Publishing. Available at: www.coe.int/lang-cefr.

Den Dikken, Marcel. 2005. A comment on the topic of topic-comment. *Lingua* 115(5). 691–710.

Dulay, Heidi C. & Marina K. Burt. 1974. Errors and strategies in child second language acquisition. *Tesol Quarterly* 8(2). 129–136.

- Finlay, Catherine. 1988. *Syntactic variation in the speech of Belfast English schoolchildren*. Belfast, Northern Ireland: University of Ulster dissertation.
- Foley, Claire & Suzanne Flynn. 2013. The role of the native language. In J. Herschensohn & M. Young-Scholten (eds.), *The Cambridge handbook of second language acquisition*, 97–113. New York: Cambridge University Press.
- Gundel, Jeanette K. 1974. *The role of topic and comment in linguistic theory*. Austin, TX: The University of Texas at Austin.
- Gundel, Jeanette K. 1988. Universals of topic-prominent structure. *Studies in Syntactic Typology* 17(1). 209–239.
- Hahn, Hye-ryeong. 2000. *UG availability to Korean EFL learners: A longitudinal study of different age groups*. South Korea: Seoul National University Unpublished doctoral dissertation.
- Halliday, Michael. 2014. *An introduction to functional grammar*, 4th edn. New York: Routledge.
- Hawkins, John. 2015. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Routledge.
- Hsieh, Fu-Tsai. 2009. The acquisition of English agreement/tense morphology and copula be by L1-Chinese-speaking learners. *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching* 3. 45–59.
- Huebner, Thom. 1983. *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma Publishers.
- Ionin, Tania & Kenneth Wexler. 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.
- Ishikawa, Shin’ichiro. 2014. Design of the ICNALE-spoken: A new database for multi-modal contrastive interlanguage analysis. *Learner Corpus Studies in Asia and the World* 2. 63–76.
- Ishikawa, Shin’ichiro. 2019. The ICNALE spoken dialogue: A new dataset for the study of Asian learners’ performance in L2 English interviews. *English teaching* 74(4). 153–177.
- Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in them interlanguage lexicon. *Language Learning* 50(2). 245–309.
- Jeon, Jeeyoung & Min-Chang Sung. 2023. L2 English learners’ verb lexicalization of motion events: Effects of proficiency and salience of manner. *Review of Cognitive Linguistics* 21(1). 266–292.
- Kim, Kitaek. 2011. Overgenerated *be* from topic marker to verbal inflection. *Foreign Language Education Research* 14. 1–22.
- Lardiere, Donna. 2008. Feature-assembly in second language acquisition. In Juana Liceras, Helmut Zobl & Helen Goodluck (eds.), *The role of formal features in second language acquisition*, 106–140. Mahwah, NJ: Lawrence Erlbaum.
- Lee, Chungmin. 2006. Contrastive topic/focus and polarity in discourse. In K. von Stechow & K. Turner (eds.), *Where semantics meets pragmatics*, 381–420. Amsterdam: Elsevier.
- Lee, Nancy & 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. *Regional Language Centre Journal* 35(2). 211–228.
- Li, Charles N. & Sandra A. Thompson. 1976. Subject and topic: A new typology of languages. In Charles N. Li (ed.), *Subject and topic*, 458–489. New York: Academic Press.
- Mikkelsen, Line. 2005. *Copular clauses: Specification, predication and equation*, Vol. 85. Amsterdam: John Benjamins.
- Montrul, Silvina. 2001. First-language-constrained variability in the second-language acquisition of argument-structure-changing morphology with causative verbs. *Second Language Research* 17. 144–194.
- Nam, Bora. 2020. *Be-insertion in interlanguage: A topic marker, a tense/agreement morpheme, or both?* Bloomington: Indiana University Unpublished doctoral dissertation.

- Nation, I. S. P. & David Beglar. 2007. A vocabulary size test. *The Language Teacher* 31(7). 9–13.
- O'Neill, Teresa. 2015. *The domain of finiteness: Anchoring without tense in copular amalgam sentences*. New York: City University of New York dissertation.
- Ortega, Lourdes. 2012. Language teaching and SLA: Understanding the limits and possibilities of the research-teaching interface. *Language Teaching Research* 16. 206–226.
- Partee, Barbara. 1986. Ambiguous pseudoclefts with unambiguous be. In *Proceedings of NELS*, Vol. 16, 354–366. Amherst: GLSA, University of Massachusetts.
- Partee, Barbara. 2010. Specificational copular sentence in Russian and English. *Oslo Studies in Language* 2(1). 25–49.
- Paul, Waltraud & John Whitman. 2008. *Shi...de* focus clefts in Mandarin Chinese. *Linguistic Review* 25. 413–451.
- Prince, Ellen F. 1981. Towards a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–155. New York: Academic Press.
- Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse description: Diverse analyses of a fund raising text*, 295–325. Amsterdam: Benjamins.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53–94.
- Roberts, Craige. 2019. 11 Topics. In Klaus von Heusinger (ed.), *Semantics-sentence and information structure*, 381–412. Berlin: De Gruyter.
- Schwartz, Bonnie D. & Rex A. Sprouse. 2000. When syntactic theories evolve: Consequences for L2 acquisition research. In John Archibald (ed.), *Second language acquisition and linguistic theory*, 156–186. Oxford: Blackwell.
- Shin, Jung-Sun. 2000. *Functional category acquisition by Korean EFL learners: The role of UG in foreign language learning*. South Korea: Seoul National University Unpublished doctoral dissertation.
- Stam, Gale. 2006. Thinking for speaking about motion: L1 and L2 speech and gesture. *International Review of Applied Linguistics in Language Teaching* 44(2). 145–171.
- Stowell, Timothy Angus. 1981. *Origins of phrase structure*. Cambridge: Massachusetts Institute of Technology dissertation.
- Talmy, Givón. 1983. Topic continuity in discourse: The functional domain of switched reference. In Pamela Munro & John Haiman (eds.), *Switch reference and universal grammar*, 51–82. Amsterdam: Benjamins.
- Taylor, Ann, Mitchell Marcus & Beatrice Santorini. 2003. The Penn treebank: An overview. In Anne Abeillé (ed.), *Treebanks*, 5–22. Dordrecht: Springer.
- Tham, Shiao Wei. 2008. The semantic category of the subject NP in Mandarin specificational copular sentences. *Journal of East Asian Linguistics* 17. 61–82.
- Zhan, Fangqiong. 2012. *The structure and function of the Chinese copular construction*. Stanford: Stanford University Unpublished doctoral dissertation.
- Zobl, Helmut & Juana Liceras. 1994. Functional categories and acquisition orders. *Language Learning* 44(1). 159–180.