First- and Second-Person Pronoun Use by East Asian EFL Learners: Corpus-Based Study of Argumentative Essays

Seonwoong Kim & Sun-Young Oh[†]
Seoul National University

ABSTRACT

Based on the ICNALE corpus, this study investigates the use of first- and second-person pronouns in argumentative essays by Chinese, Korean, and Japanese leaners of English as a foreign language (EFL) across four proficiency levels. The findings indicate that higher-proficiency learners use these pronouns less frequently, which aligns with prior studies on Western EFL learners. Additionally, Korean and Japanese learners show stronger preference for *I* over *we*, whereas Chinese learners exhibit the opposite pattern. These differences may stem from various factors, such as cultural influences or first-language (L1) rhetorical conventions. Statistical analyses suggest that L1 background exerts a stronger effect on pronoun usage than proficiency level. This challenges the prescriptive English for Academic Purposes guidelines, which discourage first-person pronouns for novice writers regardless of their linguistic background. Meanwhile, pedagogical approaches that emphasize the rhetorical functions of pronouns may better equip learners in making informed lexical choices based on their communicative intent.

Keywords: personal pronouns, East Asian EFL learners, contrastive interlanguage analysis, L2 proficiency, cultural influence

1. Introduction

First- and second-person pronouns have been extensively studied in both first language (L1) and second language (L2) contexts, particularly for their role in shaping textual characteristics in speech and writing. Prior research has shown that texts with a high frequency of these pronouns tend to be perceived as more involved (Biber, 1988; Chafe, 1982), oral (Biber, 2006), informal (Smith, 1986), or even authoritative in certain academic contexts (Ivanič & Camps, 2001). Tang and John (1999) further classified first-person pronouns based on varying degrees of authorial

[†] Corresponding author: sunoh@snu.ac.kr

presence, proposing a typology of six distinct *I*s. Among these, *I* as the originator and *I* as the opinion-holder exhibit the strongest writer presence by "formulat[ing] thoughts and ideas in writing" (p. 31), whereas weaker forms primarily guide readers or simply represent the author. This classification implies that writers can strategically adjust the level of authority they wish to project (Hyland, 2002) by carefully considering the function and frequency of different uses of *I* in their writing.

Drawing on the framework of Contrastive Interlanguage Analysis (CIA) (Granger, 1996), corpus linguistics studies have consistently found that L2 English learners over-represent *I*, *we*, and *you* in academic writing, including argumentative essays, compared to native speakers (NS) (Chang, 2015; Lee, 2008; Lee & Oh, 2018; McCrostie, 2008; Natsukari, 2012; Paquot et al., 2013; Petch-Tyson, 1998; Recski, 2004). While some studies have documented reduced pronoun usage in non-native speakers' (NNS) essays, these cases typically involve highly proficient writers–such as undergraduate students in an ESL (English as a Second Language) environment (Hong Kong) in Hyland (2002) or professional researchers in Martinez (2005)–who understand the rhetorical impact of *I* and *we* in academic registers. Consequently, such findings may not be representative of the broader EFL population.

Many EFL learners, particularly those at lower proficiency levels, struggle with register variation and the appropriate use of linguistic features across different registers. While their pronoun usage is shaped by both L1/cultural background and proficiency (Aguado-Jiménez et al., 2012; Kobayashi & Abe, 2016), effective deployment requires register awareness and familiarity with academic discourse norms (Luzón, 2009). To provide a more comprehensive understanding of non-native pronoun use, this study examines argumentative essays written by EFL learners from three East Asian countries–China, Japan, and Korea–across four proficiency levels. Specifically, it investigates how L1 background and L2 proficiency affect the use of *I*, *we*, *you*, and their variants.[1)] This study aims to address the following research questions.

(1) Is there a systematic change in the frequency of *I*, *we*, and *you*, and their variants across four proficiency levels in East Asian EFL learners' argumentative essays?
(2) Do Korean, Japanese, and Chinese EFL learners demonstrate distinct

1) The variants are *I*, *my*, *me*, and *mine*, for *I*; *we*, *our*, *us*, and *ours* for *we*; and *you*, *your*, and *yours* for *you*. For conciseness, first-person singular, first-person plural, and second-person pronouns, along with their respective variants, are collectively referred to as *I*, *we*, and *you* throughout this study.

preferences in their use of *I*, *we*, and *you* in argumentative writing?

(3) What are the most frequent collocates of *I* and *we* among the three East Asian EFL learner groups, and how do they shape rhetorical structures?

2. Literature Review

2.1. Diverse functions of first- and second-person pronouns

A large body of L1 research has attempted to distinguish between speech and writing by identifying linguistic features that differentiate these two modes. For instance, Chafe (1982) argued that spoken discourse involves greater interaction with audiences than written discourse, proposing that the dichotomous poles of *involvement* and *detachment* correspond to typical oral and written genres, respectively. According to Chafe, involvement in English is marked by features such as self-references (*I*, *me*, *my*), pragmatic markers, fuzzy references, and direct quotes, inter alia, all of which indicate the speaker's prioritization of experiential engagement over consistency. In other words, owing to the frequent use of these features including first-person pronouns, "speaking is more likely to be about personal experiences while writing typically conveys more general descriptive and explanatory information" (Redeker, 1984, p. 44).

Adopting a similar yet distinct perspective, Smith (1986) emphasized the role of *interpersonal features* in shaping the presence of both the writer and the reader within a text, thereby influencing its degree of formality. Specifically, the use of first- and second-person pronouns tends to create a more informal tone, whereas passive constructions and third-person subjects contribute to greater formality. Proposing a continuum of formality, Smith argued that any shift in tone should be gradual and carefully managed through strategic linguistic choices.

Biber (1988) offered a more empirical approach by using factor analysis to examine co-occurring linguistic features, thereby distinguishing his work from Chafe (1982) and Smith (1986), who relied on a priori assumptions about communicative functions. Analyzing 67 linguistic features, Biber found that first- and second-person pronouns had strong positive loadings on Dimension 1 (*Informational* vs. *Involved Production*), along with *that* deletion, contractions, present-tense verbs, and *do* as a pro-verb, among others.

These involvement-related features frequently co-occur in texts with a "primarily

interactive or affective purpose" (p. 105), rendering the discourse more conversational and less oriented toward the transmission of factual information. Given its functional impact, Dimension 1 was also labeled the "oral/literate dimension" (p. 162), alongside Dimensions 3 and 5, which collectively categorized the 23 genres of speech and writing analyzed in Biber's study.[2] Notably, genres with the highest mean dimension scores in Dimension 1 were telephone and face-to-face conversations, while official documents ranked the lowest. This reaffirms that texts with a higher frequency of first- and second-person pronouns more closely resemble spoken rather than written discourse.

In sum, previous L1 research—whether grounded in empirical assessments or theoretical assumptions—consistently frames first- and second-person pronouns as key indicators of involvement in discourse. Their frequency serves as a marker of a text's alignment along the speech-writing continuum.

2.2. Petch-Tyson's (1998) writer/reader visibility

With the expansion of corpus linguistics and the emergence of large learner corpora, CIA (Granger, 1996) has become a core framework in Learner Corpus Research (LCR). CIA has proven particularly valuable for two types of comparisons: (1) analyzing learner language in relation to NS data and (2) conducting L2-L2 comparisons to investigate linguistic features across different learner groups. The latter approach has extended research beyond L1 transfer, incorporating independent variables such as learning settings (foreign vs. second language), task effects, and proficiency levels (Granger, 2015).

As one of the earliest CIA-based studies, Petch-Tyson (1998) built upon Tannen's (1982) observation that speakers tend to overtly encode themselves, showing a greater degree of involvement than writers. Drawing on linguistic features outlined by Chafe (1982) and Smith (1986), she examined writer/reader visibility in argumentative essays, analyzing features such as first- and second-person pronouns, fuzziness words (e.g., *kind/sort of*), emphatic particles (e.g., *just, really*), and references to writing/reading situation. Using a preliminary version of the International Corpus of Learner English (ICLE), the study compared the frequencies of these features across five sub-corpora.

---

[2] Six dimensions were extracted and interpreted using Biber's (1988) multi-dimensional analysis framework. Among these, the third and fifth dimensions were labeled *Explicit vs. Situation-Dependent Reference* and *Abstract vs. Non-Abstract Information*, respectively.

The findings revealed sizeable variation among the five Western EFL learner groups (Swedish: 2,265; Finnish: 1,876; Dutch: 1,516; French: 1,447) and a clear discrepancy between NS and NNS texts (cf. U.S.: 607). Notably, all writer/reader visibility features, except emphatic particles, were more frequent in NNS than in NS essays. Petch-Tyson (1998) interpreted this as evidence that NNS exhibit greater interpersonal involvement in argumentative essays, likely due to weaker adherence to academic genre conventions. She also suggested that factors such as English education and essay prompts warrant further exploration in future L2-L2 comparative studies.

Beyond quantitatively identifying the NS-NNS disparity, Petch-Tyson (1998) also noted stylistic inconsistencies in NNS essays, particularly in conclusions where many learners shifted abruptly from informal to formal expressions. For instance, first-person pronouns *I* and *we* were often replaced with *one* in final sentences, indicating difficulties in maintaining a consistent interpersonal stance. This pattern aligns with broader challenges in pragmatic competence, as learners often struggle to manage interpersonal meta-discourse in a foreign language (Luukka, 1994; Markkanen et al., 1993).

Despite offering valuable insights into NNS use of personal pronouns, Petch-Tyson's study (1998) presents two key limitations. First, it did not account for variation across proficiency levels, as the ICLE primarily included upper-intermediate to advanced learners, limiting the analysis of developmental patterns in first- and second-person pronoun usage. Second, although cultural variation in discourse involvement is well-documented (Tannen, 1982), the study focused solely on Western EFL learners, leaving a gap in research on non-Western L2 learners. This study seeks to address these gaps by examining variation in first- and second-person pronoun use across both proficiency levels and East Asian L1 backgrounds.

2.3. The use of first- and second-person pronouns across proficiency levels

Building on Petch-Tyson's (1998) pioneering study of first- and second-person pronoun use among L2 English learners, subsequent research has sought to trace developmental patterns across proficiency levels. Much of this work has focused on the over-representation of these pronouns in NNS writing, often under the assumption that their frequency declines as proficiency increases. This expectation aligns with consistent findings in LCR studies, encapsulated by Gilquin and Paquot's (2008) assertion that "a linguistic form is found significantly more or less in the learner corpus than in the reference corpus" (p. 38).

One of the first studies to examine proficiency-based variation in personal pronoun

use was McCrostie (2008), who compared data from Petch-Tyson (1998) with Japanese first-year university students' essays. He found that these students used writer/reader visibility features approximately four times more frequently than native speakers in argumentative essays, but this figure was nearly halved in second-year students' essays–a decline he described as "a dramatic drop", bringing their usage closer to NS levels. Further analysis of concordance lines revealed that both first- and second-year Japanese students not only diverged from NS norms but also from their Western counterparts in Petch-Tyson (1998) by excessively using *I think* in clause-initial positions, a trend also noted by Natsukari (2012). According to McCrostie, this pattern reflects Japanese EFL learners' limited familiarity with the connotations of clause-initial *I think* in academic writing, where it conveys "a confident and expert mind in full control of the material, making judgments and passing comment on issues of concern to the discipline" (Hyland, 2000, p. 123). Moreover, the frequent use of this phrase runs counter to academic writing conventions, which often caution undergraduate students against assuming the same level of authority as established scholars (Chang & Swales, 1999).

McCrostie (2008) attributed the decline in writer/reader visibility features from first- to second-year students to the effectiveness of academic writing instruction, particularly the explicit guidance against first-person pronoun use. However, as his study did not provide details on the students' curricula, insight into the instructional context remains limited. Furthermore, grade-based comparisons do not necessarily capture proficiency-related changes, as learner proficiency was not externally assessed. In this sense, McCrostie's findings may instead reflect the strong influence of prescriptive norms that discourage overt authorial presence in academic writing. However, as Tang and John (1999) caution, strict adherence to such prescriptions may reinforce "some vague preconceived notion that academic writing should be distant and impersonal" (p. 35).

Unlike McCrostie (2008), Choung and Oh (2017) and Lee and Oh (2018) classified learners by L2 proficiency using holistic grading rather than grade-based comparisons. Drawing on the Korean Learner Corpus of English Argumentative Essays (KLCEAE) (Oh & Kang, 2013) and the Yonsei English Learner Corpus (YELC 2011) (Rhee & Jung, 2014), respectively, they compared Korean EFL learners' use of *I* with NS data from the Louvain Corpus of Native English Essays (LOCNESS) (Granger, 1998b). Their findings revealed a consistent trend: as L2 proficiency increased, the use of *I* in argumentative writing decreased, gradually aligning with NS norms. Additionally, advanced writers demonstrated greater register awareness by using *I* in

conjunction with epistemic devices more appropriately (Lee & Oh, 2018).

However, an earlier study by Ryoo (2010) found no significant difference in first-person pronoun use between beginning- and intermediate-level Korean EFL learners' opinion essays. Instead, her study revealed a notable shift in pronoun preference: lower-level learners preferred first-person plural pronouns (*we* and *our*), whereas higher-level learners favored first-person singular pronouns (*I* and *my*). Ryoo attributed this pattern to L1 transfer, as Koreans commonly use the first-person plural possessive pronoun *uri* ("our") in contexts where English requires the singular possessive *my* (*nae* in Korean). For instance, even in monogamous contexts, Korean speakers may say either *nae wife* (my wife) or *uri wife* (our wife) (Na & Choi, 2009). Ryoo further linked this tendency to Korea's collectivist orientation, tentatively suggesting that such cultural influences may contribute to lower-level Korean EFL learners' reliance on *we* for self-representation.

While Ryoo's (2010) cultural interpretation offers an intriguing perspective, her study is limited by a small sample size (N = 28) and the absence of an NS reference corpus. Without an NS benchmark, it remains unclear whether Korean EFL learners exhibit a general preference for *we* compared to NS writers. To address these limitations and reconcile the conflicting findings regarding the relationship between L2 proficiency and first- and second-person pronoun use, the present study analyzes the extensive ICNALE (International Corpus Network of Asian Learners of English) dataset (Ishikawa, 2013). This corpus includes argumentative essays of English learners classified into four CEFR (Common European Framework of Reference) bands (A2, B1.1, B1.2, and B2+), allowing for L2-L2 comparisons across different proficiency levels.

2.4. Cultural influence on the use of first- and second-person pronouns

As Ryoo (2010) noted, a writer's cultural background may influence the extent to which authorial presence is explicitly expressed in written discourse. This perspective aligns with Petch-Tyson's (1998) earlier comparison of EFL learners from four distinct L1 backgrounds, which suggested that cultural factors could contribute to differences in the use of writer/reader visibility features. Her study found considerable variation, with Swedish, Finnish, Dutch, and French EFL learners ranked from highest to lowest in their use of these features. This pattern may point to culturally influenced tendencies in the reliance on first- and second-person pronouns.

Several CIA-based studies have supported this pattern. Paquot et al. (2013) compared Norwegian, French, and NS writing across two genres–argumentative essays (ICLE & LOCNESS) and discipline-specific academic prose (Varieties of English for Specific Purposes dAtabase [VESPA] & the British Academic Written English [BAWE] corpus). Their results showed a gradual decrease in first- and second-person pronoun use from Norwegian to French to NS writers. The Norwegian preference for self-referential pronouns was linked to "differing notions of individuality across cultures" (Ramanathan & Atkinson, 1999, p. 55), with Scandinavian writers favoring a more interactive and less formal writing style.

Further evidence of Western learners' reliance on personal pronouns comes from Ringbom (1998), who found that NNS from various Western L1 backgrounds tended to over-represent first- and second-person pronouns compared to NS writers. However, it is important to note that this case study published in *Learner English on Computer* (Granger, 1998a) was based on the earliest version of the ICLE, which contained data exclusively from Western EFL learners.[3] As a result, the study offered limited scope for cross-linguistic comparisons.

Although Petch-Tyson (1998), Ringbom (1998), and Paquot et al. (2013) were pioneers in investigating potential cultural influences on personal pronoun use, their studies focused exclusively on EFL learners from European countries, largely due to the composition of the ICLE at the time. The first version (2002) incorporated data from 11 Western L1s, while the second version (2009) expanded to 14 Western L1s and only two Eastern L1s (Chinese and Japanese). As a result, previous L2 research on personal pronoun use has disproportionately centered on Western EFL learners. Recski (2004) exemplifies this trend; his study, based on six ICLE sub-corpora, analyzed essays from Czech, Brazilian, French, Spanish, Dutch, and Finnish learners.

This imbalance in data representation highlights the need for further research on EFL learners from typologically and culturally diverse L1 backgrounds, particularly in East Asia. Given that Chinese, Korean, and Japanese students collectively account for 33% of all international students in U.S. colleges and universities (Institute of International Education, 2023), addressing this gap is crucial to better understanding their academic writing practices and supporting their development in EAP programs.

A notable study extending beyond Western NNS is Luan and Zhang (2016), who compared argumentative essays from Chinese and Swedish EFL learners using the

[3] The latest ICLE version (Granger et al., 2020) has since addressed this limitation by incorporating datasets from Chinese, Korean, and Japanese EFL learners, responding to calls for greater inclusion of typologically distant languages (Tracy-Ventura, 2020).

Swedish-Chinese English Learner Corpus (SCELC). They found that Swedish learners used *I* and *we* more frequently than their Chinese counterparts. Rather than focusing solely on frequency, however, they examined the proportion of each pronoun, revealing a narrower gap for *we* (Swedish: 14.92 vs. Chinese: 10.09) than for *I* (Swedish: 11.64 vs. Chinese: 4.58). Luan and Zhang (2016) associated Chinese learners' stronger reliance on *we* with their traditional collectivist values, suggesting a culturally shaped preference for collective self-representation. This collective orientation has also been observed in studies on Chinese politeness and modesty (Chen, 1993, 2001; Gu, 1990).

Previous research has also identified a link between cultural orientation and pronoun usage (Hofstede, 2001, 2011; Kashima & Kashima, 1998; Na & Choi, 2009; Uz, 2014). These studies indicate that the use of *I* is encouraged in individualist societies, whereas languages spoken in collectivist cultures often permit or even prescribe the omission of this pronoun. Indeed, English, as a representative language of individualist cultures, uniquely capitalizes *I* to emphasize the independent self (Hofstede et al., 2010), while all three East Asian countries analyzed in this study–Korea, Japan, and China–commonly speak pro-drop languages, which are associated with collectivist cultures.

However, studies on the relationship between cultural orientation and pronoun usage have predominantly focused on L1 data, leaving its impact on L2 production relatively unexplored. Additionally, while studies have compared personal pronoun use among L2 learners with various Western L1s, to the best of our knowledge, no research has systematically examined variations across multiple Asian L1 backgrounds. As Hyland (2002) suggests, students from collectivist cultures may face challenges in adopting an individualistic authorial stance, as their cultural background tends to emphasize a more socially embedded view of the self. This underscores the importance of understanding how writers from collectivist cultures construct their authorial identity when writing in English, a language associated with individualist cultural norms.

To address these issues, the present study examines first- and second-person pronoun use among EFL learners from three East Asian countries (Korea, Japan, and China), which share geographical proximity and linguistic similarities rooted in a shared history of Chinese character influence (Song, 2006). The primary goal is to determine whether preferences for personal pronoun use vary across these learner groups, particularly in the choice between *I* and *we*.

3. Methodology

3.1. Corpus

This study draws on the International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2013), specifically its written essays module. As one of the largest publicly accessible English learner corpora, the ICNALE contains approximately 1.3 million words produced by 2,600 college students from 11 Asian countries and regions, alongside 300 native speakers. Ishikawa (2013) developed the ICNALE to address the historically limited representation of Asian English learners in corpus-based research. By including diverse Asian L1s, the ICNALE enables comparative analyses that were previously unavailable, helping to bridge longstanding gaps in learner corpus studies.

A defining feature of the ICNALE is its rigorous control over writing conditions. To minimize external influences on linguistic choices, all students were assigned two essay topics and given 40 minutes to complete each. The assigned topics were:

1) "It is important for college students to have a part-time job."
2) "Smoking should be completely banned at all the restaurants in the country."

(Ishikawa, 2013)

Beyond ensuring topical consistency, the ICNALE classifies learners into four proficiency bands (A2, B1.1, B1.2, B2+) based on standardized English test scores (TOEIC, TOEFL, IELTS) or the Vocabulary Size Test (VST; Nation & Beglar, 2007). Ishikawa applied a linear regression model to establish a conversion formula between VST and English test scores, validating the reliability of the proficiency classification. These methodological controls enhance comparability across sub-corpora, making the ICNALE a valuable resource for investigating L1- and proficiency-related variation in English usage. Table 1 provides a breakdown of the ICNALE dataset, encompassing all essays written on the two assigned topics.

Table 1. Number of tokens in the ICNALE

L1s	Proficiency levels				
	A2	B1.1	B1.2	B2	Total
EFL					
Korea[4)]	29,291	26,910	40,073	36,194	132,468
Japan	68,529	79,591	22,390	8,532	179,042
China	22,520	110,293	52,031	6,561	191,405
Indonesia	14,754	37,465	39,517	N/A	91,736
Thailand	53,506	80,758	45,840	N/A	180,104
Taiwan	12,878	40,008	28,484	11,014	92,384
ESL					
Hong Kong	N/A	14,467	24,584	7,934	46,985
Pakistan	7,929	43,542	41,533	N/A	93,004
The Philippines	N/A	5,280	87,185	6,032	98,497
Singapore	N/A	N/A	65,782	33,485	99,267
The UAE	N/A	N/A	18,629	28,351	46,980
NS[5)]					
US, UK, etc.			N/A		90,613
Total	209,407	438,314	466,048	138,103	1,342,468

Note. Sub-corpora containing fewer than ten essays are marked as N/A.

3.2. Procedures for analysis

This study systematically examines the use of first- and second-person pronouns among East Asian EFL learners by analyzing variation across proficiency levels and

4) Sixteen essays from eight Korean A2 writers were excluded from the analysis due to a high degree of phrase overlap indicating duplication.

5) Given the recognition of multiple English dialects as standards in ESL/EFL education, the ICNALE ensures a balanced representation of native speakers by including writers from the United States, the United Kingdom, Canada, Australia, and New Zealand (Ishikawa, 2013).

L1 backgrounds. The analysis is carried out in three stages, employing frequency-based, statistical, and qualitative methods, including close inspection of learner essays.

The first stage explores how personal pronoun usage varies across four CEFR levels (A2, B1.1, B1.2, B2+) in the ICNALE sub-corpora of Korean, Japanese, and Chinese EFL learners. Frequency counts for *I*, *we*, *you*, and their variants were extracted using AntConc 4.3.1 (Anthony, 2024). Since each participant wrote two essays on assigned topics, both were included to maintain a balanced 50:50 topic distribution. Frequency counts were normalized per 1,000 tokens to allow meaningful comparisons.

Since the data did not follow a normal distribution (Shapiro-Wilk test, $p < .001$), three Scheirer-Ray-Hare tests–the nonparametric equivalent of a two-way analysis of variance–were conducted to examine the effects of L2 proficiency and L1 background on pronoun usage. Dunn's test with Bonferroni correction was used for post hoc multiple comparisons, and effect sizes were measured using epsilon squared (ε^2). Statistical analyses were performed in R (RStudio, 2025) on 1,092 essays from the Korean, Japanese, and Chinese ICNALE sub-corpora. To reduce the impact of topic variation on language use (Hinkel, 2002, 2009; Yang et al., 2015; Yoon, 2017, 2021), and to meet the assumption of independent observations, only essays on the topic of part-time jobs were included.

The final stage investigates how first-person pronoun choice (*I* vs. *we*) influences rhetorical strategies, with a particular focus on Japanese and Chinese EFL learners, who demonstrated the highest reliance on *I* and *we*, respectively. A qualitative analysis examines how pronoun preferences shape sentence construction, argument structure, and persuasive techniques. To explore these patterns, a collocational analysis was conducted to identify the most frequently occurring words following *I* and *we* across the Korean, Japanese, Chinese, and NS sub-corpora. This stage serves to bridge quantitative findings with qualitative interpretations, providing deeper insight into the rhetorical functions of personal pronouns in EFL argumentative writing.

4. Results and Discussion

4.1. Variation in the use of *I*, *we*, and *you* across proficiency levels

Table 2 presents the normalized frequency counts of *I*, *we*, and *you* in argumentative essays written by Korean, Japanese, and Chinese EFL learners at four profi-

ciency levels, alongside an NS reference group.

Table 2. Frequency of *I*, *we*, and *you* in ICNALE-Korea, Japan, China, and NS

L1 background & L2 proficiency	*I*	*we*	*you*	Sum
Korea				
A2	23.73	6.96	10.41	41.10
B1.1	30.10	8.21	11.82	50.13
B1.2	21.26	8.33	9.46	39.05
B2+	16.99	6.55	5.94	29.48
Average	22.44	7.52	9.19	39.14
Japan				
A2	31.72	19.26	5.65	56.63
B1.1	32.08	18.72	5.80	56.60
B1.2	31.93	12.68	5.76	50.38
B2+	25.55	16.53	4.22	46.30
Average	31.61	18.07	5.66	55.34
China				
A2	12.88	21.71	8.66	43.25
B1.1	12.58	25.63	9.98	48.20
B1.2	11.34	22.95	9.05	43.34
B2+	8.99	18.44	4.27	31.70
Average	12.16	24.19	9.38	45.73
NS	27.20	5.11	5.36	37.68
Average	22.03	14.70	7.41	44.14

Note. All values represent normalized frequencies per 1,000 tokens.

Overall, pronoun usage trends align with previous studies, suggesting that higher-proficiency learners tend to be less overtly present in their writing, using fewer first- and second-person pronouns (Choung & Oh, 2017; Lee & Oh, 2018; McCrostie, 2008). However, the ICNALE's inclusion of four distinct proficiency levels offers a more detailed view of developmental trajectories. While pronoun counts generally declined at higher proficiency levels, an initial increase was observed between A2 and B1.1, potentially indicating a tentative U-shaped pattern in the earliest stages of acquisition. This trend appeared in seven out of nine cases, except for *we* among Japanese learners and *I* among Chinese learners. Although minor fluctuations were also found at higher levels, they were relatively infrequent, occurring in only 2 out of 18 instances–a slight increase in *we* between Korean B1.1 and B1.2 and between Japanese B1.2 and B2+.

To determine the statistical significance of these proficiency-related trends and to assess the influence of L1 background, three Scheirer-Ray-Hare tests were conducted. *I*, *we*, and *you* were treated as dependent variables, with L1 and proficiency serving as independent variables. The dataset included 1,092 essays from Korean, Chinese, and Japanese learners, all written on the topic of part-time jobs.

Table 3. Scheirer-Ray-Hare test results and effect size for *I*

Source	*df*	Sum of squares	*H*	*p*	ε^2
L1	2	24,099,941	242.355	.00000***	.222
Proficiency	3	1,739,603	17.494	.00056**	.016
L1 * Proficiency	6	458,330	4.609	.59483	.0042
Residuals	1,080	79,973,625			

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

Table 4. Scheirer-Ray-Hare test results and effect size for *we*

Source	*df*	Sum of squares	*H*	*p*	ε^2
L1	2	10,648,782	111.734	.00000***	.102
Proficiency	3	489,352	5.135	.16220	.0047
L1 * Proficiency	6	504,957	5.298	.50616	.0048
Residuals	1,080	88,602,797			

***$p < .0001$.

Table 5. Scheirer-Ray-Hare test results and effect size for *you*

Source	*df*	Sum of squares	*H*	*p*	ε^2
L1	2	5,555,112	71.311	.00000***	.0654
Proficiency	3	1,377,665	17.685	.00051**	.0162
L1 * Proficiency	6	942,116	12.094	.05991	.0111
Residuals	1,080	77,797,449			

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

The results in Tables 3-5 indicate that L1 background has a more pronounced impact on pronoun usage than L2 proficiency. Specifically, L1 exerted a significant influence on all three pronoun types, with the strongest effect observed for *I* (ε^2 = .222, p < .001), followed by a medium effect for *we* (ε^2 = .102, p < .001), and a small-to-medium effect for *you* (ε^2 = .0654, p < .001). These findings suggest that L1 contributes to shaping how L2 learners express themselves, particularly in their use of first-person pronouns.

While L2 proficiency had a statistically significant effect on *I* and *you*, its influence was comparatively weaker. The effect sizes were small for both *I* (ε^2 = .016, p = .00056) and *you* (ε^2 = .0162, p = .00051), indicating only modest shifts in usage with increasing proficiency. Meanwhile, proficiency did not significantly affect the use of *we* (ε^2 = .0047, p = .162), suggesting that *we* remains relatively stable across proficiency levels.

Furthermore, the absence of significant interaction effects across all three pronouns indicates that L1 influences pronoun usage consistently, regardless of proficiency level. While L2 proficiency contributes to some variation in pronoun use, its overall impact appears relatively minor compared to the more robust and persistent influence of L1.

Table 6. Post hoc results: Effects of proficiency on pronoun usage

Pronoun	Comparison	*Z*	Adjusted *p*-value	Significance
I	A2-B1.1	4.18	.0002	***
	A2-B1.2	4.96	.0000	***
	A2-B2	5.39	.0000	***
	B1.1-B1.2	1.51	.7856	
	B1.1-B2	2.77	.0339	*
	B1.2-B2	1.52	.7660	

Table 6. Continued

Pronoun	Comparison	*Z*	Adjusted *p*-value	Significance
we	A2-B1.1	-2.96	.0186	*
	A2-B1.2	0.00	1.0000	
	A2-B2	4.08	.0003	***
	B1.1-B1.2	2.85	.0263	*
	B1.1-B2	6.46	.0000	***
	B1.2-B2	4.02	.0004	***
you	A2-B1.1	-0.03	1.0000	
	A2-B1.2	-0.71	1.0000	
	A2-B2	2.41	.0953	
	B1.1-B1.2	-0.77	1.0000	
	B1.1-B2	2.59	.0570	
	B1.2-B2	2.91	.0215	*

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 6 shows the results of Dunn's post hoc test, examining the effects of proficiency on pronoun usage. Adjusted *p*-values were calculated using the Bonferroni correction to control for multiple comparisons.

For *I*, A2 learners used it significantly more than all higher proficiency levels ($p < .001$ for B1.1, B1.2, B2). However, no significant differences were observed among the higher levels, except for a single contrast between B1.1 and B2. This result suggests that *I* usage stabilizes beyond A2, with early-stage learners exhibiting pronounced differences, but further proficiency gains having little impact on its frequency.

In contrast, the most notable differences for *we* appeared at higher proficiency levels, with B2 learners using it significantly less than all lower proficiency groups ($p < .001$). Notably, A2 learners used *we* significantly less than B1.1 learners ($p = .0186$), aligning with the B1.2 level and indicating an initial increase at the earliest stages of acquisition. However, a significant decrease was found between B1.1 and B1.2, followed by another decline at B2, providing statistical support for a U-shaped developmental trajectory in *we* usage across proficiency levels.

For *you*, no significant proficiency-related differences emerged, except for a decrease between B1.2 and B2 ($p = .0215$). This suggests that the use of *you* remains

relatively stable across proficiency levels compared to *I* and *we*, with L2 proficiency exerting only minimal influence.

4.2. Variation in the use of *I*, *we*, and *you* across L1 backgrounds

Beyond proficiency-level variations, the statistical analysis in Tables 3-5 reveals significant differences in the use of *I*, *we*, and *you* among Korean, Japanese, and Chinese EFL learners, with L1 background exerting a stronger influence on pronoun selection than proficiency level. Table 7 presents six key values related to the interquartile range, while Figure 1 visually illustrates these variations using box plots.

Table 7. Descriptive statistics of *I*, *we*, and *you* across three East Asian L1s

	Korean			Japanese			Chinese		
	I	*we*	*you*	*I*	*we*	*you*	*I*	*we*	*you*
Min.	0	0	0	0	0	0	0	0	0
Q1	8.56	0	0	16.34	0	0	4.33	4.80	0
Q2	14.60	0	0	28.47	9.43	0	9.09	25.13	3.26
Mean	22.44	7.52	9.19	31.61	18.07	5.66	12.16	24.19	9.38
Q3	34.78	9.13	13.82	51.69	46.44	0	14.77	54.13	14.48
Max.	165.0	113.1	127.1	132.3	157.0	132.4	85.6	123.6	107.5

Note. Q1, Q2, and Q3 represent the 25th, 50th (median), and 75th percentiles, respectively.

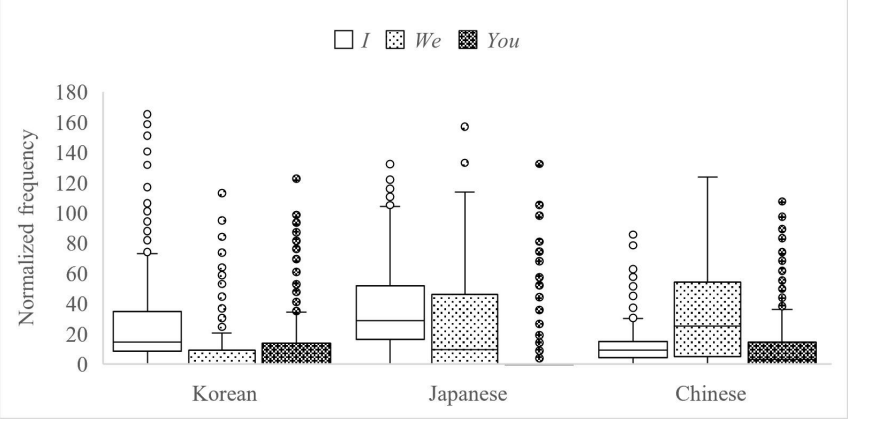

Figure 1. Variability of *I*, *we*, and *you* across three East Asian L1s[6)]

Among the three groups, Japanese learners exhibit the highest frequency of *I* usage, whereas Chinese learners show a preference for *we*. Korean learners' pronoun usage patterns fall between these two L1 groups, aligning with Japanese learners in their preference for *I* over *we* while sharing Chinese learners' relatively high use of *you*. To statistically validate these observations, Table 8 presents post hoc results from Dunn's test, assessing the impact of L1 background on pronoun usage.

Table 8. Post hoc results: Effects of L1 background on pronoun usage

Pronoun	Comparison	*Z*	Adjusted *p*-value	Significance
I	CHN-JPN	-16.27	.0000	***
	CHN-KOR	-7.51	.0000	***
	JPN-KOR	7.44	.0000	***
we	CHN-JPN	5.20	.0000	***
	CHN-KOR	12.27	.0000	***
	JPN-KOR	7.50	.0000	***
you	CHN-JPN	7.38	.0000	***
	CHN-KOR	0.87	1.0000	
	JPN-KOR	-5.91	.0000	***

***$p < .001$.

The post hoc test results confirm significant differences in pronoun usage across the three L1 groups, highlighting the influence of L1 background on pronoun preferences. Specifically, Chinese learners use *I* significantly less frequently than both Japanese and Korean learners ($p < .001$), while *we* appears far more frequently in Chinese learners' writing than in the texts of Japanese and Korean learners ($p < .001$).

Regarding the pronoun *you*, Japanese learners use it significantly less frequently than both Korean and Chinese learners ($p < .001$), while no significant difference is observed between the latter two groups. This finding suggests that Japanese learners may be more reluctant to directly address the audience with *you*, as it "separates readers as a different group and sounds detached and command-like" (Kuo, 1999, p. 136).

6) A box plot displays the interquartile range (IQR), which represents the middle 50% of the data. A line within the box marks the median, a central point of reference. Whiskers extend from the box to show data points within a typical range, usually defined as 1.5 times the IQR beyond the first and third quartiles. Data points outside this range appear as individual dots, representing potential outliers (Frigge et al., 1989).

Overall, statistically significant differences are observed among all three L1 groups in their use of personal pronouns, except for *you* between Chinese and Korean learners. Notably, Chinese learners contribute substantially to the observed variation, as their greater use of *we* and lower reliance on *I* distinguish them from Japanese and Korean learners. This suggest that removing the Chinese component from the analysis would considerably reduce the variance in *I* and *we* usage among East Asian learners.

To interpret these findings within a cultural framework, Hofstede et al.'s (2010) Individualism Index Values (IDV) offer useful context. Although all three countries are classified as collectivist societies (IDV<50), Japan has the highest individualism score (IDV = 46), setting it apart from China (IDV = 20) and Korea (IDV = 18). These scores align with the relatively high frequency of *I* among Japanese learners and the preference for *we* among Chinese learners, supporting Uz's (2014) finding that the relative frequency of *I* over *we* correlates with a country's level of individualism. This trend is further corroborated by earlier studies reporting frequent use of *I* in Japanese learners' writing (Akahori, 2007; Ishikawa, 2009; McCrostie, 2008; Natsukari, 2012) and *we* in Chinese learners' essays (Chen, 2020; Li & Lee, 2013; Luan & Zhang, 2016; Xia, 2018).

However, the IDV framework does not fully account for all findings. For instance, despite having an IDV score similar to China, Korean learners exhibit pronoun usage patterns more comparable to Japanese learners'. Moreover, as illustrated by the outliers in Figure 1, intra-group variability in pronoun choice is evident, with some learners deviating from the general tendencies of their respective L1 groups. These discrepancies suggest that cultural orientation alone cannot sufficiently explain L2 learners' pronoun choices. Additional factors, such as L1 linguistic structures, educational practices, and individual differences, may also play a significant role in shaping these patterns.

Korean and Japanese are typologically more closely related to each other than to Chinese, exhibiting significant lexical and phonological similarities (Song, 2006). Both languages also differ from Chinese in lacking a clearly defined, closed set of personal pronouns (Kashima & Kashima, 1998; Kitagawa & Lehrer, 1990), a structural feature that may shape learners' pronoun preferences. Furthermore, Korea's English education has been historically influenced by Japan's, particularly during the colonial period (1910-1945), when the grammar-translation method (GTM) was adopted and remained dominant even after liberation (Kwon, 2000). These linguistic and educational parallels may partially account for the similarities in pronoun use

between Korean and Japanese learners, distinguishing them from their Chinese counterparts.

Additionally, it is noteworthy that Chinese academic conventions discourage the use of the first-person singular pronoun *wo/wo de* in formal writing (Xia, 2018), which may further reinforce Chinese learners' tendency to eschew *I* and instead favor *we* in English essays. Supporting this, Chen (2020) observed greater use of *we* in Chinese students' master's and doctoral theses compared to undergraduate theses, indicating a departure from academic English conventions, where *I* is typically preferred (Hyland, 2001). Chen attributes this pattern to the influence of academic Chinese writing norms, as interviews revealed that writers minimized direct authorial presence to avoid appearing conceited and to maintain modesty. These findings suggest that differences in pronoun usage among the three L1 groups are likely shaped by a complex interplay of cultural, linguistic, and educational factors, rather than by a single explanatory framework.

4.3. Qualitative analysis of learner essays

This section examines how the most frequent collocates of *I* and *we* influence sentence construction and rhetorical choices among EFL learners, with particular attention to Japanese and Chinese writers. Table 9 presents the three most commonly used collocates immediately following *I* and *we* across the Korean, Japanese, Chinese, and NS sub-corpora in the ICNALE.

Table 9. Three most frequent collocates of *I* and *we* in the ICNALE

L1s	*I*	*we*
Korea	*think, agree, have*	*can, have, should*
Japan	*think, agree, have*	*can, have, should*
China	*think, have, am*	*can, should, have*
NS	*think, have, don't*	*have, can, should*

The results indicate that the frequent collocates of *I* and *we* are largely consistent across all four L1 groups, including NS writers. However, notable distinctions emerge between the two pronouns. *I* is often followed by mental verbs (*think*, *agree*) that express opinions and attitudes, while *we* is typically followed by modal auxiliaries (*can*, *should*) that convey ability or obligation (Biber et al., 1999). The verb *have*

serves dual functions: it can denote possession or appear in the semi-modal *have to*, expressing obligation.

To illustrate these patterns, two sample essays written by Japanese and Chinese EFL learners are analyzed below. Each excerpt contains the three most frequent collocates of either *I* or *we*, providing insight into differing rhetorical strategies shaped by pronoun choice. Both essays respond to the common ICNALE prompt: "It is important for college students to have a part-time job."

(1) Japanese learner essay

"**I agree** the idea that it is important for college students to have a part-time job. Certainly, some people disagree this idea and think college students should spend the time instead the time of a part-time job. But, **I have** three reasons of agreement. First, **I think** a part-time job makes my independence. I notice the independence of money by earning by myself. So I thank my parents since I started a part-time job. Second, I made friends for a part-time job. (…) Finally, I noticed the real independence of word 'Arigatou!!'. My part-time job is waiting on customers. Before I started my part-time-job, I thought I can't speak my customers well. But I could speak my customers now. And I could notice the independence of 'Arigatou'. (…)"

(ICNALE_JPN_A2_PTJ_319)

(2) Chinese learner essay

"Nowadays, as most college students have no considerable social experience, what they have learned and what they are going to learn seem to become alienated from reality. Thus, having a part-time job for college students has aroused widely attention. And **we do should** attach more importance to the part-time jobs. College students can benefit a lot from the part-time job. To begin with, **we can** earn some money in order to lighten our parents' financial load. (…) Besides, in the course of the job, **we have to** cope with the mixed people and say various words. (…) Last but not the least, **we can** adjust ourselves to the society in advance. (…) Apparently, **we can** draw the conclusion that it is important for college students to have part-time jobs."

(ICNALE_CHN_B1_2_PTJ_014)

A comparative analysis of the two excerpts reveals pronoun-driven differences in rhetorical strategies. The Japanese writer exclusively uses *I* to frame the argument

from a personal perspective, opening the essay with *I agree* and introducing the structure with *I have three reasons*. This pattern aligns with previous findings that Japanese EFL learners frequently use the phrase *I have* (Ishikawa, 2009). The two most frequent trigrams containing *I have* in the Japanese corpus are *I have two* and *I have three*, which shape a formulaic argumentative structure grounded in personal reasoning and experience. In addition, as reported in several learner corpus studies (Aijmer, 2001; Hasselgård, 2009; Paquot et al., 2013), the repeated use of *I think* further reinforces strong dependence on subjective opinion as a rhetorical strategy.

In contrast, the Chinese writer employs *we* to construct a collectivist argument, situating the discussion within a broader social context while modestly appealing to the acceptance of readers and the wider community (Kuo, 1999). The phrase *we do should* explicitly stresses obligation, and modal verbs such as *can*, *should*, and *have to* frequently appear to convey shared responsibility. Whereas the Japanese writer presents part-time jobs as a means of individual growth and independence, the Chinese writer emphasizes social adaptation and collective benefit. Notably, *we have to* is the most frequent trigram containing *have* in the Chinese sub-corpus, reflecting a preference for expressing social obligation through collective identity.

In sum, the qualitative analysis of learner essays confirms that pronoun preferences shape distinct rhetorical strategies. Japanese writers tend to structure their arguments around personal experiences and individualized reasoning, while Chinese writers are more likely to employ collective appeals and obligation-based discourse. However, these patterns should not be overgeneralized, as frequent hybrid constructions such as *I think we should* complicate a strictly cultural interpretation. Nevertheless, the findings suggest that L1 discourse norms strongly influence EFL learners' use of first-person pronouns, reinforcing the view that authorial stance in L2 English writing is shaped by an interplay of linguistic and cultural factors.

5. Conclusion

This study investigated the use of first- and second-person pronouns (*I*, *we*, and *you*) among Korean, Chinese, and Japanese EFL learners across four proficiency levels, identifying L2 proficiency and L1 background as key influences on pronoun selection. Two notable findings emerged: (1) higher-proficiency learners generally used fewer first- and second-person pronouns, except at the earliest developmental stage (A2 to B1.1) and (2) Chinese learners preferred *we* over *I*, whereas Korean

and Japanese learners showed the opposite tendency. These results underscore the interplay of developmental, linguistic, and cultural factors in shaping authorial stance in English argumentative writing, extending prior research on cultural variability in L2 personal pronoun use (Luan & Zhang, 2016; Paquot et al., 2013; Petch-Tyson, 1998) to East Asian learners, an understudied group in this area.

For NNS, academic and professional writing poses unique challenges (Luzón, 2009), particularly regarding the use of first-person pronouns, which shape authorial presence and identity (Hyland, 2002). Many EAP instructors and textbooks discourage the use of *I* in academic writing, often advocating a blanket ban on first-person pronouns (Chang, 2015; McCrostie, 2008). However, our findings challenge such prescriptions, demonstrating that personal pronoun use is not merely a marker of informality or weak academic rigor, but rather a practice shaped by L2 proficiency and cultural background.

Rather than enforcing rigid rules, educators should help students recognize their own pronoun preferences and understand their rhetorical impact (Tang & John, 1999). This requires shifting the focus from avoidance to strategic use, teaching students when and how to integrate personal pronouns effectively in academic writing. As Hyland (2001) argues, self-mention is not simply a stylistic choice but a fundamental component of academic discourse, enabling writers to assert authority and position themselves within their discourse communities. Thus, instead of prioritizing strict adherence to Western academic norms, students should cultivate rhetorical awareness and develop writing styles that align with both their cultural background and personal identities.

This study has several limitations that future research could address. An interesting finding was the temporary increase in pronoun usage between A2 and B1.1 learners, which may stem from intermediate learners experimenting with rhetorical strategies, overgeneralizing learned structures (e.g., *I think*), or being influenced by instructional practices and L1 discourse patterns. Future studies could explore how writing strategies evolve across proficiency levels and how pedagogical approaches shape pronoun use. In addition, this study did not account for intra-group variation potentially related to individual differences such as personality, aptitude, and motivation (Dörnyei, 2005), or demographic factors such as gender, age, and educational level (Gablasova et al., 2017). As the presence of outliers suggests their potential impact, future studies could draw on available metadata to conduct more detailed investigations of these variables. Lastly, since this study focused solely on argumentative essays, examining other genres could provide deeper insights into how

genre conventions affect pronoun choice. Despite these limitations, the findings contribute to a more comprehensive understanding of how L1 background and proficiency influence EFL learners' use of personal pronouns, offering valuable implications for both language research and pedagogy.

References

Aguado-Jiménez, P., Pérez-Paredes, P., & Sánchez, P. (2012). Exploring the use of multi-dimensional analysis of learner language to promote register awareness. *System*, *40*(1), 90-103.

Aijmer, K. (2001). *I* think as a marker of discourse style in argumentative student writing. In K. Aijmer (Ed.), *A wealth of English: Studies in honour of Göran Kjellmer* (pp. 247-259). Acta Universitatis Gothenburgensis.

Akahori, N. (2007). The excessive use of the first-person pronoun *I* in English compositions by Japanese students. *Nihonjin eigo gakushusha no hanashikotoba kakikotoba no corpus sakusei to sono goyoron teki taishobunseki* [Building learner corpora of Japanese learners of English and the contrastive analyses]. (Report for Grant-in-Aid for Scientific Research (B) No. 15320059). Brainsnetwork.

Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntConc

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (2006). *University language* (Vol. 10). John Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. Longman.

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-54). Ablex.

Chang, J. Y. (2015). A comparison of the first-person pronoun *I* in NS and Korean NNS corpora of English argumentative writing. *English Teaching*, *70*(2), 83-106.

Chang, Y.-Y., & Swales, J. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In C. Candlin & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 145-167). Longman.

Chen, R. (1993). Responding to compliments: A contrastive study of politeness strategies between American English and Chinese speakers. *Journal of Pragmatics*, *20*(1), 49-75.

Chen, R. (2001). Self-politeness: A proposal. *Journal of Pragmatics*, *33*(1), 87-106.

Chen, R. (2020). Single author self-reference: Identity construction and pragmatic competence. *Journal of English for Academic Purposes*, *45*, 100856.

Choung, Y., & Oh, S. Y. (2017). A systemic functional study of thematic organization in the English writing of Korean college students. *English Teaching*, *72*(3), 119-144.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language*

acquisition. Routledge.

Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some implementations of the boxplot. *The American Statistician*, *43*(1), 50-54.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning*, *67*(S1), 130-154.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41-61.

Granger, S. (1996). From CA to CIA and back: an integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund University Press.

Granger, S. (1998a). *Learner English on computer.* Routledge.

Granger, S. (1998b). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). Longman.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7-24.

Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020). *The international corpus of learner English* (Version 3). Presses Universitaires de Louvain.

Gu, Y. (1990). Politeness phenomena in modern Chinese. *Journal of Pragmatics*, *14*(2), 237-257.

Hasselgård, H. (2009). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer (Ed.), *Corpora and language teaching* (pp.121-139). John Benjamins.

Hinkel, E. (2002). *Second language writers text: Linguistic and rhetorical features*. Lawrence Erlbaum.

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, *41*(4), 667-683.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Sage Publications.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, *2*(1).

Hofstede, G., Hofstede G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind revised and expanded 3rd edition*. McGraw-Hill.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing.* Longman.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, *20*(3), 207-226.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, *34*(8), 1091-1112.

Institute of International Education (IIE). (2023). *Open doors 2023 fast facts*. Retrieved from https://opendoorsdata.org/fast_facts/fast-facts-2023/

Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. In K. Yagi & T. Kanzaki (Eds.), *Phraseology, corpus linguistics, and lexicography: Papers from phraseology* (pp. 87-100). Kwansei Gakuin University Press.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner Corpus Studies in Asia and the World*, *1*(1), 91-118.

Ivanič, R., & Camps, D. (2001). I am how I sound: Voice as self-representation in L2 writing. *Journal of Second Language Writing*, *10*(1-2), 3-33.

Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology*, *29*(3), 461-486.

Kitagawa, C., & Lehrer, A. (1990). Impersonal uses of personal pronouns. *Journal of Pragmatics*, *14*(5), 739-759.

Kobayashi, Y., & Abe, M. (2016). A corpus-based approach to the register awareness of Asian learners of English. *Journal of Pan-Pacific Association of Applied Linguistics*, *20*(2), 1-17.

Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, *18*(2), 121-138.

Kwon, O. (2000). Korea's English education policy changes in the 1990s: Innovations to gear the nation for the 21st century. *English Teaching*, *55*(1), 47-91.

Lee, E. K., & Oh, S. Y. (2018). Korean EFL use of *I* in English argumentative writing: Focusing on genre-specific and proficiency-specific characteristics. *Language Research*, *54*(2), 357-381.

Lee, S. W. (2008). Investigating ideational thematic content in Korean university student essay writing. *English Teaching*, *63*(2), 71-94.

Li, N., & Lee, T. K. (2013). Authorial voice in academic writing: A contrastive perspective. *Journal of PLA University of Foreign Languages*, *36*, 17-23, 40, 127. (In Chinese)

Luan, J., & Zhang, Y. (2016). A comparative study of the use of first-person subject pronouns in Chinese and Swedish students' argumentative essays. A corpus-based research. *International Journal of Education and Research*, *11*, 341-352.

Luukka, M. R. (1994). Metadiscourse in academic texts. In B. L. Gunnarsson, P. Linell, & B. Nordberg (Eds.), *Text and talk in professional context* (pp. 77-88). ASLA (The Swedish Association of Applied Linguistics).

Luzón, M. J. (2009). The use of *we* in a learner corpus of reports written by EFL engineering students. *Journal of English for Academic Purposes*, *8*(3), 192-206.

Markkanen, R., Steffensen, M., & Gismore, A. (1993). Quantitative contrastive study of metadiscourse: Problems in design and analysis of data. *Papers and Studies in Contrastive Linguistics, 28*, 137-151.

Martínez, I. A. (2005). Native and non-native writers' use of first-person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing*, *14*(3), 174-190.

McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study.

ICAME Journal, *32*(1), 97-114.

Na, J., & Choi, I. (2009). Culture and first-person pronouns. *Personality and Social Psychology Bulletin*, *35*(11), 1492-1499.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9-13.

Natsukari, S. (2012). Use of *I* in essays by Japanese EFL learners. *JALT Journal*, *34*(1), 61-78.

Oh, S.-Y., & Kang, S.-J. (2013). The effect of English proficiency on Korean undergraduates' expression of epistemic modality in English argumentative writing. *The Journal of Asia TEFL, 10*(4), 97-132.

Paquot, M., Hasselgård, H., & Ebeling, S. O. (2013). Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead. Proceedings of the First Learner Corpus Research Conference (LCR 2011)* (pp. 377-387). Presses Universitaires de Louvain.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107-118). Longman.

R Studio Team. (2025). R Studio: Integrated development environment for R (Version 2024.12.1. +563.pro5) [Computer software]. Posit PBC. https://posit.co/download/rstudio- desktop/

Ramanathan, V., & Atkinson, D. (1999). Individualism, academic writing, and ESL writers. *Journal of Second Language Writing, 8*(1), 45-75.

Recski, L. J. (2004). Expressing standpoints in EFL written discourse. *Revista Virtual de Estudos da Linguagem, 3*(2).

Redeker, G. (1984). On differences between spoken and written language. *Discourse Processes*, *7*, 43-55.

Rhee, S.-C., & Jung, C. K. (2014). Compilation of the Yonsei English Learner Corpus (YELC) 2011 and its use for understanding current usage of English by Korean pre-university students. *The Korea Contents Society, 14*(11), 1019-1029.

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A crosslinguistic approach. In S. Granger (Ed.), *Learner English on computer* (pp. 41-52). Longman.

Ryoo, M.-L. (2010). Use of the pronouns 'I' and 'we' by Korean EFL students of differential writing skill levels. *English Language Teaching*, *22*(3), 77-97.

Smith, E. L. (1986). Achieving impact through the interpersonal component. In B. Couture (Ed.), *Functional approaches to writing: Research perspectives* (pp. 108-119). Frances Pinter.

Song, J. J. (2006). *The Korean language: Structure, use, and context*. Routledge.

Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first-person pronoun. *English for Specific Purposes*, *18*, 23-39.

Tannen, D. (1982). The myth of orality and literacy. In W. Frawley (Ed.), *Linguistics and literacy* (pp. 37-50). Plenum Press.

Tracy-Ventura, N. (2020). The future of corpora in SLA. In N. Tracy-Ventura & M. Paquot (Eds.), *The Routledge handbook of SLA and corpora* (pp. 411-426). Routledge.

Uz, I. (2014). Individualism and first-person pronoun use in written texts across languages.

Journal of Cross-Cultural Psychology, *45*(10), 1671-1678.
Xia, G. (2018). A cross-disciplinary corpus-based study on English and Chinese native speakers' use of first-person pronouns in academic English writing. *Text & Talk*, *38*(1), 93-113.
Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53-67.
Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130-141.
Yoon, H. J. (2021). Interactions in EFL argumentative writing: Effects of topic, L1 background, and L2 proficiency on interactional metadiscourse. *Reading and Writing, 34*(3), 705-725.

Seonwoong Kim
Graduate Student
Department of English Language Education
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
Email: coolbaugh1@snu.ac.kr

Sun-Young Oh
Professor
Department of English Language Education
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
Email: sunoh@snu.ac.kr

Received: February 28, 2025
Revised version received: March 27, 2025
Accepted: April 4, 2025