

Teaching English Articles to Japanese Learners: Insights from Corpus Data and Classroom Applications

Ayako Aizawa¹

Graduate School, Rikkyo University, Japan

ABSTRACT

English articles are widely considered one of the most challenging grammatical features for Japanese learners of English, whose first language lacks an article system. This paper draws on corpus data from the Longman Learners' Corpus (LLC) and British and American English reference corpora to illustrate how definite article usage patterns vary across proficiency levels among first language (L1) Japanese learners. The data reveal an underuse of the definite article at lower proficiency levels, with usage frequency increasing as proficiency develops. Qualitative analysis of concordance lines identifies specific error patterns, including article omission before superlatives and inappropriate use before proper nouns. Based on these findings, this paper proposes practical, corpus-informed classroom activities that teachers in Japan can implement, including step-by-step procedures and freely available tools. Throughout the discussion, the paper adopts a World Englishes perspective, framing corpus-based norms as one reference point among many rather than as a prescriptive standard.

¹ Ayako Aizawa is a doctoral candidate at the Graduate School of Intercultural Communication, Rikkyo University. Her main area of research is vocabulary testing. She teaches English at several universities in Tokyo. Correspondence should be sent to Ayako Aizawa, Graduate School of Intercultural Communication, Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan. E-mail: a.aizawa@rikkyo.ac.jp.

INTRODUCTION

The English article system is widely regarded as one of the most difficult grammatical features for second language (L2) learners to acquire (Butler, 2002; Master, 1990; Park, 2023). This difficulty is particularly acute for learners whose first language (L1) does not include an article system. Japanese, for instance, lacks articles entirely, although particles such as *wa* (は) and *ga* (が) can serve some related functions in marking information status (Snape et al., 2013). Research has shown that L1 Japanese learners acquire English articles at a later stage than learners whose L1s have article systems, such as Spanish, German, and French (Murakami & Alexopoulou, 2016), and that even advanced Japanese learners continue to differ from L1 English users in their article use (Butler, 2002).

A corpus, in the context of linguistics, is a large, systematically collected body of authentic language stored in electronic format (Callies & Paquot, 2015). A learner corpus is a specific type of corpus that contains language produced by L2 learners. When researchers or teachers examine corpus data, they often use concordance lines, which display every occurrence of a target word or phrase within its surrounding context, typically in a Key Word In Context (KWIC) format. This format aligns the target word in the center of each line with several words of context on either side, allowing users to quickly identify patterns in how the word is used. For example, a concordance line for the word *the* might show *...he was best soccer player in **the** class...*, with the node word aligned in the center and surrounding context visible on either side (see Figure 2 for a fuller example). Learner corpora have been used to study various aspects of L2 language use, including the overuse and underuse of specific language features and the influence of the L1 on learner production (Leech, 1998).

While corpus data have contributed to the understanding of how learners use English articles, relatively little attention has been given to how these findings can be translated into practical classroom activities, especially for teachers working in the Japanese English as a Foreign Language (EFL) context (Mizumoto & Chujo, 2015; Poole, 2022). Data-driven learning (DDL), an approach in which learners explore corpus data to discover language patterns inductively, has been shown to be effective across a range of language learning contexts. A meta-analysis by Boulton and Cobb (2017) found a large overall effect for DDL on language learning outcomes, suggesting that direct engagement with corpus data can facilitate acquisition of grammatical and lexical features. However, applications of DDL to article instruction for Japanese learners remain underexplored.

An important consideration in any corpus-based study of learner language is the question of norms. Comparing learner production with L1 English user corpora is a common practice in learner corpus research (Flowerdew, 2015), but it does not imply that any particular variety of English should serve as the sole target. As Gilquin (2022) has argued, corpus-derived norms should be approached critically, recognizing that no single variety of English holds inherent superiority. This paper adopts this perspective throughout, treating the comparison with L1 English user data as one useful reference point while acknowledging learners' right to make informed choices about the kind of English they wish to use.

This paper has two aims. First, it presents corpus data on how L1 Japanese learners use the definite article *the* across proficiency levels, drawing on the Longman Learners' Corpus and reference corpora of British and American English. Second, it proposes concrete, corpus-

informed classroom activities with step-by-step procedures, accessible to teachers who may not have prior experience with corpus tools. The paper begins with a description of the corpora and analytical procedures, then presents quantitative and qualitative findings, and finally outlines classroom applications with practical guidance for implementation.

CORPUS DATA ON JAPANESE LEARNERS' USE OF THE DEFINITE ARTICLE

Corpora and Procedures

Three corpora were used in this study. The Longman Learners' Corpus (LLC; Longman Corpus Network, n.d.) served as the primary data source. The LLC is a 10-million-word corpus consisting of essays and exam scripts written between 1990 and 2002 by learners of 20 different L1 backgrounds at eight levels of proficiency. Each essay is coded for country of data collection, environment of learning, learner proficiency level, target variety of English, and task type (Xiao, 2008). For this study, the task type was restricted to essays (set essay, free essay, and project essay), the country of data collection was restricted to Japan, and the learner's L1 was restricted to Japanese. Seven proficiency levels were queried: beginners, elementary, pre-intermediate, intermediate, upper-intermediate, advanced, and proficiency. There were no texts by L1 Japanese learners at the eighth level (English for Academic Studies). The LLC was selected because, unlike more recent learner corpora, it records essays by L1 Japanese learners across a full range of proficiency levels rather than a single level or a narrow band, which was necessary for examining how article use develops over the course of acquisition.

Two reference corpora were used for comparison. British English 2006 (BE06; Baker, 2009) and American English 2006 (AmE06; Baker, 2017; Potts & Baker, 2012) are each approximately one-million-word corpora of written English published around 2006. Both are members of the Brown Family of corpora, designed to be comparable to each other using the same sampling frame of 15 genres (Xiao, 2008). From these corpora, 80 texts each from text category J (academic prose in various disciplines) were selected, as academic prose was considered the most appropriate register for comparison with learners' written academic essays in the LLC. BE06 and AmE06 were chosen, rather than other available reference corpora, because their comparable design allowed article use to be benchmarked against both major national varieties of English rather than a single standard, and because their genre categories made it possible to isolate academic prose specifically for a like-for-like comparison with the LLC essays.

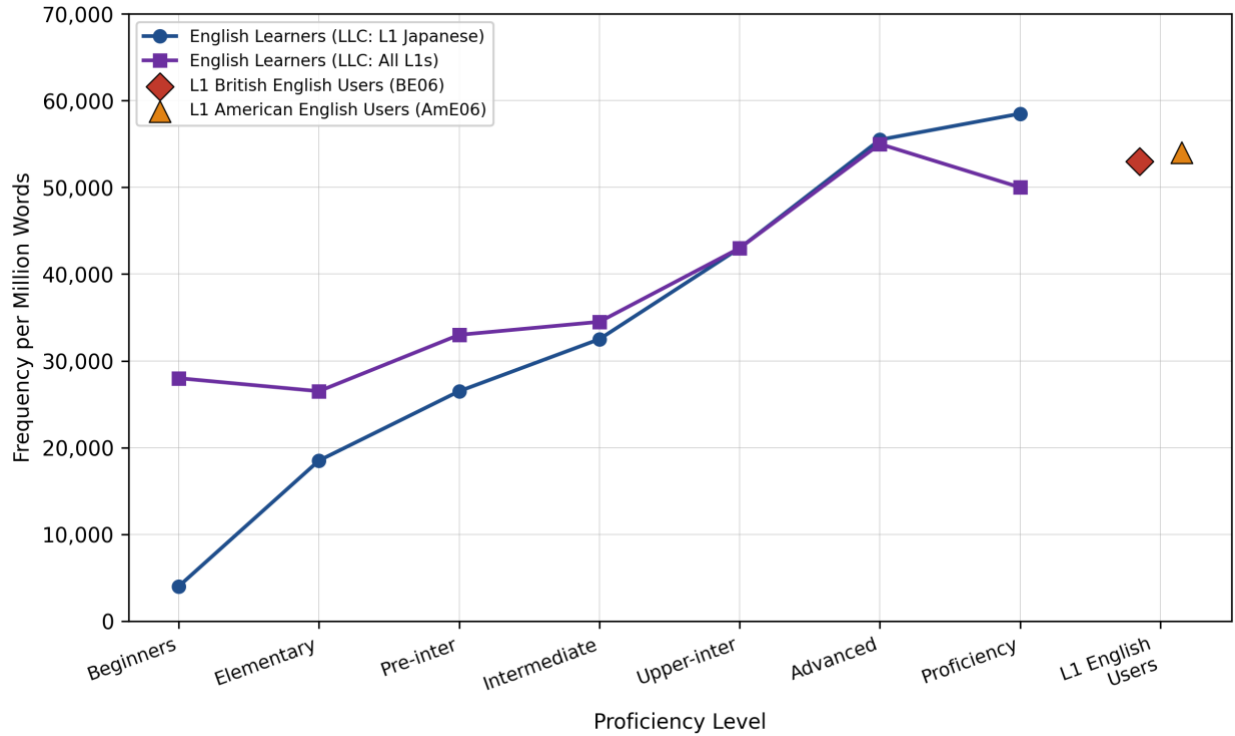
All queries were conducted using the restricted query function of CQPweb (Hardie, 2012), a freely available web-based corpus analysis platform. Concordance lines were examined in KWIC format, and relative frequencies (per million words) were calculated for comparison across subcorpora of different sizes.

Quantitative Findings

Figure 1 illustrates the relative frequency of *the* across proficiency levels for L1 Japanese learners, all learners in the LLC, and L1 English users in BE06 and AmE06. Among L1 Japanese learners, the frequency of *the* increased consistently from beginners to proficiency level. At the

beginner level, *the* appeared only 4,180 times per million words, which represents approximately 7.8% of the frequency observed in BE06 (53,366 per million words). Beginners in the LLC as a whole used *the* 6.7 times more frequently than L1 Japanese beginners, suggesting that the underuse at the beginner level is particularly characteristic of L1 Japanese learners rather than a universal feature of beginning learners. By the proficiency level, L1 Japanese learners' frequency of use was comparable to that of L1 English users in both BE06 and AmE06.

Figure 1
Relative Frequency of “the” in Academic Writing Across Proficiency Levels



This developmental pattern is consistent with Butler (2002), who found that L1 Japanese learners with higher proficiency performed more target-like on article tasks. However, the quantitative similarity at the proficiency level does not necessarily mean that learners are using *the* in the same ways as L1 English users. Quantitative frequency alone cannot reveal whether the contexts of use are appropriate, which is why qualitative analysis of concordance lines is also essential.

To examine whether the quantitative findings reflect consistent article use across individual learners or are driven by a few heavy users, the dispersion of *the* was also examined. Table 1 presents the number and percentage of essays containing the definite article at each proficiency level. Only 36% of L1 Japanese beginner essays contained *the*, compared with 73.5% of all beginner essays in the LLC and 100% of texts in BE06 and AmE06. The dispersion data confirm that the underuse at the beginner level is widespread across individual learners rather than being an artifact of a few outliers. From the upper-intermediate level onward, 98.5% or more L1 Japanese learners used the definite article in their essays.

Table 1
Dispersion of “the” in Academic Writings

Proficiency	LLC: Japanese Dispersion	%	LLC: All L1s Dispersion	%
Beginners	9/25	36.0	1,023/1,392	73.5
Elementary	746/950	78.5	2,993/3,686	75.1
Pre-intermediate	436/572	76.2	3,596/4,052	88.7
Intermediate	413/462	89.4	5,948/6,221	95.6
Upper-intermediate	725/736	98.5	4,802/4,900	98.0
Advanced	95/96	99.0	3,263/3,289	99.2
Proficiency	18/18	100.0	1,763/1,764	99.9
L1 English Users			160/160	100.0

Note. Dispersion shows the number of texts containing the out of the total number of texts at each proficiency level. L1 English user data are from BE06 and AmE06 (80 texts each).

To further confirm whether the underuse was specific to the definite article, the indefinite articles *a* and *an* were also searched among L1 Japanese beginners. They appeared in only 12 out of 25 essays, while they appeared in all 80 texts in both BE06 and AmE06. This suggests that the underuse of articles by L1 Japanese beginners extends to the article system as a whole, not only the definite article.

Qualitative Findings: Error Patterns in Concordance Lines

To move beyond frequency counts and understand how L1 Japanese learners use the definite article, concordance lines were examined qualitatively at different proficiency levels. Concordance lines display each occurrence of a search term within its surrounding textual context, making it possible to identify patterns of use and misuse. Figure 2 shows an example of KWIC concordance lines from CQPweb.

Figure 2
Concordance Lines (KWIC Format) for “the” in L1 Japanese Beginner Essays

1	11464	I want to see you again . Yamagiwa had played good at the baseball . He had a bad head . He had been deingerous
2	11466	Manabu were going to jounior high school . He was always playing the basketball . Takizawa was my jounior high school teacher . He had
3	11467	have been studying math for a long time . Yasuaki He was the best clever friend . He has been going to Shouwa Tetudo .
4	11467	and write three short sentences about each friend [fig] Nobumasa He was the best funy friend : we liked " mortal kombat " " Fatality
5	11476	write three short sentences about each friend [fig] 1 . Tamura was the best my friend . He was basketball club . We played TV.
6	11466	was a number of volleyball club with me . And he was the best of player . I and Manabu were going to jounior high
7	11474	He practices every day . Shinsei likes rock music . He plays the drums very well . He goes to music school . Kujiro has
8	11460	Ohkaba was my friend . He wants to be a doctor in the future . Perhaps he wo n't to be doctor . because he
9	11464	rich . So he had many cars . And he had practiced the karatedo He is strong . I want visted to his house .
10	11468	" Thuyoshi Nagabuthi " he is very strong he want to become the rock'n rooler . It's a his dream . Jaikuda was same
11	11476	hard work . Now he is best my friend . Saitoh is the storongest in Fluchu . He is going to Showadaichi high school .
12	11477	very good player . Kumagai is my best friend . He is the tallest boy in my friends . He is kind boy . Make
13	11455	He is same school with me in Toyo . He was playing the Torrnbone in elementary school . Akash was going to school with me

Note. Screenshot from CQPweb (Hardie, 2012) showing all 13 occurrences of *the*, sorted by one word to the right of the node word.

Among L1 Japanese beginners, only 13 instances of *the* were found across all essays. Examination of these concordance lines revealed that six instances (lines 3, 4, 5, 6, 11, and 12 in Figure 2) involved *the* correctly used before superlatives. Two instances (lines 7 and 13) were correctly used before musical instruments (drums and trombone). However, three instances involved misuse: *the* was placed before sports nouns (e.g., *the baseball*, *the basketball*, *the karatedo*), where a zero article is required in target-like English. This means that of only 13 uses of *the* by beginners, three (23%) were inappropriate in the context of written English conventions.

However, searching for *the* alone can give a misleading picture. To investigate whether beginners were consistently applying the rule of using *the* before superlatives, a separate query was conducted searching for *most* and **est* (where *** is a wildcard allowing any preceding characters) among L1 Japanese beginner essays. As Figure 3 shows, 38 superlative uses were found in total, and 8 of them lacked the required preceding definite article or possessive noun. This demonstrates that even when *the* appears to be used correctly in some instances, many learners at the same level are omitting it in obligatory contexts.

Figure 3
Examples of Superlative Constructions Without Preceding Articles in L1 Japanese Beginner Essays

1	hard work . Now he is	<u>best</u>	my friend . Saitoh is the
2	best friends . He was Kuwana	<u>best</u>	friends too . He was baseball
3	Kuwana was my and Maeda	<u>best</u>	friends . He was best soccer
34	to soccer club . He was	<u>best</u>	of junior high school . Mr
35	Maeda best friends . He was	<u>best</u>	soccer player . He was studying
36	was my friends . He was	<u>Best</u>	basedall player . He was studying
37	watching baseball game . He was	<u>Best</u>	soccer player . Katou was my

Note. Sorted by one word left of the node words **est* or *most*. Line numbers correspond to position in the full concordance output. These examples show cases where the superlative appears without a preceding definite article or possessive noun.

At the elementary to upper-intermediate levels, additional error patterns emerged. Learners showed a tendency to add *the* before Japanese words written in the Roman alphabet without translation (e.g., *the kendo*, *the kimono*, *the oshogatsu*). More notably, learners from elementary through upper-intermediate levels deviated from conventions by adding *the* before proper nouns combined with common nouns such as station or airport. In BE06, neither Heathrow Airport nor Paddington Station were preceded by *the*, but L1 Japanese learners used *the* before proper noun + station/airport combinations in 6.2% of relevant instances (23 out of

371 occurrences), producing forms such as *the Narita Airport*, *the Haneda Airport*, and *the Asakusa Station*.

These errors may reflect multiple sources. Butler (2002) noted that lower proficiency learners tend to rely on textbook rules and apply them without sufficient attention to exceptions. English textbooks used in Japanese secondary education typically introduce articles with a simple rule that the indefinite articles *a* and *an* indicate unspecified singular nouns while *the* refers to something already mentioned or clearly known. No exceptions are typically mentioned at this stage, which may lead some learners to believe that all nouns must be preceded by an article. Mizuno (1985) suggested that the underuse of articles by L1 Japanese learners may also reflect avoidance strategies, in which learners deliberately omit features they are uncertain about. Japanese uses numerals (e.g., 1 個の; *ikko-no*) and demonstratives (e.g., その; *sono*) that function similarly to English articles, but unlike in Mandarin, their use is optional in Japanese, which may lead learners to treat English articles as similarly optional.

Proficient Learners' Use

In contrast to the patterns observed at lower proficiency levels, a random sample of 200 concordance lines from proficiency-level L1 Japanese learners showed entirely appropriate use of *the*. The same superlative query found that all 31 occurrences at this level were preceded by the required article. No misuse before proper noun + station/airport was found. These findings support the view that article use by L1 Japanese learners becomes increasingly consistent with English conventions as proficiency develops, both in terms of frequency and the contexts in which *the* is used.

CLASSROOM APPLICATIONS

The corpus findings described above point to specific areas where teachers can provide targeted support. The following activities are designed for teachers in Japan, including those who may not have prior experience with corpus tools. Each activity includes step-by-step procedures and suggestions for materials.

Activity 1: Noticing Article Patterns Through Concordance Lines

This activity uses concordance lines to help learners notice how the definite article is used in authentic English. It is suitable for intermediate and above learners and takes approximately 20 to 30 minutes.

Preparation: The teacher selects 10 to 15 concordance lines from a reference corpus showing common patterns of the + noun (e.g., *the government*, *the university*, *the end of*). These can be obtained using freely available tools such as SkELL (Sketch Engine for Language Learning; <https://skell.sketchengine.eu>), which provides concordance lines from a large English corpus without requiring registration or payment. The teacher also selects 5 to 10 concordance lines from learner writing (either from a learner corpus or from students' own essays) that show either appropriate or inappropriate article use.

Procedure: (1) The teacher distributes the concordance lines from the reference corpus and asks learners to underline *the* and identify what kind of noun follows it. (2) Learners work in pairs to categorize the patterns they find (e.g., the + previously mentioned noun, the + unique noun, the + superlative). (3) The teacher then distributes the learner corpus concordance lines and asks learners to identify which uses match the patterns they found and which do not. (4) The class discusses why certain uses may differ from the reference corpus patterns, considering the possibility that some differences may reflect legitimate variation rather than errors.

Activity 2: Identifying Zero-Article Contexts

This activity focuses on contexts where articles should not be used, addressing the error patterns identified in the corpus data. It is suitable for elementary and above learners and takes approximately 15 to 20 minutes.

Preparation: The teacher creates a worksheet containing 15 to 20 sentences, some with correct article use and some with common errors drawn from the patterns identified in the corpus analysis. For example, sentences can include cases such as the following:

Sentence	Article Use
She arrived at the Narita Airport.	Unconventional Use
She arrived at Narita Airport.	Conventional Use
He plays the baseball every day.	Unconventional Use
He plays baseball every day.	Conventional Use
She is best student in the class.	Unconventional Use (Missing article)

Procedure: (1) Learners work individually to identify whether the article use in each sentence follows English conventions. (2) In pairs, they discuss their answers and try to formulate a rule for when *the* is not used (e.g., before sports, before proper noun + station/airport). (3) The teacher facilitates a class discussion, presenting additional examples and helping learners refine their understanding. The teacher can note that these conventions reflect patterns commonly found in published English writing, while acknowledging that variation exists across different English-speaking contexts.

Activity 3: Learning “the” Through Frequent Phrases

This activity introduces the definite article through high-frequency formulaic sequences, reducing the cognitive burden of applying abstract grammatical rules. It is suitable for intermediate and above learners and takes approximately 20 minutes.

Background: Research has shown that a substantial proportion of natural language use is formulaic, with fixed or semi-fixed phrases occurring with high frequency (Erman & Warren, 2000). Data from the British National Corpus show that *the* appears in many of the most frequent n-grams (multi-word sequences) and phrase frames (sequences with one variable slot). Table 2 lists examples of high-frequency four-word phrase frames containing *the*, drawn from the British National Corpus via the Phrases in English tool (Fletcher, n.d.; <https://phrasesinenglish.org>).

Rather than teaching articles as isolated grammatical items, these phrases can serve as vocabulary chunks that learners acquire as whole units, naturally embedding article use within meaningful expressions.

Table 2
High-Frequency Four-Word Phrase Frames Containing “the” From the British National Corpus

Phrase Frame	Frequency	Top Variable Words
the * of the	241,612	end, rest, top
in the * of	80,186	case, middle, form
at the * of	34,699	end, time, top
one of the *	29,402	most, first, main
for the * of	27,590	sake, rest, purpose
on the * of	34,699	basis, top, part
the * and the	19,931	state, role, nature

Note. The asterisk (*) indicates a variable slot where different words can be inserted. Frequency refers to total occurrences in the British National Corpus. Data obtained from Phrases in English (Fletcher, n.d.).

Procedure: (1) The teacher selects three to five phrase frames appropriate for the learners’ level. (2) Learners are given the phrase frames with blank slots and asked to guess what words might fill the variable position. (3) The teacher reveals the most frequent words from the corpus data and provides example sentences using each phrase. (4) Learners practice using the phrases in their own sentences or short paragraphs. This approach allows learners to acquire article use within meaningful chunks rather than applying abstract rules to individual nouns.

Freely Available Corpus Tools for Teachers

Teachers who wish to explore corpus data further or create their own DDL materials can access several freely available tools. Table 3 lists tools that can be used without charge and without specialized technical knowledge.

Table 3
Freely Available Corpus Tools for Teachers

Tool	URL	Description
SkELL	https://skell.sketchengine.eu	Concordance lines, word sketches, and examples from a large English corpus. No registration required.
COCA	https://www.english-corpora.org/coca/	Corpus of Contemporary American English. Free account allows limited searches.
Phrases in English	https://phrasesinenglish.org	N-gram and phrase frame statistics from the British National Corpus.
AntConc	https://www.laurenceanthony.net/software/antconc/	Free downloadable concordance software for analyzing small corpora, such as student essays.

Using AntConc, teachers can build small corpora from their own students' writing and examine article use patterns specific to their classroom context. Even a collection of 20 to 30 student essays can reveal patterns that are common among learners in the same class, enabling teachers to design lessons that address shared difficulties (Reppen, 2010). By collecting student writing over time, teachers can also monitor developmental changes in article use.

DISCUSSION

The corpus data presented in this paper reveal a developmental trajectory in L1 Japanese learners' use of the definite article, from underuse at the beginner level to L1 English-like frequency and usage at the proficiency level. The qualitative analysis identified specific error patterns that can inform targeted instruction, including article omission before superlatives, inappropriate use before sports and proper noun + station/airport combinations, and a tendency to attach *the* to Japanese words written in the Roman alphabet.

These results extend Butler's (2002) finding that L1 Japanese learners' article knowledge grows more target-like with proficiency by showing where, specifically, that growth is uneven: frequency of *the* converges on L1 English user norms well before certain contexts of use, such as proper noun + station/airport combinations, are fully mastered. This distinction matters for teachers, who may otherwise assume that once learners are producing *the* at a rate comparable to L1 English users, article instruction is no longer necessary. The concordance-based error patterns identified here, particularly the overextension of *the* to Romanized Japanese words and to proper noun + common noun combinations, give teachers concrete, learner-relevant targets for instruction rather than the more general guidance available from textbook rules alone. In this respect, the contribution of this study is to translate corpus findings on article use, which have mostly circulated within corpus linguistics research, into a form that is directly usable in the L1 Japanese EFL classroom, consistent with the aim of DDL approaches described by Boulton and Cobb (2017).

These findings should be interpreted with some important caveats. First, the LLC data were collected between 1990 and 2002, and the BE06 and AmE06 corpora reflect language published around 2006. Language use and English education in Japan have evolved since then, and more recent data might reveal different patterns. Second, the number of essays by L1 Japanese beginners (25) and proficiency-level learners (18) is relatively small, limiting the generalizability of the findings. A larger corpus, such as the Japanese section of the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2014), could provide more robust data. Third, the LLC essays were written under exam conditions, which may not reflect learners' typical writing performance. These limitations point to directions for future research. Replicating this analysis with a more recent, larger corpus such as ICNALE, and with writing produced outside exam conditions, would help establish whether the developmental pattern reported here still holds for current learners. Future work could also track individual learners longitudinally, rather than comparing proficiency levels cross-sectionally, to confirm that the trajectory observed here reflects genuine development rather than cohort differences, and could extend the qualitative error analysis to other article-related contexts beyond superlatives, sports nouns, and proper noun + common noun combinations.

From a World Englishes perspective, it is important to note that the classroom activities proposed in this paper are not intended to enforce conformity with British or American English

norms. The comparison with BE06 and AmE06 serves as a descriptive reference point for understanding how article use differs across learner and L1 English user populations. In classroom practice, teachers should present these patterns as information that helps learners make informed choices about their own language use. As Gilquin (2022) has emphasized, the goal of corpus-informed teaching should be to expand learners' linguistic repertoires rather than to impose a single standard. Some learners may choose to aim for article use that closely resembles published written English, while others may develop their own patterns as competent users of English as a lingua franca. Both choices should be respected.

Taken together, the quantitative and qualitative findings suggest that instruction on the definite article should shift focus as learners advance: at lower proficiency levels, the priority is encouraging use of *the* at all, given the marked underuse documented among beginners; at intermediate levels, the priority shifts to the specific contexts identified in this study where L1 Japanese learners diverge from L1 English user conventions. Framed this way, corpus-informed instruction is not a one-time intervention but a set of tools that can be revisited as learners' article use develops, with the World Englishes perspective adopted throughout this paper ensuring that such instruction supports rather than overrides learners' own choices about their English.

CONCLUSION

This paper has drawn on corpus data to illustrate how L1 Japanese learners' use of the English definite article develops across proficiency levels, and to propose practical classroom activities that teachers in Japan can implement. The underuse at the beginner level, the identifiable error patterns at intermediate levels, and the developmental trajectory toward L1 English-like use all provide a foundation for corpus-informed teaching. The activities presented here require no special resources beyond freely available corpus tools and can be adapted to suit learners at various levels. By incorporating DDL approaches, formulaic sequence instruction, and a critical awareness of language norms, teachers can support learners in developing their article use in ways that are both research-informed and respectful of learner agency.

REFERENCES

- Baker, P. (2009). The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics*, 14(3), 312–337. <https://doi.org/10.1075/ijcl.14.3.02bak>
- Baker, P. (2017). *American and British English: Divided by a common language?* Cambridge University Press. <https://doi.org/10.1017/9781316227589>
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Butler, Y. G. (2002). Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system. *Studies in Second Language Acquisition*, 24(3), 451–480. <https://doi.org/10.1017/S0272263102003042>

- Callies, M., & Paquot, M. (2015). Learner corpus research: An interdisciplinary field on the move. *International Journal of Learner Corpus Research*, 1(1), 1–6.
<https://doi.org/10.1075/ijlcr.1.1.001edi>
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, 20(1), 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Fletcher, W. H. (n.d.). *Phrases in English*. Retrieved December 31, 2021, from <https://phrasesinenglish.org/index.html>
- Flowerdew, L. (2015). Learner corpora and language for academic and specific purposes. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 465–484). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139649414.021>
- Gilquin, G. (2022). One norm to rule them all? Corpus-derived norms in learner corpus research and foreign language teaching. *Language Teaching*, 55(1), 87–90.
<https://doi.org/10.1017/S0261444821000185>
- Hardie, A. (2012). CQPweb—Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380–409.
<https://doi.org/10.1075/ijcl.17.3.04har>
- Ishikawa, S. (2014). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner Corpus Studies in Asia and the World*, 1, 91–118.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv–xx). Longman.
- Longman Corpus Network. (n.d.). *The Longman Learners' Corpus*. Retrieved December 28, 2021, from <http://www.pearsonlongman.com/dictionaries/corpus/learners.html>
- Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly*, 24(3), 461–478. <https://doi.org/10.2307/3587230>
- Mizuno, H. (1985). A psycholinguistic approach to the article system in English. *JACET Bulletin*, 16, 1–29.
- Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, 22, 1–18.
- Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition*, 38(3), 365–401. <https://doi.org/10.1017/S0272263115000352>
- Park, S. (2023). Corpus analysis of L2 English article usage patterns & pedagogical implications. *Cogent Education*, 10(1), Article 2197662. <https://doi.org/10.1080/2331186X.2023.2197662>
- Poole, R. (2022). “Corpus can be tricky”: Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, 35(7), 1620–1641.
<https://doi.org/10.1080/09588221.2020.1825095>
- Potts, A., & Baker, P. (2012). Does semantic tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics*, 17(3), 295–324.
<https://doi.org/10.1075/ijcl.17.3.01pot>

- Reppen, R. (2010). Building a corpus: What are the key considerations? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 31–37). Routledge. <https://doi.org/10.4324/9780203856949.ch3>
- Snape, N., Mayo, M. D. P. G., & Gürel, A. (2013). L1 transfer in article selection for generic reference by Spanish, Turkish and Japanese L2 learners. *International Journal of English Studies*, 13(1), 1–28. <https://doi.org/10.6018/ijes/2013/1/138701>
- Xiao, R. (2008). Well-known and influential corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 383–457). Mouton de Gruyter.