

日本人EFL学習者の発話コーパスにおける冠詞の使用
The Use of Articles in Japanese EFL Learners' Spoken Corpus

日吉 佑太
Yuta Hiyoshi

Abstract

This research surveys Japanese EFL learners' errors related to articles. In particular, the following are investigated; (1) how often the errors occur, (2) what types of errors are more frequent, and (3) how the frequency and the types of errors change over learners' development are investigated. The error-tagged component of NICT JLE Corpus, the transcription of Japanese EFL learners' speech in the English oral proficiency test, was used in the present research, and the errors were detected and analyzed by using AntConc, a concordancer software. As a result, the learners were found to make about 31 errors related with articles in every 1,000 words. Errors regarding indefinite articles (*i.e.*, *a/an*) occurred more frequently than those regarding the definite article (*i.e.*, *the*), and the omission of articles was more common than substitution and addition of them among the learners. It was also revealed that the group of the most proficient learners made approximately one-third as many errors as beginners did. In general, the percentage of errors related to the definite article and that of the addition errors increased as the learners' proficiency level rose.

1. Introduction

I have been teaching English to junior high, high school, and college students for more than nine years. In the course of my teaching, I often notice that my students omit, add, and substitute articles and that they complain about the difficulty of learning English articles. In my perception,

the difficulty has been increased by the lack of time spent on teaching articles, as it has been generally placed near the end of textbooks used for lessons.

Articles have been hypothesized to be one of the morphemes that are learned relatively early by both L1 and L2 learners. Brown [1973] argued that an article is the sixth of nine morphemes that L1 learners acquire. Similarly in L2, Dulay and Burt [1974] reported that English learners who spoke various L1 also followed the similar sequences of morpheme acquisition. Following the series of studies, Krashen [1977] formulated the Natural Order Hypothesis of L2 morpheme acquisition, with articles being, along with the auxiliary *be*, the second of four groups of morphemes that are acquired early by L2 learners.

Judging from my teaching experience, I have had doubts about the above proposed easiness of learning articles until I learned recently that there are research papers focusing on Japanese learner's acquisition of morphemes. These papers have indicated, the order of acquisition of Japanese learners is basically consistent with Krashen's theory, except for possessive *-s* and an article, *i.e.* the former was learned relatively early, and the latter was learned late [Hakuta 1976; Shirahata 1988; Tono 2002 as cited in Izumi, Uchimoto and Isahara 2005; Luk and Shirai 2009].

However, it is still unclear how learners' use of articles change over time. I am here concerned with how the frequency of errors change, and what types of errors Japanese EFL learners commit in regard to articles. I will begin this paper with the details of research on morpheme acquisition order of Japanese learners, followed by the statement of research questions. I will thereafter describe the methodology of this paper, along with the results and discussion. In the conclusion, I will summarize and raise further questions for study.

2. Literature Review

As mentioned in the introduction, a number of studies were conducted to investigate the acquisition order of morphemes in the 1970s. The research

by Brown [1973] was the starting point of a series of papers, although it was L1 learners, not L2 learners, that were surveyed in the paper. He reported that L1 learners manifested the following order of acquisition: (1) Present progressive *-ing*, (2) plural *-s*, (3) irregular past, (4) possessive *'s*, (5) copula *be*, (6) an article, (7) regular past *-ed*, (8) third person singular present *-s*, and (9) auxiliary *be*.

Subsequently, similar results were found for L2 English learners irrespective of their L1, such as Spanish and Chinese [Dulay and Burt 1974; Bailey, Madden and Krashen 1974]. However, ranking all morphemes has potential risk that they disguise the actual difference in accuracy between them [Ellis 2008, p. 83], meaning that actual distances of difficulty levels between the morphemes are ignored. To deal with this issue, Krashen [1977] finally put morphemes into groups that are learned at the same stage of development. This is known as the Natural Order which consists of four groups: (1) *-ing*, plural, and copula, (2) auxiliary and article, (3) irregular past, and (4) regular past, third person singular, possessive *-s*.

Some studies have been conducted to explore the acquisition order of Japanese learners of L2 English. They have shown that Japanese learners basically share similar sequences of acquisition with the Natural Order, although there were some differences. In particular, articles have been shown to be more difficult for these learners (*i.e.*, they are learned later) than it was predicted in the Natural Order. On the contrary, morphemes such as possessive *'s* were learned earlier [Hakuta 1976; Sihrahata 1988; Tono 2002 as cited in Izumi, Uchimoto and Isahara 2005; Luk and Shirai 2009]. For example, Luk and Shirai [2009] reviewed eight studies of morpheme acquisition by Japanese learners and found that all the studies except two indicated Japanese learners acquired articles relatively late compared with the Natural Order, where articles belong to the second group learned. These findings were consistent with the recent research that was carried out from the perspective of learner corpus research. Tono [2002 as cited in Izumi, Uchimoto and Isahara 2005] which surveyed the corpus named JEFFL (Japanese EFL Learner) Corpus that consists of written production derived for Japanese high school and college students. He

likewise found that the acquisition of articles was delayed in comparison with the Natural Order of acquisition while possessive 's was learned earlier.

The reason for these differences has been attributed to the influence of L1. As the Japanese language does not have any system equivalent to articles, a negative transfer occurs, forcing learners' difficulty. On the other hand, possessive 's has its L1 equivalent that allows learners to learn easily [Luk and Shirai 2009].

3. Research Questions

Following the discussion above, it seems to be obvious that articles impose a heavy burden on Japanese learners compared with other L1 speakers. It is thus advised that teachers offer substantial help to learners. For that reason, it is important to focus on the details of the acquisition of articles, *i.e.*, how often the errors occur, what errors are prevalent among Japanese EFL learners, or how those features change with the learners' development. However, there is yet little research that directly answers to these questions. Thus, the following research questions are set for the present study:

1. How often do Japanese EFL learners make errors regarding articles?
2. What types of errors do Japanese EFL learners tend to make?
3. How do the frequency of errors and the error types change as learners develop their English proficiency?

4. Methodology for the Present Study

The corpus used for this project is taken from the NICT JLE Corpus. This corpus consists of 1,281 samples and 1.2 million words which were elicited from Japanese EFL learners in an English oral proficiency test (ACTFL-ALC SST). One of the advantages of using this corpus is that 167 out of 1,281 transcripts are error-tagged and easy for analyses using a concordancer software. It is also beneficial that each learner's proficiency level is specified, from Level 1 (lowest) to Level 9 (highest). To maximize these advantages, I only used 167 error-tagged transcripts and divided them

into 8 sub-corpora according to the proficiency level. As there was only one transcript from Level 1 learner, Levels 1 and 2 were combined and formed a sub-corpus. Table 1 shows the summary of the sub-corpora.

The errors were detected by utilizing AntConc, a free concordancer software which was developed and distributed by Dr. Laurence Anthony. To answer Research Question 1, the number of errors were counted in each sub-corpora and subsequently frequency rates per 1,000 words were calculated.

For Research Question 2, the detected errors were further classified in two ways. First, the errors were categorized into indefinite article (*i.e.*, *a*, *an*) and definite article (*i.e.*, *the*) based on which article is obligatory grammatically. Secondly, the errors were categorized into three types: omission, substitution, and addition. In this project, they were defined as follows: (a) Omission: An error in which article was omitted when it was required; (b) Substitution: An error in which indefinite article (*a*, *an*) was used instead of the definite article (*the*), or in which the definite article was used instead of an indefinite article; (c) Addition: An error in which an article was unnecessarily used although it was not required.

For Research Question 3, the frequency data and the categorized data are compared between sub-corpora in order that the change of article use according to proficiency can be observed.

Table 1

Summary of eight Sub-Corpora

Level	Number of learners	Total words	Distribution
1 & 2	8	1,680	4.78%
3	28	12,506	16.76%
4	43	30,244	25.74%
5	30	26,547	17.96%
6	28	28,016	16.76%
7	16	18,562	9.58%
8	9	10,507	5.38%
9	5	6,682	2.99%
Total	167	134,699	–

5. Results

5.1 Research Question 1 (How often do Japanese EFL learners make errors regarding articles?)

The frequency of the errors regarding the use of articles is shown in Appendix A. The number of errors regarding articles adds up to 4,177 times in 134,699 words. This means the learners committed 31.01 errors in every 1,000 words.

5.2 Research Question 2 (What types of errors do Japanese EFL learners tend to make?)

The numbers of errors regarding indefinite articles (*a, an*) and definite articles (*the*) are shown in Appendix B. The learners made 2,044 times of mistakes regarding indefinite articles (60.1%), and 1,356 times regarding definite articles (39.9%). Moreover, the numbers of each type of errors and its percentage are shown in Appendix C. When we look at the whole subject corpus, 2,779 errors out of 4,177 were omission of articles, which was 66.5% of the total. Substitution occurred 503 times, and addition 895 times, which accounted for 12.0% and 21.4% each. Thus the most common type of error among Japanese EFL learners was omission of articles. The results show that Japanese learners make errors for

indefinite articles more often than definite articles, and they tend to omit articles more frequently than the other types of errors, substitution and addition.

5.3 Research Question 3 (How do the frequency of errors and the error types change as learners develop their English proficiency?) To begin with, for overall frequency of errors, Japanese learners tend to improve as their language proficiency develops (See Appendix A). The learners at Levels 1 and 2 made 50.00 errors per 1,000 words; for the most proficient group of learners, Level 9, the figure dropped down to 17.36. In other words, the most advanced learners showed error frequency that was about one-third that of the beginners.

Next, the percentage data in Appendix B was transformed into a chart (see Figure 1). Here we can find that more than 75% of the errors made by the beginners are with indefinite articles. Throughout the proficiency levels, however, errors regarding indefinite articles are more common than those of the definite article. Additionally, the ratio of proficient learners' errors with the definite article was higher than that of beginning learners.

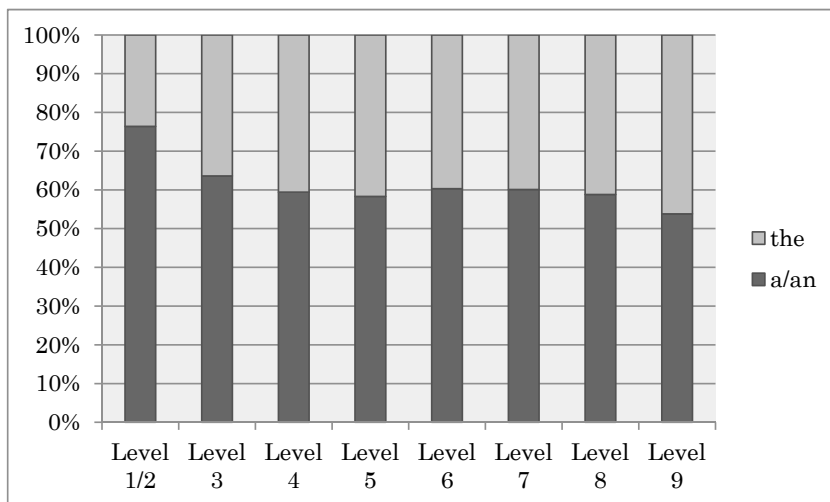


Figure 1. Rate of Errors Regarding Indefinite/Definite Articles

Finally, The percentage data in Appendix C were organized as a chart (see Figure 2). It is obvious that learners tend to make frequent omission of articles from the beginning of learning. As they improve their proficiency, learners move away from omission to addition. However, omission was still the most frequent type of error throughout the proficiency levels. Even for learners who belong to Level 9, the most proficient group, omission accounted for 58.6% of the total number of errors.

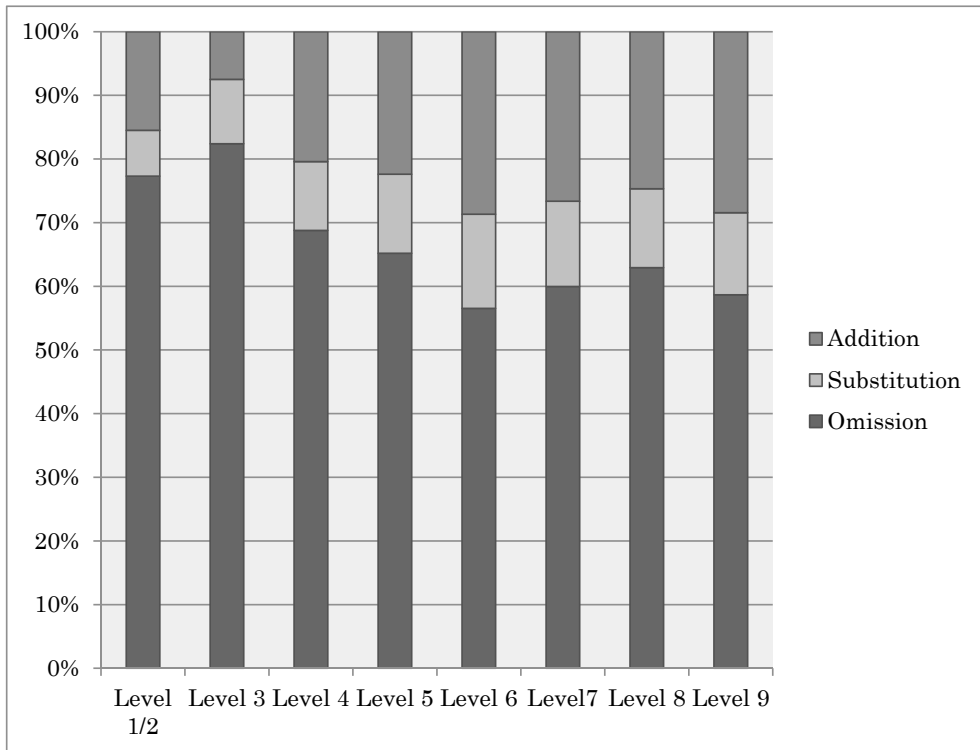


Figure 2. Rate of Error Types

6. Discussion

As we have seen in the previous section, Japanese EFL learners as a whole appear to make about 31 errors about articles in every 1,000 words according to the subject corpus. Although beginning learners make about 50 article errors in every 1,000 words, the rate should ultimately decrease to less than 20 per 1,000 words if they keep on learning.

According to the study, indefinite articles (*a, an*) are prone to errors more frequently than the definite article (*the*). This tendency seems to persist even after the learners reach high proficiency, although the rate of errors with the definite article errors slightly increased. It can be assumed that indefinite articles are more susceptible to Japanese EFL learners' errors because the proportion of obligatory contexts requiring indefinite and definite articles could have been consistent throughout the levels. However, we cannot directly conclude that, as the accuracy rate of each article type was not calculated in this paper.

With regard to the types of the errors, omission is the most frequent for all groups of learners. This result might have reflected the influence of negative transfer from Japanese to English. That is, the learners probably do not even notice that there should be articles, as the Japanese language does not include any system that is comparable to articles. Bryan [1984] is supportive of this view, mentioning that incorrect omission of articles by Japanese learners results is caused by the 'intrusion of L1' (p. 2). The fact that the articles are function words, and the misuse of them does not necessarily affect communication might also account for this tendency. Here again, the accuracy rate of articles was not counted either. The same problem also occurs here so that we cannot make a conclusion. We can nevertheless presume that beginning learners are likely to make more omission than substitution and addition errors, while advanced learners make less omission and more addition.

In class, teachers should inform learners of the importance of articles from the beginning stage of learning, against frequent disregard for them in textbooks and syllabi. The recent trends toward communication-oriented teaching make the situation more difficult, as communication can be carried

out without articles to some extent. However, brief but frequent references to when and how articles are used might help learners' understanding. Teachers furthermore need to focus on the indefinite article rather than the definite article if they want to improve the overall article use of their learners.

7. Conclusion

It is true that Japanese EFL learners frequently make errors regarding articles. However, such a difficult morpheme of English can be acquired over time gradually, not to say rapidly. It might be a good idea for teachers to show their learners such data derived from learner corpus research so that learners are properly encouraged and motivated. It was also found that the beginning learners make more errors with indefinite articles than with the definite article, and more omission than substitution or addition. As learners improve, the rate of errors with the definite article and addition type errors increase.

There are certainly some limitations in this study. The small sample size of the study may have influenced the results. In addition, the total number of occurrence of articles in learners' production should have been counted, which enables calculation of an accuracy rate.

Further research is required. Analyzing other types of corpora and comparing the results from various analyses should help us understand morpheme acquisition of Japanese learners. For instance, the use of morphemes in written learner corpus (e.g., the JEFFL Corpus) should be investigated to see whether there is a different tendency of errors between oral production and written production because writing allows more planning time for learners. If there is enough time for monitoring, learners can pay more attention, which leads to higher accuracy in turn (*i.e.*, style-shifting) [Tarone 1985]. It is thus predicted that the error frequency should fall. Research involving a corpus gained from younger learners should also benefit classroom teaching at junior-high and high school in Japan.

References

- Bailey, N., Madden, C., & Krashen, S.: Is there a “natural sequence” in adult second language learning?, *Language Learning*, 24(2), 1974, p.p.235-243
- Brown, R.: *A First Language: The Early Stage*, Harvard University Press, 1973, Cambridge, Mass.
- Bryant, W. H.: Typical errors in English made by Japanese ESL students, *JALT Journal*, 6, 1984, p.p.1-18
- Dulay, H., & Burt, M.: A new perspective on the creative construction process in child second language acquisition, *Language Learning*, 24, 1974, p.p.37-53
- Hakuta, K.: A case study of a Japanese child learning English as a second language, *Language Learning*, 26, 1976, p.p.321-351
- Izumi, E., Uchimoto, K., & Isahara, H.: Investigation into Japanese learners' acquisition order of major grammatical morphemes using error-tagged learner corpus, *Journal of Natural Language Processing*, 12, 2005, p.p.211-225
- Krashen, S.: Some issues relating to the Monitor Model, In H. Brown, C. Yorio, & R. Crymes (Eds.), *On TESOL '77*, TESOL, Washington, D.C., 1977, p.p.144-158
- Luk, Z. P., & Shirai, Y.: Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles, and possessive 's, *Language Learning*, 59(4), 2009, p.p.721-754
- Shirahata, T.: The learning of English grammatical morphemes by Japanese high school students, *JACET Bulletin*, 19, 1988, p.p.83-102
- Tarone, E.: Variability in interlanguage use: A study of style-shifting in morphology and syntax, *Language learning*, 35(3), 1985, p.p.373-403

Appendix A

Frequency of Article Errors

Level	Total words	Total errors	Errors per 1,000 words
1 & 2	1,680	84	50.00
3	12,506	562	44.94
4	30,244	1,196	39.55
5	26,547	896	33.75
6	28,016	686	24.49
7	18,562	394	21.23
8	10,507	243	23.13
9	6,682	116	17.36
Total	134,699	4,177	31.01

Appendix B

Number of Errors Regarding Indefinite/Definite Articles

Level	Indefinite (<i>a, an</i>) (%)	Definite (<i>the</i>) (%)
1 & 2	55 (76.4)	17 (23.6)
3	302 (63.6)	173 (36.4)
4	585 (59.4)	400 (40.6)
5	434 (58.3)	310 (41.7)
6	320 (60.3)	211 (39.7)
7	185 (60.1)	123 (39.9)
8	114 (58.8)	80 (41.2)
9	49 (53.8)	42 (46.2)
Total	2,044 (60.1)	1,356 (39.9)

Appendix C
Number of Error Types

Level	Total errors	Omission (%)	Substitution (%)	Addition (%)
1 & 2	84	65 (77.3)	6 (7.2)	13 (15.5)
3	562	463 (82.4)	57 (10.1)	42 (7.5)
4	1,196	822 (68.7)	130 (10.8)	244 (20.4)
5	896	584 (65.2)	111 (12.4)	201 (22.4)
6	686	388 (56.6)	101 (14.8)	197 (28.7)
7	394	236 (59.9)	53 (13.4)	105 (26.6)
8	243	153 (63.0)	30 (12.4)	60 (24.7)
9	116	68 (58.6)	15 (12.9)	33 (28.4)
Total	4,177	2,779 (66.5)	503 (12.0)	895 (21.4)