

# A Corpus-Based Analysis of Japanese EFL Learners' Linking Adverbial Use

Shusaku Nakayama

Meiji Gakuin University, Tokyo, Japan.

**How to cite this paper:** Shusaku Nakayama. (2021). A Corpus-Based Analysis of Japanese EFL Learners' Linking Adverbial Use. *The Educational Review, USA*, 5(6), 164-171.  
DOI: 10.26855/er.2021.06.002

**Received:** May 6, 2021

**Accepted:** May 29, 2021

**Published:** June 7, 2021

**Corresponding author:** Shusaku Nakayama, Meiji Gakuin University, Tokyo, Japan.

**Email:** shusaku-nakayama@outlook.jp

---

## Abstract

This research investigates to what extent Japanese EFL learners' linking adverbial (LA) in writing is identical to that of native speakers from two different perspectives: frequency of use and the distribution of LAs toward three different sentence positions, namely, initial, medial, and final positions. To this end, the International Corpus Network of Asian Learners' English is adopted. Frequency analysis identifies that Japanese learners overuse LAs relative to native speakers at a level that is statistically significant. Looking at results at an individual LA level, it is found at a statistically significant level that Japanese learners overuse "second(ly)," "for example," and "so" but underuse "then" and "rather". For results of LA placement analysis, Japanese learners statistically significantly overuse LAs in sentence-initial position but underuse LAs in sentence-medial/final positions. Research findings suggest that language teachers or teaching materials should provide learners with other equivalent expressions to LAs and expose learners to natural language patterns occurring in the real world in order to prevent Japanese learners from relying heavily on certain expressions to introduce their ideas or indicate relationships between two units of discourse and having a false impression that LAs rarely occur in sentence-medial/final positions.

## Keywords

Corpus-based analysis, Japanese EFL learners, linking adverbials, the ICNALE corpus

---

## 1. Introduction

In their paper, Halliday and Hasan (1976) considered cohesion to be an important textual factor to produce well-constructed texts. If one wants to be perceived as a good writer, it would be indispensable to become familiar with this factor. Linking adverbials (hereafter, LA), which are capable of indicating relationships between two units of discourse, play an important role in creating textual cohesion (Biber et al., 1999). Writers use such items appropriately, thereby allowing readers to expect what writers intend to express next. In other words, LAs can serve as a landmark to make readers follow the flow of discourse and are therefore among the most important items for learners to convey their own ideas consistently and make their arguments understood by readers.

According to my experience of teaching English at senior high school, it is often the case that Japanese non-native speakers (hereafter, JNNS) frequently use certain LAs such as "first," "second," or "in conclusion" in sentence-initial position to introduce their ideas. This trend may be partly attributed to their learning experience. Specifically, they might have been taught to use such expressions in order to get a high score in writing sections of Eiken tests, which are one of the most commonly used tests in Japan to gauge learners' English proficiency. Referring to books for preparing for Eiken tests, we can find many example answers using LAs in sentence-initial position as below. In this example,

“first,” “second,” and “in conclusion” are LAs:

I think more people will spend a lot of money on their pets. First, many people care about their pets' health (...). Second, many people notice this trend and there are more new pets services. (...). In conclusion, I believe the number of people who spend a lot of money on their pets will increase. (Obunsha, 2019, pp. 54-55)

In 2017, the Ministry of Education in Japan officially announced that they will be introducing a new entrance examination system that students can utilize scores on private English tests for university entrance exams, meaning that helping students obtain a good score on Eiken tests would largely contribute to their future. It can be, therefore, expected that junior/senior high schools in Japan will pay much more attention to the development of learners' writing proficiency. In reaction to this trend that is highly likely to take place, evaluating JNNSs' LA use will be of high value for people involved in English education to impart such important items in writing as LAs to learners effectively.

## 2. Corpus-based Analysis of JNNSs' LA Use

In this chapter, we review how corpus-based research has investigated JNNSs' LA use.

Narita et al. (2004) compared the use of LAs between advanced JNNSs and native speakers (hereafter, NS). The International Corpus of Learner English and the Louvain Corpus of Native English Essays were adopted as language data of JNNSs and NSs respectively. They found many interesting facts regarding the use of LAs by JNNSs. When compared against NSs, JNNSs overused LAs in sentence-initial position at a level that was statistically significant. In addition to JNNSs' preferences for certain LAs, they found a notable trend among JNNSs that they preferred using LAs for indicating enumeration or addition such as “first” or “moreover” or those for indicating apposition such as “for example” but avoided using LAs for indicating contrast such as “yet” or “instead”.

Ishikawa (2010) questioned whether or not writing conditions of written corpora used in previous studies on the LA use had been better controlled and thus investigated differences in the use of LAs between Asian non-native speakers including Japanese and Chinese learners and NSs within a corpus he developed, the Corpus of English Essays Written by Asian University Students (CEEAAUS). According to his findings, overall, Asian learners and NSs similarly used LAs in terms of variety and frequency. Frequency analysis at an individual LA level, however, exposed a number of LAs overused or underused by Asian learners. Furthermore, his analysis of LAs in terms of semantic categories identified that Asian learners statistically significantly overused LAs indicating addition such as “also” or “moreover”. Because in this corpus, learners are stratified by their levels of English proficiency, he further investigated the use of LAs between learners' proficiency levels. However, it seems that learners' levels of English proficiency were not closely correlated with the production of LAs.

## 3. Research Question

Ishikawa (2013) developed a new corpus representing Asian EFL learners' language use, namely, the International Corpus Network of Asian Learners' English (ICNALE), for the reason that although the CEEAAUS gained some attention from local scholars, it was clear that the size, variety of writers, and data control were far from satisfactory. (...) Therefore, the author reconsidered the entire data collection scheme and expanded the CEEAAUS to cover a greater diversity of writers in Asia and to enable it to be used as a more reliable database for international contrastive studies. (Ishikawa, 2013, p. 94)

Reevaluating JNNSs' LA use by using his new better-controlled corpus can potentially offer more reliable insights into JNNSs' use of LAs. What is better, participants in this corpus and the corpora used in the two previous studies were required to write their essays argumentatively. That is, investigating JNNSs' LA use within the ICNALE corpus can corroborate what previous studies have identified, allowing us to better understand how JNNSs use LAs in the style of argumentative writing.

In the ICNALE corpus, learners are stratified by CEFR-linked bands (B2+, B1-2, B1-1, A2), which allows the ICNALE corpus users to closely investigate characteristics of language use that are particular to a certain level of learners where researchers want to focus. According to Negishi et al. (2013), more than 80% of JNNSs fall within A1 and A2 levels on the CEFR. This indicates that language use produced by learners at these levels should be examined in great detail, rather than attempting to obtain a comprehensive picture of JNNSs' LA use, so that research findings can represent a typical pattern of JNNSs' LA use and help people involved in education apply them to the classroom. The current study, therefore, decided to put primary focus on LAs produced by A2 JNNS. The following two research questions were posed:

1. What LAs do A2 JNNSs prefer or avoid using relative to NSs?
2. In which sentence positions do A2 JNNSs tend to use LAs relative to NSs?

As for the first research question, the frequency of use of LAs was compared between A2 JNNSs and NSs. Following

the two previous studies (Ishikawa, 2010; Narita et al., 2004), frequency analysis took semantic categories into consideration. For the second research question, Biber et al. (1999) argued that LAs were not necessarily placed in sentence-initial position, and they could show up in sentence-medial/-final positions, too, which was an important characteristic of adverbials. This research thus investigated the distribution of LAs toward sentence positions as Narita et al. (2004) did.

## 4. Methodology

### 4.1 Corpus under Analysis

As mentioned above, this research focused on A2 JNNSs' LA use in writing. To this end, the ICNALE corpus, which consists of topic-controlled speeches and essays produced by NSs as well as college students in Asia, was adopted. Although several modules are available in this corpus, only the Written Essays 2.3 (Ishikawa, 2013) was adopted for the current study.

When learner language is compared with NS language, what we need to consider is that language to be analyzed is produced by what kind of NSs. According to Ishikawa (2013), language produced by native-speaking teachers (hereafter, NST) can be deemed to be a language model and was thus chosen as a benchmark language data. By comparing A2 JNNS language with NST language, this research sought to identify to what extent A2 JNNS' LA use derived from NS norms. The following table (Table 1) shows the corpus information.

**Table 1. Corpus Information**

	Number of participants	Number of tokens
NST	44	19,867
A2 JNNS	154	67,925

### 4.2 LAs under Investigation

For investigation, LAs to be analyzed should occur enough so that we can fairly infer and generalize A2 JNNSs' LA use from examples extracted. First, let us start with looking at the frequency of occurrence of LAs appearing in this corpus and their normalized frequencies in Table 2. WordSmith 7 (Scott, 2016) was asked to count how many times each LA occurred.

As can be seen from Table 2, several LAs occurred only once or twice, implying that analyzing such LAs may provide a biased interpretation. To avoid this possibility, this research consulted Ishikawa's (2010) research. Following his approach, I calculated the averages of the two normalized frequencies. The top 25 LAs were then chosen for investigation as shown in Table 3.

LAs analyzed in this study fell into four types of semantic category: "enumeration/addition," "apposition," "result/inference," and "contrast/ concession."

## 5. Results

### 5.1 Frequency Analysis

Table 4 illustrates how many times each LA was used by NSTs as well as A2 JNNSs respectively; numbers in parenthesis refer to the frequency of use of the LA in each group. In addition to raw frequencies of the LAs, to see whether or not differences in frequency of occurrence between the two groups were statistically significant, results of a log-likelihood test are also provided.

Before looking at results of the statistical test, it could be interesting to note that relatively high-frequency LAs in A2 JNNSs' essays, "first(ly)," "second(ly)," "also," "for example," and "so," each were produced more than 120 times, implying that these LAs might have been used by almost all A2 JNNSs since the number of them were 154 as shown in Table 1. Of course, we cannot disregard the possibility that one writer produced the LA several times.

For the results of a log-likelihood test, please note that they were interpreted with Bayes factors. That is, the differences whose Bayes factors were less than 2 including negative numbers will not be discussed since it means evidence is not worth discussing (Wilson, 2013).

A2 JNNSs used the LAs more frequently than NSTs with strong evidence against the null hypothesis. Looking at the data at a level of individual LAs, the differences in the use of "second(ly)," "for example," "so," "then," and "rather" were statistically significant. Among these, all the overused LAs, "second(ly)," "for example," and "so," were what almost all A2 JNNSs might have used in their essays as mentioned above. This fact may imply that these LAs can be what JNNSs strongly prefer using in their essays. When it comes to the semantic categories, a significant difference was indicated for all the categories except for contrast/concession.

**Table 2. List of LAs Observed in this Study**

LA	NST		A2 JNNS		Average of normalized frequencies
	Raw frequency	Normalized <sup>a</sup> frequency	Raw frequency	Normalized frequency	
Enumeration/addition					
also	67	337	241	355	346
first(ly)	15	76	123	181	129
second(ly)	15	76	127	187	132
in addition	8	40	37	54	47
furthermore	6	30	10	15	23
third(ly)	5	25	57	84	55
fourth(ly)	n/a	n/a	6	9	5
lastly	n/a	n/a	6	9	5
for another	n/a	n/a	1	1	1
in the first place	3	15	4	6	11
next	2	10	5	7	9
first of all	1	5	20	29	17
to begin with	1	5	n/a	n/a	3
moreover	1	5	28	41	23
similarly	n/a	n/a	1	1	1
what is more	n/a	n/a	1	1	1
Summation					
all in all	n/a	n/a	1	1	1
in conclusion	n/a	n/a	5	7	4
Apposition					
for example	9	45	120	177	111
for instance	2	10	28	41	26
that is	2	10	1	1	6
NST					
A2 JNNS					
LA	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency	Average of normalized frequencies
namely	2	10	2	3	7
in other words	n/a	n/a	7	10	5
that is to say	n/a	n/a	8	12	6
Result/inference					
then	43	216	36	53	135
so	21	106	423	623	365
therefore	14	70	70	103	87
consequently	2	10	2	3	7
thus	2	10	15	22	16
as a result	2	10	12	18	14
Contrast/concession					
however	24	121	93	137	129
though	9	45	15	22	34
anyway	8	40	6	9	25
yet	4	20	4	6	13
after all	3	15	8	12	14
conversely	2	10	1	1	6
still	2	10	n/a	n/a	5
instead	2	10	2	3	7
alternatively	1	5	n/a	n/a	3
besides	1	5	14	21	13
on the other hand	1	5	9	13	9
rather	1	5	2	3	4
in spite of	n/a	n/a	4	6	3
nevertheless	n/a	n/a	2	3	2
in contrast	n/a	n/a	1	1	1
NST					
A2 JNNS					
LA	Raw frequency	Normalized frequency	Raw frequency	Normalized frequency	Average of normalized frequencies
on the contrary	n/a	n/a	1	1	1
Transition					
by the way	n/a	n/a	2	3	2
incidentally	n/a	n/a	1	1	1

Note. <sup>a</sup> Normalized frequency was calculated by raw frequency/the size of population \*100,000.

**Table 3. LAs to Be Analyzed**

Semantic category	LA
Enumeration/addition	first, second, third, in the first place, first of all, next, in addition, besides furthermore, also, moreover
Summation	n/a
Apposition	for example, for instance
Result/inference	therefore, thus, so, then, as a result
Contrast/concession	On the other hand, though, anyway, however, yet, after all, rather
Transition	n/a

**Table 4. Frequency of Use of each LA and Results of a Log-likelihood Test**

LA	NST	A2 JNNS	Log-likelihood value	Bayes factor
Enumeration/addition				
first(ly)	15	123	+ 12.81*	1.43
second(ly)	15	127	+13.96*	2.57
third(ly)	5	57	+9.35*	-2.04
in the first place	3	4	-1.41	-9.98
first of all	1	20	+5.19*	-6.19
next	2	5	-0.13	-11.25
in addition	8	37	+0.64	-10.74
besides	1	14	+2.81	-8.57
furthermore	6	10	-1.79	-9.59
also	67	241	+0.14	-11.25
moreover	1	28	+8.64*	-2.74
Total	124	666	+23.59*	12.20
Apposition				
for example	9	120	+23.04*	11.66
LA	NST	A2 JNNS	Log-likelihood value	Bayes factor
for instance	2	28	+5.62*	-5.77
Total	11	148	+28.65*	17.27
Result/inference				
therefore	14	70	+1.83	-9.55
thus	2	15	+1.33	-10.06
so	21	423	+110.32*	98.94
then	43	36	-37.36*	25.98
as a result	2	12	+0.62	-10.76
Total	82	556	+39.55*	28.16
Contrast/concession				
on the other hand	1	9	+1.09	-10.29
though	9	15	-2.69	-8.69
anyway	8	6	-7.73*	-3.65
however	24	93	+0.31	-11.08
yet	4	4	-2.85	-8.53
after all	3	8	-0.13	-11.25
rather	11	2	-22.55*	11.17
Total	60	137	-6.42	-4.96
Total	277	1,507	+56.03*	44.65

Note. \*p<.05.

## 5.2 LA Placement Analysis

Next, we look at in which sentence positions NSTs and A2 JNNSs used the LAs. Table 5 summarizes the distribution of each LA toward three sentence positions, by which I mean, initial, medial, and final. In each column, the left and right sides distinguished by a slash mark respectively refer to the number of times A2 JNNSs and NSTs used the LA in



up, when compared against NSTs, A2 JNNSs overused the LAs in sentence-initial position whereas they underused them in sentence-medial/-final positions.

## 6. Discussion

In the frequency analysis, a log-likelihood test indicated statistical significance for many LAs. As mentioned above, however, this research interpreted this statistical test results with Bayes factors. In the end, it was found that only five LAs were worth discussing: “second(ly),” “for example,” “so,” “then,” and “rather.” The LAs, “then” and “rather,” were underused. “Then” and “rather” were respectively one of the underused LAs in Narita et al.’s (2004) study and Ishikawa’s (2010) study. If we consider one of the goals of language teaching is to bring learners’ language use close to native-like use as Granger suggested (2002), maybe language teachers can encourage JNNSs to use these LAs.

One interesting finding regarding the three overused LAs, “second(ly),” “for example,” “so,” was that they were produced so frequently that almost all the A2 JNNSs might have used them. Although this was behind one fact that their frequency of occurrence was very close to the number of A2 JNNSs, we cannot strongly argue this point anymore because this research cannot rule out the possibility that one author used these LAs several times. Overreliance on a certain LA among LAs belonging to the same semantic category can be explained by the *one-to-one principle* (Anderson, 1984), an idea that language learners give one form to one meaning in order to avoid getting confused by adding extra rules to their existing interlanguage system and establish an internal consistent interlanguage system. Ishikawa (2010) argued that Asian learners tended to rely on a relatively limited range of LAs. Hence, A2 JNNSs might have relied largely on “so” to indicate results or inferences, or on “for example” to show examples.

This principle can be somewhat applicable to the interpretation of A2 JNNSs’ overuse of the LAs. That is, A2 JNNSs might have relied heavily on the LAs to indicate relationships between two units of discourse even if it had been possible to convey almost the same propositions without LAs as exemplified by the following two sentences:

- (a) She does not like the rich. In other words, she hates you.
- (b) She does not like the rich. Put differently, she hates you.

To help learners to be able to extend their range of expressions for linking two units of discourse, language teachers should provide learners with other equivalent expressions to LAs.

The frequency analysis also identified several notable findings regarding the four semantic categories. First and most importantly, A2 JNNSs statistically significantly overused the LAs for indicating enumeration or additive, which reasonably accorded with what Narita et al. (2004) and Ishikawa (2010) found. This fact can allow us to strongly argue that JNNSs prefer using this type of LA in argumentative essays. Another notable finding was that among the four semantic categories in question, only the difference in the overall frequencies of the LAs for indicating contrasts and concessions was not statistically significant. Although “rather” was statistically significantly underused, it seems that A2 JNNSs were able to use this type of LA in the same way as NSTs.

In contrast to Biber et al.’s (1999) findings that the distribution of LAs to the three sentence positions was not quite different from each other in academic writing, the results of the LA placement analysis offered a strong indication that A2 JNNSs preferred using the LAs in sentence-initial position over the other positions, which was identical to Narita et al.’s (2004) findings. This research sees one of the causes of this trend in learners’ experience.

In 2005, Hoey proposed *lexical priming* theory, an idea that every word has inclinations toward particular linguistic features, and multiple encounters with them in multiple contexts lead us to acquire such tendencies. In this theory, he explains lexical priming concerning sentence positions: ‘every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations.’ (Hoey, 2005, p. 13). As mentioned earlier, LAs are used primarily in sentence-initial position in Eiken test-preparation books. It is possible that A2 JNNSs were primed for the LAs to occur in sentence-initial position and avoid the other sentence positions. Fortunately, he indicates a possibility that an individual’s existing priming can be replaced with new priming if one meets unknown usage and adopts it. To cause this replacement, for example, language teachers can possibly ask their students to read texts extensively or conduct lessons based on a data-driven learning approach. These can expose learners to a number of natural language patterns occurring in the real world, giving them a chance to notice how LAs are actually used and replace their existing negative priming that LAs can rarely occur in sentence-medial/-final positions with positive priming that LAs can occur in a variety of sentence positions.

## 7. Conclusion

This corpus-based study has examined A2 JNNSs’ LA use from two different perspectives: frequency of use and the distribution of LAs toward the three different sentence positions (i.e., initial, medial, and final). Considering A2 JNNSs’ overuse of the LAs positively, even the lowest-level learners in the corpus might not have had an aversion to using the

LAs. On the other hand, the fact that they rarely used the LAs in sentence-medial/-final positions relative to NSTs can be an issue to address soon. Hence, what language teachers should do is to explain that it is possible to use LAs in a variety of positions depending on the flow of passages.

Further research should include data produced by proficiency levels other than A2, namely, B1-1, B1-2, and B2. Comparing the use of LAs between proficiency levels would offer a comprehensive picture of JNNSs' LA use and allow language teachers to teach LAs according to their students' proficiency levels more effectively.

## References

- Anderson, R. W. (1984). The one to one principle of interlanguage construction. *Language Learning*, 34(4), 77-95. <https://doi.org/10.1111/j.1467-1770.1984.tb00353.x>.
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.
- Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. Longman.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Ishikawa, S. (2010). A corpus-based study on Asian learners' use of English linking adverbials. *Themes in Science and Technology Education*, 3(1-2), 139-157.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world*, 1 (pp. 91-118). Kobe University. [http://www.lib.kobe-u.ac.jp/handle\\_kernel/81006678](http://www.lib.kobe-u.ac.jp/handle_kernel/81006678).
- Ministry of Education, Culture, Sports, Science and Technology. (2017). *Basic guidelines for standardized tests*. [https://www.mext.go.jp/a\\_menu/koutou/koudai/detail/1420499.htm](https://www.mext.go.jp/a_menu/koutou/koudai/detail/1420499.htm).
- Mizumoto, A. and Takeuchi, O. (2008). Basics and considerations for reporting effect sizes in research papers. *Eigokyouiku kenkyu*, 31, 57-66. <http://hdl.handle.net/10112/5019>.
- Mizumoto, A. (n.d.). *Langtest.jp* (Computer program). <http://langtest.jp/>.
- Narita, M., Sato, C., and Sugiura, M. (2004). Connector usage in the English essay writing of Japanese EFL learners. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, 1171-1174.
- Negishi, M., Takada, T., and Tono, Y. (2013). A progress report on the development of the CEFR-J. In E. D. Galaczi & C. J. Weir (Eds.), *Exploring language frameworks: Proceedings of the ALTE Krakow Conference* (pp. 135-163). Cambridge University Press.
- Obunsha. (Eds.). (2019). *A workbook of the Eiken grade 2 to complete preparation in a week: A separate volume*. Obunsha.
- Scott, M. (2016). WordSmith Tools version 7 (Computer software). Lexical Analysis Software. [https://www.lexically.net/publications/citing\\_wordsmith.htm](https://www.lexically.net/publications/citing_wordsmith.htm).
- Wilson, A. (2013). Embracing Bayes factors for key item analysis in corpus linguistics. In M. Bieswanger & A. Koll-Stobbe (Eds.), *New approaches to the study of linguistic variability* (pp. 3-11). (Language Competence and Language Awareness in Europe; Vol. 4). Peter Lang.