Students' Use of Vocabulary and *Kanji* in Essays: Corpus Data-Based Measurement

Ai Sumirah Setiawati[1], Dwi Puspitosari[2], Danny Buyung Yudha Prasetya[3], Yulia Agustina[4]
Semarang State University[1,2,3,4]
ai.sumirah@mail.unnes.ac.id*

Abstract: This research was conducted to describe 1) the level of Vocabulary and Kanji that students tend to use in essays measured based on JLPT standards and 2) the suitability of the level of Vocabulary and Kanji used by students in Japanese essays with the learning level that has been taken, namely N4. These two measurements utilize a corpus of data that has only been used for reading-learning purposes. This research uses a qualitative approach, and the data source is 53 students with data collection techniques, namely assignments to write essays with predetermined themes and questionnaires. The data analyzed is the vocabulary and Kanji students use in their essays. This research shows that students use 68.8% of N5-level vocabulary and 61% of N5-level Kanji in their essays. The findings from assessing the use of vocabulary and Kanji differ from their current learning level, namely level N4, which the highest percentage at this level should indicate. Thus, the target of students' ability to use N4-level vocabulary and Kanji at their current learning level has not been achieved.

Keywords: Corpus data, JJLPT, *Kanji*, Japanese language, Vocabulary

INTRODUCTION

Writing skills in Japanese involve understanding and using the Japanese writing system, namely *Hiragana, Katakana*, and *Kanji* letters. These three letters are the main letters used in Japanese writing. Japanese can be written using Latin letters, but this will make the meaning difficult to understand because many words in Japanese sound the same but have different meanings. The essay's meaning and writing will be quickly understood using these three letters.

For example, the sentence "*Kamisama ame kudasai*" is not easy to understand, whether it is "God, give me rain" or "Lord, give me candy" if the sentence is written in Latin letters. Because the word ame can be interpreted

as 'rain' or 'candy.' It is different if the sentence is written using Japanese characters. The *Kanji* for the word "*ame*" (雨) in the first sentence means rain, while the Kanji in the second sentence (飴) means candy. In this way, what the sentence means can be easily understood.

(1) 神様、雨ください。

"*Kamisama, ame kudasai*"

"Lord, give me rain."

Or

(2) 神様、飴ください。

"*Kamisama, ame kudasai*"

"Lord, give me candy."

Apart from letters, understanding and use of vocabulary must be addressed and cannot be ignored. Mastery of Japanese vocabulary is crucial in writing because it will help writers convey information precisely and accurately and form coherent and cohesive sentences and paragraphs. Choosing the right words can significantly impact the understanding of the person reading our paper. Choosing the correct vocabulary will improve understanding, quality, attractiveness, positive impression, and writing skills. Therefore, choosing the correct vocabulary when writing in any language, especially Japanese, is imperative.

Vocabulary and letters, including *hiragana, katakana, and Kanji,* are one of the test items in the *Nihongo nouryoku shiken* or JLPT (Japanese Language Proficiency Test). These two materials in the JLPT questions become one in the *moji-goi* test (letters and vocabulary) and are classified into the knowledge aspect test. Other test items are *dokkai-bunpou* (reading and grammar) and *choukai* (listening).

Based on what has been explained, it can be understood that mastery of letters and vocabulary is essential for smooth communication, both oral and written. It is also vital to master because it is one of the test subjects that determines whether or not to pass the JLPT, which consists of five levels, namely N1 to N5. N1 is the highest level of Japanese language proficiency test that applies internationally. Possession of a JLPT pass certificate is also often used as a standard for recruitment for internship programs, apprenticeships, scholarships, and even when applying for work at Japanese companies in Indonesia. Apart from being a prerequisite, the abilities demonstrated by the JLPT results also influence income and position in the workplace. There is a significant relationship between the level of JLPT and the position of the participants. Participants with higher JLPT levels tend to hold managerial and professional positions in JRCSM (Japanese Related Companies in Malaysia), and participants with lower JLPT levels tend to hold other positions, which include general workers under the "others" category (Su et

al., 2020, p.964). Therefore, efforts to improve teaching by using media or methods and research related to letters and vocabulary are also significant in Japanese language education.

So far, research and studies in Japanese language education, especially in learning to write, have focused more on aspects of methods or media (Aneros et al., 2020; Lee et al., 2016; Nakano, 2014; Sakoda et al., 2016; Sutjiati & Rasiban, 2017). The other is the use of language and analysis of errors (Ahmed, 2022; Kanai, 2014; Kyōya, 2020; Peng, 2019; Sugimura, 2013; Wan, 2017; Wang, 2016, 2017; Yamamoto, 2016; Yoshida, 2011), and evaluation or assessment of writing learning outcomes(Setiawati, 2020).

Regarding evaluating learning outcomes or measuring students' abilities, so far, this is still often done by carrying out tests, both verbally and in writing. For example, the web version of the *Kanji* SPOT test and the *Kanji* proficiency diagnostic test are then compared with the results of student self-evaluation using the Ca-do statement format about *Kanji* (Kanō, 2015). SPOT (Simple Japanese Proficiency Test) is a test that places a blank space in the *Kanji* vocabulary part of a particular sentence, and participants are asked to select a *Kanji* character and insert it into the blank space while listening to the audio at a natural speed (Kanō, 2009). Research related to corpus data has been widely carried out but is still limited to the development or use of corpus data as a data source (Honda, 2019; Kobernyk, 2012; Sangawa, 2019; Sugimura, 2010; Takahara, 2017; Wan, 2017; Wang, 2016).

Based on the background explained above, this research discusses the level of Vocabulary and *Kanji* that students tend to use in writing Japanese essays, measured based on JLPT standards, and their suitability to the expected abilities and level of learning that has been learned by intermediate level students, namely at the level of JLPT N4. These two measurements utilize an existing corpus of data which so far has only been used for reading-learning purposes (Honda, 2019; Kawamura, 2012; Kitamura, 2013).

RESEARCH METHODOLOGY

This research used qualitative descriptive methods to analyze the data of students' vocabulary and *Kanji* usage ability measurement in writing an essay. Measurements were carried out using a data corpus, the Japanese Language Reading Tutorial System, which can be accessed on the page *https://chuta.cegloc.tsukuba.ac.jp/index_e.html*. The subjects in this research were 53 students of the 2021 year Japanese Language Education Study Program (23 men and 30 women). This research was applied to fifth-semester students at the beginning of the odd semester of 2023. These students have completed Japanese language learning at level N4 of the JLPT standard (Japanese Language Proficiency Test). An assignment, namely an essay writing

assignment, was used to obtain the data. A questionnaire was also used to find out the conditions and reasons behind the use of vocabulary and *Kanji* in the essay.

RESULT AND DISCUSSION

After the data was collected, it was measured using the Japanese Language Reading Tutorial System corpus data. In this study, the level-checking feature was used to identify the level of vocabulary and *Kanji* used by students in essays based on JLPT standards or internationally accepted Japanese language proficiency tests.

Results

After the data was processed and generated using the Japanese Language Reading Tutorial System corpus data, generated data was tabulated to see whether students used vocabulary and *Kanji* in their essays to see whether they were appropriate to their learning level. Based on the initial survey, of the 53 respondents, most had taken the JLPT N5, N4, and N3 exams. 17 respondents have passed N5, 17 people have passed N4, and 3 have passed N3. The remaining 16 people were those who had taken the JLPT but did not pass or had never taken the test at all. The ta and *Kanji* used by students in their essays were based on JLPT standards or internationally accepted Japanese language proficiency tests.

The level of vocabulary used by students in essays

The following are the results of statistical data regarding the number of uses and the percentage of vocabulary levels students use in writing essays. The essays written are in the form of opinions or arguments, which are called *ikenbun* in Japanese. Students wrote *ikenbun* with the theme "How students' opinions regarding JLPT certificates affect companies' high and low salary levels".

Table 1. JLPT Vocabulary Level Result

No	JLPT Level	Quantity of words	Percentage
1	N5 Vocabulary	3630	68,81%
2	N4 Vocabulary	425	8,06%
3	N3-N2 Vocabulary	730	13,84%
4	N1 Vocabulary	67	1,27%
5	Others	423	8,02%
	TOTAL	5275	100%

Table 1 shows how many vocabulary words or *goi* students use in writing *ikenbun* or opinion essays. Furthermore, the table shows the number

of conversions to the percentage of vocabulary used in student essays. The "other" column results from the assessment of vocabulary used by students but is not included in any JLPT level category.

Column N5 has the most significant percentage. The vocabulary students use is 3630 vocabularies, with a percentage of 68.81%. Then, the N3-N2 column was ranked second with 730 vocabularies and a usage percentage of 13.84%. Followed in the third rank with the percentage of vocabulary used by students totaling 425 vocabularies, amounting to 8.06% in column N4. The "other" column totaled 423 vocabularies with a percentage of 8.02%, and column N1 totaled 67 vocabularies with a 1.27% percentage.

Kanji levels students tend to use in essays

The following are the results of statistical data regarding the number of uses and the percentage of *Kanji* levels used by students in writing essays.

Table 2. JLPT *Kanji* Level Result

No	JLPT Level	Quantity of words	Percentage
1	N5 *Kanji*	729	61%
2	N4 *Kanji*	330	27,61%
3	N3-N2 *Kanji*	125	10,46%
4	N1 *Kanji*	10	0,83%
5	Others	1	0,01%
	TOTAL	1195	100%

Table 2 shows how many *Kanji* letters students write when writing *ikenbun* or opinion essays. The "Other" column results from the assessment of *Kanji* letters used by students but not included in the JLPT level category in the Base corpus data.

Column N5 has the most significant percentage regarding the number of *Kanji* used by students, amounting to 729 *Kanji*, a percentage of 61%. Then, column N4 was ranked second with 330 *Kanji* with a usage percentage of 27.6%. Followed in third place with the percentage of *Kanji* used by students totaling 125 *Kanji*, amounting to 10.46% in columns N3-N2. The N1 level column has 10 *Kanji* letters with a percentage of 0.83%, and then the "Other" category has 1 *Kanji* letter with a percentage of 0.1**%**.

Discussion

Table 1 and Table 2 data show that the vocabulary and *Kanji* used by students are still prominently concentrated at the N5 level, with a percentage of 68.81% for vocabulary and 61% for *Kanji.* Table 1 shows the student's use of vocabulary, where the second ranking is at the N3-N2 vocabulary level with a

percentage of 13.84% and then, followed by the third ranking at the N4 vocabulary level with a percentage of 8.06%. However, in contrast to the vocabulary level, the second ranking for the ability to use Kanji letters is considerably higher at the N4 Kanji level with a percentage of 27.6% and followed by the third ranking at the N3-N2 *Kanji* level with a percentage of 10.46%.

With the results, students' ability to use *Kanji* shows relatively more remarkable results than their ability to use vocabulary, indicated by the use of *Kanji* in the N4 and N3-N2 level range amounts to 38.06%. Furthermore, the data of both tables also indicates that the ability to understand Japanese vocabulary and *Kanji* reaches the N3-N2 level, although the data in both tables shows that the ability of students to understand Japanese vocabulary and *Kanji* is still prominent at the N5 level. This inequality occurs because the primary causal factor is that students who are not native Japanese, like those who study English or other languages, will experience difficulties. Many students think that learning English is difficult since it is not their mother tongue (Aminatun & Putri, 2023).

This research shows significant results regarding the description of the ability to use vocabulary and *Kanji* letters in students of the class of 2021 of the Semarang State University Japanese Language Education Study Program, who are in their third year. Then, provide an overview of the suitability between the learning targets and the reality in the field, namely that the ability to use vocabulary and *Kanji* when writing essays differs from the learning targets set. The Japanese language proficiency test called Nihongo Nouryoku Shiken, places more emphasis on knowledge and understanding of the language, so there are indications that students have a higher level of understanding of Japanese but have a lower ability to use Japanese. This hypothesis certainly requires further study to prove it. However, this initial hypothesis is supported by the material taught in the third year, which is equivalent to the JLPT N3 level, thus indicating that the material students study or have studied has exceeded the JLPT N5 and N4 levels.

When collecting data, students cannot use tools such as printed dictionaries, online dictionaries, or various artificial intelligence instruments to help them look for words and letters or translate. When students have difficulty expressing ideas in Japanese, the only thing they can do is look for synonyms. Furthermore, as a result, students end up using more vocabulary than they master, namely lower-level vocabulary. Therefore, N5-level vocabulary dominates in the essays of N4-level students.

The data in this study show that students' vocabulary abilities have not reached the target or are not per their learning level, namely the N4 level. This factor is caused by students' being accustomed to using Google Translate or other tools and applications when encountering words they do not know the equivalent of in Japanese. Of all respondents, 100% used the application with different frequencies ranging from those who said they always (33.3%), often (61.9%), and

rarely used it (4.8%). This is similar to research conducted by Moriya (2019), that when writing essays, 85% of students used Google Translate with details of 45% (30 people) answering often, 40% (27 people) answering sometimes, and 15% (7 people) answered that they rarely use it. Then, students usually need to remember the vocabulary they have searched for equivalents using either a regular dictionary or Google Translate. The author often finds students who need help remembering the meaning of the Japanese they write. Even after they reread it to get the meaning from the context of the reading, they need help to explain the meaning of the sentences they write. This factor causes students to be unable to reproduce the vocabulary for which they have searched for the Japanese meaning.

The frequency of using applications or other artificial intelligence to search for words or even translate sentences when writing essays makes students lack the motivation to master vocabulary by rote. Students can efficiently utilize existing tools that are practical and fast every time they need information or translation. However, the dependent use of the application weakens their abilities as a result. According to research by Ulfa (2023),the frequency of using Google Translate has little contribution to increasing translation ability. Google Translate only correlates around 0.6% with increasing translation skills (in this case, writing in a foreign language is the same as translating ideas in the mother tongue into the target language). That finding proves that using Google Translate or others is not very useful for improving language skills. On the contrary, it makes students neglect to apply their abilities.

Moreover, many students can only write a few sentences on the given theme. Many sentences were found to be more appropriate as answers to questions. When collecting data, students are given guided questions, which are expected to be developed into essays. Even though the theme of the essay given is very familiar to the daily life of students studying Japanese, it is not one of the topics usually written about as an exercise or writing assignment. It can also be predicted that this theme is different from a theme that students like. This condition also has a minor influence on students' fluency when writing. These non-linguistic factors influence writing, including interest in writing, motivation, discipline, intelligence, and reading (Amalisa et al., 2022).

Lastly, the ability to write *Kanji* is usually lower than the ability to read it. To write *Kanji*, learners need to remember the number and order of the strokes. In addition, many homophones in Japanese are represented by the same *kana* letters but have different meanings and can be differentiated by writing in *Kanji* (Otsuka & Murai, 2020). Even though Otsuka and Murai say that the meaning of homophone *Kanji* can be distinguished by how it is written in *Kanji*, there are many similar letters, which adds to the difficulty of using *Kanji*.

CONCLUSIONS

Based on the analysis results, it can be understood that students still need to be able to use vocabulary and *Kanji* letters in writing essays according to their targeted learning level, namely the JLPT N4 level. Based on the data, it is known that the majority of students still use N5-level vocabulary and *Kanji.* This reflects that Japanese language learning abilities targeted or expected at the N4 learning level have yet to be achieved. Students need a dictionary to write essays, so they can only use beginner-level vocabulary or *Kanji* in their writing.

This research also indicates a difference between students' ability to understand and use *Kanji.* Even though the material and knowledge students studied in the third year have reached Level N3, the student's ability to use vocabulary and *Kanji* letters has yet to show the same results.

This research can be an initial overview and a mitigation plan for improving policies and implementing education in institutions. Based on the findings and indications described above, follow-up plans need to carry out more detailed research regarding the factors that caused the student's ability to use vocabulary and *Kanji* not yet align with the learning target set and what efforts the institution can take to overcome this problem. Apart from that, in this study, new ability measurements were carried out on vocabulary and *Kanji*; as a continuation, it is also necessary to measure the level of students' use of grammar. Alternatively, further research is required regarding the correlation between students' ability to understand Japanese and their ability to use Japanese.

References

Ahmed, H. (2022). Analysis of Japanese Articles "Ni" and "De" Acquisition Process and Errors made by Arabic-speaking Egyptian Japanese Language Learners. *Majalat Wadi Alniyl Lildirasat Walbuhuth al'iinsaniat Waliajtimaeiat Waltarbawayhi*, *33*(33), 1–34.

Amalisa, T. R., Anggraini, E., & Nurchurifiani, I. (2022). The Correlation of Diction Mastery and Writing Interest With Students' Writing Ability. *International Journal of Language, Humanities, and Education*, *5*(1), 1–7.

Aminatun, D., & Putri, N. (2023). Applying Picture Series to Promote Students' Writing Achievement on Recount Text. *International Journal of Language, Humanities, and Education*, *6*(1), 53–60.

Aneros, N., Permatasari, I., Bachri, A. S., & Juangsih, J. (2020). Penerapan Media Audiovisual dalam Pembelajaran Menulis Karangan Bahasa Jepang

(Sakubun). *Journal of Japanese Language Education and Linguistics*, *4*(2), 132–146. https://doi.org/10.18196/jjlel.4239

Honda, Y. (2019). Kōpasu ni motodzuku "dokkai kihon goi 1 man-go" no sentei. *Nihongo Kyōiku*, *172*, 118–133.

Kanai, H. (2014). Sakubun no goyō bunseki no tame no nihonjin gakusei to gaikokujin ryūgakusei no konzai-gata jugyō. *Saitamadaigaku Nihongo Kyōiku Sentā Kiyō*, *8*, 25–34.

Kanō, C. (2009). Kanji goi no onsei shori nōryoku o saguru: Kanji supotto no kaihatsu to kadai. *Tsukubadaigaku Ryūgakusei Sentā Nihongo Kyōiku Ronshū*, *24*, 1–17.

Kanō, C. (2015). Nōryoku kijutsu bun ni yoru kanji-ryoku no jiko hyōka no kanōsei. *JSL Kanji Gakushū Kenkyūkai-Shi*, *7*, 86–93.

Kawamura, Y. (2012). Nihongo dokkai gakushū shisutemu "rīdingu-chūta" no torikumi. *U~ebumagajin "Ryūgaku Kōryū," 19*, 1–11.

Kitamura, T. (2013). Nihongo dokkai gakushū shien shisutemu "rīdingu-chūta." *Kōnandaigaku Kiyō. Chinōjōhō-Gaku-Hen*, *6*(2), 243–253.

Kobernyk, N. (2012). Roshiago Ukuraina-go bogo washa ni yoru Nihon gogakushūsha kōpasu goyō bunseki. An Analysis of Vocabulary and Grammar Errors, Usually Made by Russian and Ukrainian Learners of Japanese. *Language, Area and Culture Studies*, *18*, 173–180.

Kyōya, M. (2020). Nihongo shōchūkyū gakushū-sha no sakubun ni okeru goyō bunseki to shidō-hō no ichikōsatsu: Betonamu-go bogo washa no goyō rei o tōshi. *Sano Nihon Daigaku Tankidaigaku Kenkyū Kiyō*, *31*, 37–54.

Lee, B., Iinkai, S., Fujimura, I., & Ōshima, Y. (2016). *Chūgokujin Nihon go Gakushūsha no tame no Korokēshon Gakushū no Shidōhō ni kansuru Kisoteki Kenkyū: Dakubun dēta ni Motozuku "Meishi + o + Dōshi" no Korokēshon o Chūshin ni*. Nagoya University Graduate School.

Moriya, M. (2019). Daigakusei no kikai hon'yaku no riyō to nihongo kyōiku. *Jiyeogsan-Eob-Yeongu I*, *42*(11), 133–151.

Nakano, W. (2014). Mentātekisuto o Mochiita Sakubun Shidō-hō ni Kansuru Kenkyū. *Mimasaka Daigaku Kiyō*, *47*, 105–112.

Otsuka, S., & Murai, T. (2020). The multidimensionality of Japanese kanji abilities. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-59852-0

Peng, Y. (2019). Fukugō joshi "ni kanshite" to "ni taishite" no goyō bunseki. *Tsukuba Nihongo Kenkyū*, *23*, 15–25.

Sakoda, K., Konishi, M., Sasaki, A., Suga, W., & Hosoi, Y. (2016). Ta gengo bogo no Nihon gogakushūsha ōdan kōpasu. *Kokugo Kenkyuu Rebyuu*, *6*(3), 93–110.

Sangawa, K. H. (2019). *"Nihongo Hibogo Washa no Dokkai Kōpasu" kara Mieru Hikanji-ken Nihongo Gakushūsha no Jisho shiyō*. 351–358.

Setiawati, A. S. (2020). Kemampuan Menulis Mahsiswa PBJ UNNES Semester 3 Tahun 2019/2020: Penilaian BErdasarkan JF Standard. *CHI'E Jurnal Pendidikan Bahasa Jepang (Journal of Japanese Learning and Teaching)*, *8*(2), 85–91.

Su, Y. L., Manjet, K., & Tengku-Sepora, T. M. (2020). The Importance of Japanese Language Skills Proficiency among Malaysian Graduates in Japanese-based Companies. *Pertanika Journal of Social Sciences and Humanities*, *28*(2), 957–978.

Sugimura, Y. (2010). Kōpasu kara mita chūgokujin Nihon gogakushūsha no kakujoshi ni kansuru mondaiten ni tsuite. *Gengo Bunka Kenkyū Sōsho*, *9*, 137–152.

Sugimura, Y. (2013). Chūgokugo washa ni okeru nihongo no yū tai dōshi no jidōshi tadōshi ukemi no sentaku ni tsuite: Jin'i-teki jitai no baai. *Nihongo/ Nihongo Kyōiku Kenkyū*, *4*, 21–38.

Sutjiati, N., & Rasiban, L. M. (2017). Project-Based Learning Sebagai Upaya Meningkatkan Kreativitas Menulis Karangan Pendek Bahasa Jepang. *Barista: Jurnal Kajian Bahasa Dan …*, *4*. https://journal.stp-bandung.ac.id/index.php/barista/article/view/145

Takahara, M. (2017). "Chīsai" "ōkī" o chūshin to shita kihon keiyōshi no rui Yoshi kankei: Kōpasu o mochiita bunseki o tōshite. *Nihongo Kyōiku Hōhō Kenkyūkai-Shi*, *23*(2), 10–11.

Ulfa, M. (2023). The Correlation Between Students' Translation Ability and The Frequency of Using Google Translate of English Study Program Students at Riau University. *Madah: Jurnal Bahasa Dan Sastra*, *14*(1), 1–12. https://doi.org/10.31503/madah.v14i1.518

Wan, C. (2017). Chūgokugo o bogo to suru Nihon gogakushūsha no ukemi bun no goyō bunseki: Sakubun kōpasu o dēta to shite. *Kumamotodaigaku Shakai Bunka Kenkyū*, *15*, 105–121.

Wang, C. (2016). Chūgokugo o bogo to suru Nihon gogakushūsha no shiekibun no goyō bunseki: Sakubun kōpasu o dēta to shite. *Kumamotodaigaku Shakai Bunka Kenkyū*, *14*, 77–91.

Wang, C. (2017). Chūgokugo o bogo to suru Nihon gogakushūsha no kanō kōbun no goyō bunseki: Sakubun kōpasu o dēta to shite. *Ari Ake Kumamotodaigaku Gengo-Gaku Ronshū, 16, *, *16*, 67–90.

Yamamoto, Y. (2016). Nihongo bijinesu bunsho raitingu ni mirareru nihonjin daigakusei no goyō bunseki. *Kyōei Daigaku Kenkyū Ronshū*, *12*, 87–99.

Yoshida, M. (2011). Iken bun sanshutsu katei no hōryaku no bunseki: Sakubun hyōka ga takai gakushū-sha to hikui gakushū-sha no hikaku. *Akademikku Japanīzu Jānaru*, *3*, 21–32.