

Errors in L2-Chinese Orthography for L1-Japanese and L1-Korean Learners: A Corpus Study

Kun Yu, The Hong Kong University of Science and Technology, Hong Kong SAR

The Asian Conference on Education & International Development 2024
Official Conference Proceedings

Abstract

Both Korea and Japan are countries within the Sinosphere. Nowadays, the use of sinograms in Korean has become increasingly rare, while they are still retained in Japanese. While previous studies have investigated L2-Chinese character errors separately for L1-Japanese and L1-Korean students at different levels of L2-Chinese proficiency, a direct comparison is needed to find out L1-specific differences given their differential preference in the use of sinograms. This study focuses on L1-Japanese and L1-Korean students at the intermediate level of Chinese language proficiency, which is considered the stage where the writing ability of international students is relatively stable. Individual Chinese characters were randomly selected and grouped into pre-established frequency bands A and B, with 60 characters in each band. Sentences containing target characters were extracted from Chinese continuous interlanguage corpus. Errors were classified into two types: orthographic errors (writing the wrong character), and substitution errors (writing a different character than intended). A total of 5007 sentences were analyzed. Results show that L1-Japanese learners tend to make more substitution errors for low-frequency words, indicating a potential influence of confusability from their larger inventory of sinograms.

Keywords: Chinese Orthography, Word Frequency, L2 Acquisition, Cross-Linguistic Comparison

iafor

The International Academic Forum
www.iafor.org

Introduction

Chinese characters are symbols used to record and write the Chinese language, and are the oldest writing system still in use today. In history, Chinese characters were the only international communication system in East Asia, and were even the official written language of countries like Japan, Korea until the 20th century. Japan has established the "List of Commonly Used Chinese Characters," while Korea has established the "Basic Chinese Characters for Education." Today, we can still see the influence of Chinese characters in the written language and daily life of Japan and Korea.

Chinese characters are known as kanji in Japan and hanja in Korea. In Japan, instruction of 996 kanji begins in the 1st grade of elementary school and is officially completed by the end of the 6th grade. Another 949 characters are covered in grades 7-9, and collectively, these 1945 characters are labelled jooyoo kanji. In Korea, formal hanja instruction begins in the 7th grade and continues through the subsequent five years of secondary schooling. According to the Korean Ministry of Education, 900 characters are taught in middle school, and another 900 in high school (Brown, 1990).

Chinese Character Acquisition in Korean and Japanese

Because of this connection between Chinese and Korean/Japanese writing systems, it has been proposed that Chinese characters are easily learned by Korean/Japanese speakers by means of positive transfer. For example, China and Japan share some Chinese characters with identical shapes that have remained unchanged since their inclusion in the Japanese language system, despite changes in pronunciation and meaning. Second, while most Chinese characters used in Japan have undergone changes in pronunciation, some still have similar or nearly identical pronunciations to their Chinese counterparts. The only difference is whether they are pronounced with tones or not. Third, most characters shared by China and Japan have consistent meanings without any obvious difference (Hu, 2012).

However, literature has identified potential difficulties by Japanese/Korean speaker despite this connection. For example, a recent study by Liu (2020) compared Japanese and Korean students studying abroad, and found that both Japanese and Korean students made a large number of common or similar mistakes. Stroke deformation, stroke spatial position, and stroke deficiency are among the most common errors made by both groups of students.

At the same time, some potential group differences have been found as well. For example, it has been identified that Korean students make more stroke deformation and homophonic wrong characters than Japanese students. Liu (2020) further stated that the Chinese proficiency of Japanese students studying abroad seems to be higher than that of Korean students on 2 bases: first, the proportion of Japanese students who obtain Chinese language certificates is higher than that of Korean students; second, the average scores of Japanese students in HSK essays were higher than Korean students. This provided anecdotal evidence that Chinese character learning may be easier for Japanese students than Korean students.

One possible reason for the apparent higher difficulty of Chinese orthography acquisition by Korean speakers may be the increasingly rare usage of hanja in Korea. After the Korean language was reformed in the 1940s and 1950s, the use of Hanja was gradually reduced, and today, it is not commonly used in everyday communication. While Hanja is still taught in

schools as a part of Korean language education, its usage is limited to academic, legal, and historical contexts.

Despite empirical evidence, some scholars have maintained that learning characters may be simpler in Korea than in Japan on the basis of linguistic comparison. In Korea, characters have only one pronunciation, i.e. single syllable approximations to the original Chinese. On the other hand, in Japan, 60 percent of the jooyoo kanji have both 'Chinese' as well as native Japanese pronunciations, often several of each. Moreover, the Korean system is less complicated than the Japanese system, with Korean hanja only used to write nouns of Chinese or Sino-Korean origin (Brown, 1990). Given this divergence in predictions, more evidence is needed to establish the relative difficulty of Chinese character acquisition in Korean and Japanese speakers. This will be one of the aims of the current study.

Finally, Liu (2020) also investigated character errors made by students of different L2 proficiencies. It was found that the number of different types of errors decreases significantly as students' Chinese proficiency level increases. This gives some insight into the individual differences underlying Chinese character acquisition (i.e. target L2 proficiency being a factor). Since L2 proficiency may be an index of how familiar the learners are with Chinese characters, the current study investigates whether another proxy of familiarity – frequency of target character – may also be predictive of Chinese character acquisition in Japanese and Korean speakers. This is the second aim of the current study.

We conclude by explicitly stating our two research questions: (1) Are there differential character error rates of Chinese characters by L1-Korean and L1-Japanese speakers? (2) Are any differences further modulated by word frequency of the characters?

Methods

Chinese characters were selected and categorized into pre-established frequency bands A to B in descending order of word frequency (Ministry of Education, 2010), resulting in a total of 60 characters in each band. Example characters for each Band are listed in Table 1 below.

Band A	Band B
百 [hundred]	伴 [companion]
饱 [full]	辈 [generation]
比 [compare]	波 [wave]
布 [cloth]	灿 [bright]
草 [grass]	潮 [tide]

Table 1: Example Chinese characters from each frequency band

Learner data was extracted from the Chinese Continuous Interlanguage Corpus, which contains written data by foreign learners of Chinese in more than 110 different countries and regions that total more than 5 million words.

Sentences containing the target characters written by L1-Japanese and L1-Korean learners were analyzed. Data was filtered so that learners were at the intermediate level of Chinese language proficiency (equiv. HSK Level 4-5), which is considered the stage where the writing ability of international students is relatively stable (Zhu, 2021). This yielded a total of 5007 target sentences. Character errors were classified into two types: orthographic errors (writing a

character wrong), and substitution errors (writing a different character than intended). For each error type, accuracy was binary-coded (1 for correct, 0 for incorrect).

Results and Discussion

Figure 1 shows the results for orthographic errors. General observations reveal a global higher rate of error for low-frequency words than high-frequency words.

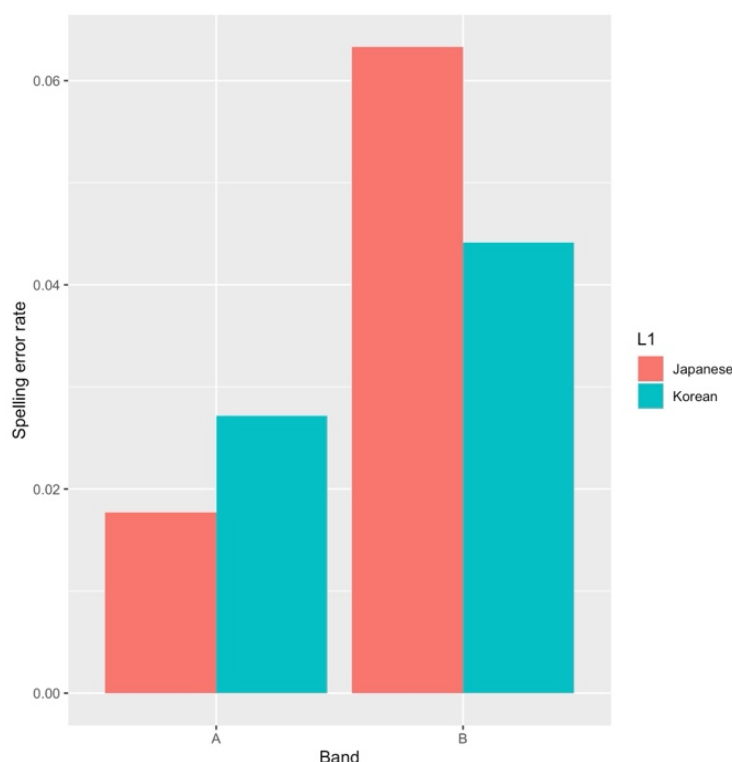


Figure 1: Proportion of orthographic errors by L1-Japanese and L1-Korean learners

A 2 (L1: Korean vs. Japanese) x 2 (Band: A vs. B) independent-measures ANOVA was performed on the orthographic error rate. We found a significant main effect of Band ($F(1,5003)=8.49$, $p=0.004$), confirming that both L1 groups exhibit higher error rates for low-frequency words than high-frequency words. This is expected because less exposure to target characters elicited more errors in general for both groups. However, neither the main effect of L1 nor the interaction between L1 and Frequency band reached significance, indicating no statistical evidence that Japanese and Korean speakers differed in terms of orthographic errors, or any further modulating effect of character frequency.

Figure 2 shows the results for substitution errors. General observations again reveal a global higher rate of error for low-frequency words than high-frequency words, and may also suggest a potential differential influence where Japanese speakers are disproportionately affected by low-frequency characters than Korean speakers.

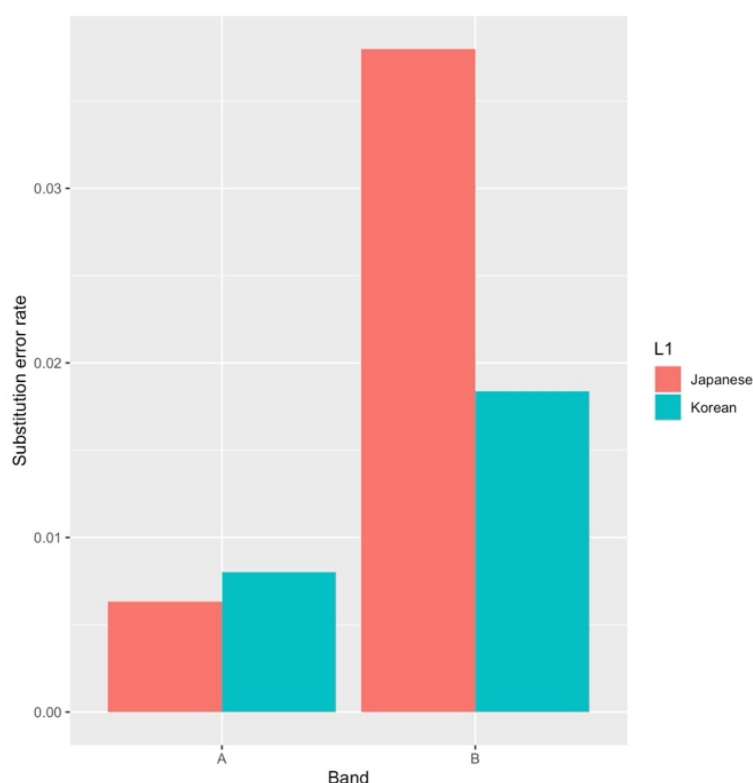


Figure 2: Proportion of substitution errors by L1-Japanese and L1-Korean learners

A 2 (L1: Korean vs. Japanese) x 2 (Band: A vs. B) independent-measures ANOVA was performed on the substitution error rate which partially confirmed the observations. We found a significant main effect of Band ($F(1,5003)=6.68, p=0.010$), confirming that both L1 groups exhibit higher substitution error rates for low-frequency words than high-frequency words. Crucially, the interaction between L1 and Band was marginally significant ($F(1,5003)=2.92, p=0.088$). Combined with the numerical trend of interaction for orthographic errors (albeit not reaching even marginal significance), this may provide some initial evidence that L1-Japanese learners indeed tend to make more substitution errors for low-frequency words than Korean speakers.

Conclusion

The current results have provided initial evidence of differential difficulty experienced by L1-Japanese and L1-Korean speakers in writing Chinese characters, a difficulty which is modulated by character frequency. In particular, Japanese speakers seem to be especially affected by low-frequency characters, exhibiting a disproportionately higher error rate than Korean speakers. This may indicate a potential influence from the larger inventory of sinograms in Japanese. Namely, the more frequent use of sinograms played a negative role for Japanese speakers. Thus, due to negative transfer of similarly-written L1 and L2, L1-Japanese learners may have shown higher error rate because of easily confusable sinograms between Chinese and Japanese (example: 廌 vs 𧈧, both meaning spacious).

References

- Brown, R. A. (1990). Chinese character education in Japan and South Korea. *Language and Communication*, 10(4), 299-309.
- Hu, X. (2012). Teaching and Learning Chinese Characters: The Advantages and Disadvantages of Japanese Students.
- Liu, T. (2020). 基于 BCC 语料库的日韩高级水平留学生汉字偏误对比研究 (Master's thesis, 河北大学).
- Zhu, Y. (2021). 中高级阶段汉字文化圈留学生的汉字书写偏误比较研究 (Master's thesis, 上海师范大学).