A Study of the Mutual Phonetic Resemblance Between Japanese and Chinese:
Quantification of the Difficulty of Phonetic Cross-Comprehension

Yuji Obataya, University of Geneva, Switzerland

The Asian Conference on Education 2019
Official Conference Proceedings

Abstract
This study investigated the mutual phonetic resemblance of Chinese ideograms between Japanese and Chinese by using a database of 1078 kanji (Chinese ideograms in Japanese) extracted from the two volumes of the Japanese grammar textbook (Second Edition, 2013 and 2015) used in the department of East Asian Studies at University of Geneva. The initial aim of this analysis is to help learners from non-kanji backgrounds to study simultaneously these two languages. Firstly, since most of kanji have multiple readings (on’yomi [Chinese reading] and kun’yomi [Japanese reading]), the rate of use of on’yomi in each 1078 kanji was calculated :59.72% [1], by taking into account the factor of frequency of all words (total 9233 words) who contain these kanji and are classified in the JLPT word list. Secondly, the basic phonetic resemblance has be figured out at 19.6% [2] according to the result of survey of twelve Chinese native speakers, who teach Chinese to Japanese people, nine of which passed N1 and three of which, N2 level of Japanese-Language Proficiency Test (JLPT). In comparison with the shape resemblance (71%), semantic resemblance (about 90 %) in the same database, analyzed in my former two studies (OBATAYA 2018a, 2018b), this relative low rate of phonetic similarity (12.4 %, judging from these values [1] and [2]) visualize significantly a gap between the “interdependence” of the shape and the meaning aspects and the “independence” of the phonetic aspect, and emphasize the importance of phonetic cross-comprehension for learners of these two East Asian languages.

Keywords: Asian languages, Japanese, Chinese, phonetic, logographic, cross-comprehension, simultaneous learning

---

I should add that when I wrote the abstract, the number of respondents to the survey was 9, but 3 joined after that. In the end, the survey in the current study was completed by 12 respondents.
Introduction

This study investigates the mutual phonetic resemblance in Chinese ideograms between Japanese and Chinese by using a database of 1,078 kanji (Chinese ideograms in Japanese) extracted from the two volumes of the Japanese grammar textbook, Minna no Nihongo (Second Edition, 2013 and 2015) used in the department of East Asian Studies at University of Geneva.

The initial aim of this analysis was to help the students of Japanese (or learners of Chinese, or both simultaneously) at University of Geneva to learn how to read Chinese characters and pronounce them correctly. It is hoped that this study may be useful not only for the students at University of Geneva but also for simultaneous Japanese and Chinese learners in non-kanji areas. For instance, it can be effective for beginner Chinese learners who have prior knowledge of Japanese.

For this purpose, my previous research studies were based on the characters required to pass each language’s proficiency test: the Japanese-Language Proficiency Test (JLPT) for Japanese and the Hanyu Shuiping Kaoshi (HSK) Chinese Proficiency Test for Chinese. Figure 1 illustrates the approximate correspondence of the proficiency levels between the JLPT and the HSK.

![Figure 1: Correlative table between JLPT and HSK proficiency levels.](image)

The increasing number of candidates for both the JLPT and the HSK reflects the global interest in these two East Asian languages; in 2013, a total of 571,075 examinees took the JLPT, while in 2017, that number rose to 887,380. For the HSK, a total of 189,691 examinees took the test in 2013, while in 2017, 470,807 took it.

---

2 For the purpose of these studies, I took the HSK exams and passed with a proficiency level of HSK 6—the highest level—in 2014.

3 In comparison, the number of people who participated in the DELF / DALF exams was 389,120 in 2015.
The increased interest has encouraged approximately 20% of students at department of East Asian Studies at University of Geneva to select them as their two majors required to complete their bachelor’s degree, even though a survey\(^5\) has revealed that learners from non-\textit{kanji} backgrounds find learning Japanese and Chinese simultaneously to be a heavy burden.

### The Difficulties in Learning Japanese and Chinese Simultaneously and the Introduction of Chinese Characters to the Japanese Writing System

One of the difficulties that students face in learning these two languages at the same time is the complexity of the Chinese ideograms used in both Japanese and Chinese. For example, the simplified form of Chinese ideograms used in Japanese—called \textit{kanji}—sometimes differ in shape, meaning, and pronunciation from the simplified Chinese characters as used in the People’s Republic of China (Yoshida, 2014, p.19).\(^6\)

Contemporary Japanese has three main graphic systems: \textit{kanji}, Chinese ideographs used in Japanese, and two syllabaries (hiragana and katakana).\(^7\) In order to better understand the relationship between Japanese \textit{kanji} and Chinese characters, it is necessary to review the history of the three graphic systems of Japanese.

The Japanese, having no writing system of their own, imported Chinese characters, or ‘\textit{kanji}’, as early as the 4\textsuperscript{th} or 5\textsuperscript{th} century. Some three centuries later, a cursive and simplified form of \textit{kanji} appeared, chosen for its phonetic value: the \textit{manyōgana}, the name of which was derived

\[^4\] The JLPT’s numbers are quoted from the JLPT’s homepage (https://www.jlpt.jp/e/statistics/archive.html). The HSK’s numbers were taught by the Confucius Institute of Geneva University (2018.07.11).
\[^6\] This paragraph is taken from Obataya Y. (2018a), p.2.
\[^7\] This does not take into account the romaji, or Latin alphabet.
from the title of an anthology of poetry written using this simplified script during the Nara era (710–794), ‘Collection of Ten Thousand Leaves’, or Manyōshū. Manyōgana would give rise to the 9th century hiragana syllabary, as well as to the katakana syllabary, which was also created in the 9th century. Its current form, however, became fixed in the 12th century. Figures 3 and 4 show examples of the three writing systems as well as their proportion of use in Japanese.

![Diagram of three writing systems: Kanji, Hiragana, Katakana](image.png)

**Figure 3:** Three writing systems in Japan

![Proportion of use of writing systems](image.png)

**Figure 4:** Approximate proportion of the use of the three systems in contemporary Japanese

Japanese continued to borrow Chinese characters according to their meaning; however, since the pronunciation differed between Japanese and Chinese, Japanese kanji could be read by Japanese speakers in the Japanese way (‘kun’yomi’, henceforth kun-reading) as well as in the (original) Chinese way (‘on’yomi’, henceforth on-reading). Furthermore, a kanji sometimes has two, or three or more types of on-reading (go-on-reading, kan-on-reading, and tō-on-reading, among others9), because kanji’s phonetics have evolved in various ways over the centuries in the different eras and regions from China.10

---

8 Document used for the annual “open campus” session at Geneva University for high school student
9 These different on-reading are indicated in my database for the analysis of the graphic field (Otabata, 2018a, 9th note).
10 ‘Phonetic value of kanji’ (...) Go-on is considered to be the oldest kanji sound from China. Go is the name of the old State of Wu, which is the current area of Jiangsu Province (Nanjing). It broadly includes the Yangtze River area in South China and refers to the sounds that were used in that region. Japan traded in the 5th and 6th centuries, before the Nara era, with the
In addition, a simplification of the Chinese characters of the People’s Republic of China took place in the 1950s, following the Japanese government’s simplification of kanji in 1926. These simplifications resulted in the lists of Tôyô kanji in 1946 and Jôyô kanji in 1981.

These different points may cause great confusion for simultaneous learners, especially complete beginners, and are likely to disturb their learning of ideograms. Therefore, this study aims to support such students’ effective learning by building a database that identifies the commonalities and differences between kanji and Chinese characters. The database comprises all 1,078 kanji that appear in the two volumes of the Japanese grammar textbook that the department of East Asian Studies at University of Geneva has adopted for instruction.
Previous Studies on the Graphic and Semantic Resemblance Between Kanji and Chinese Characters

The analysis of the graphic aspect and the semantic aspect of this database has been effectuated in Obataya (2018a) and Obataya (2018b), respectively.

Table 1: The number of kanji covered in the textbooks (I and II)\textsuperscript{11}

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
JLPT N5 & 79 & 100\% \\
JLPT N4 & 166 & 100\% \\
JLPT N3 & 321 & 90\% \\
JLPT N2 & 241 & 60\% \\
JLPT N1 & 248 & 20\% \\
More & 23 & \\
More & 1078 & \\
\hline
HSK1 & 120 & 70\% \\
HSK2 & 122 & 73\% \\
HSK3 & 176 & 60\% \\
HSK4 & 243 & \\
More & 433 & \\
\hline
\end{tabular}
\end{center}

The first study demonstrated that the textbook not only covers 100\% of the kanji of the JLPT N5 (the lowest level) and the N4 but also approximately 70\% of the Chinese characters of the HSK1 (the lowest level) and HSK2, as well as 60\% of the HSK3.

In addition, the results of an analysis of the degree of mutual similarity in the list indicated that 71\% of Chinese characters are identical (or have only a slight difference) in the two languages.

After this analysis of graphical resemblance, we effectuated the analysis of semantic resemblance, with the results revealing that 89\% of the characters are identical or have only a slight difference.

The current study aims to quantify this final field—namely, phonetic resemblance—by using the same database.

Previous Studies on Phonetic Resemblance

Several studies exist on the phonetic resemblance between Japanese and Chinese words and characters: Kayamoto (1995), Matsushita (2009), and Gi (2017). While the current study highlights ‘parallel characters’ in terms of phonetic resemblance, Matsushita (2009)\textsuperscript{12} and Gi (2017) focused mainly on ‘parallel words’.

\textsuperscript{11} This is the ‘Table 1’ in Obataya (2018a) that was modified and some images were added.

\textsuperscript{12} This study on the phonetic resemblance of Matsushita (2009) is based on the data provided by Kayamoto (1995).
Analytical Approach

In the current study, the scores calculated by the multiplication of two values are considered to be the real phonetic resemblance of Chinese ideograms utilised in both Japanese and Chinese. These two values are (1) the frequency of a kanji in on-reading and (2) the rate of the phonetic resemblance of a kanji between on-reading and the pronunciation in Chinese. In order to calculate the latter value, I applied the method of Kayamoto (1995) by distributing questionnaires concerning the 1,078 kanji of my database to Chinese native speakers who have considerable knowledge of the Japanese language and calculating the mean values. In our study, we did not use audio recordings of Japanese or Chinese pronunciation (unlike Gi’s study, which used audio recordings of both languages’ pronunciation) by pointing to a website that contains the sound sources of Japanese pronunciation.

Calculation of the Frequency of On-Reading Usage

No previous study has taken into consideration the frequency of the on-reading usage of a kanji, which is one of the most important criteria in this work. A kanji is pronounced in various ways, including on-reading, kun-reading, and other ways of reading, such as an idiomatic way of pronunciation. In this research, the values that signify the usages of on-reading in the target kanji data are identified.

First, the usage of each kanji in words was investigated. Here, 9,233 words in the list of the previous JLPT\textsuperscript{13} were categorised into ‘on-reading’, ‘kun-reading’, and ‘other’, according to their pronunciation.

In order to reflect the frequency of on-reading usage according to the levels of difficulty designated in JLPT, all five levels were counted differently. For example, a kanji in N1, the highest proficiency level in JLPT, was calculated as one point, whereas a kanji in N5, the easiest in JLPT level, was calculated as five points. It was considered that the easier a kanji is, the more frequently it is used.

\textsuperscript{13} In the previous JLPT (= Old Japanese Language Proficiency Test Levels 1 to 4), before the renewal of the current JLPT (= Levels N1 to N5) in 2010, ‘Test Content Specifications’ was published for helping the candidates. This book contained a word list, a kanji list, and grammar list, respectively classified in four levels. The current JLPT does not publish such a book, so for comparing the new JLPT levels to current work, I reference the Jisho.org website, and Jonathan Waller’s JLPT Resources page. According to Waller, New JLPT N1 is equivalent to the old JLPT 1; JLPT N2, to the old JLPT2; JLPT N3, to halfway between the old JLPT 2 and JLPT 3; JLPT N4, to the old JLPT 3; JLPT N5, to the old JLPT4.
Figure 6: Example of calculation of the on-reading usage

For example, with a *kanji*, 社, there are 12 words that use this *kanji* with the /sha/ pronunciation. All of the scores for these words were counted according to the levels of difficulty and added to be the total score. There is only one word that uses the *kun*-reading: /yashiro/, categorised as N1. Therefore, the score of *kun*-reading for this *kanji* was 1. This means that the frequency of on-reading for this *kanji* was designated as 97%.

The *kanji* that tend to be written in hiragana at present were not counted. Instead, they were listed in the section of Figure 7 emphasised in red. For example, a *kanji*, 何, was not included in the JLPT list with the pronunciation of /ka/. Therefore, ‘Not on the JLPT list for on-reading’ was indicated (see Figure 7).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>JP</td>
<td><em>Kun</em>-reading</td>
<td>JLPT Word list</td>
<td>JLPT Word list (levels)</td>
<td>Calculation of on-reading (L2 levels)</td>
<td>Percentage of on-reading (%)</td>
</tr>
<tr>
<td>社</td>
<td>やしろ</td>
<td>97</td>
<td>シャ</td>
<td>[209x520]sha</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>kaisha</td>
<td>会社</td>
<td>かしゃ</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>shakai</td>
<td>社会</td>
<td>しゃかい</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>shachō</td>
<td>社長</td>
<td>しゃちょう</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>jingai (jinsha)</td>
<td>神社</td>
<td>じんじゃ</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>shinbunsha</td>
<td>新聞社</td>
<td>しんぶんしゃ</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>shakaihakusho</td>
<td>社会科学</td>
<td>しゃかいかがく</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>hasetsu</td>
<td>社員</td>
<td>しゃいん</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>shōsha</td>
<td>職社</td>
<td>しょうしゃ</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>nyūsha</td>
<td>新社</td>
<td>しんしゃ</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>shakó</td>
<td>社報</td>
<td>しゃきょう</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>shataku</td>
<td>社宅</td>
<td>しゃたく</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>shussha</td>
<td>出社</td>
<td>しゅっしゃ</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>yashiro</td>
<td>社</td>
<td>やしろ</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7: Cases of the *kanji* that tend to be written in hiragana at present
Questionnaires Concerning Phonetic Resemblance Between Japanese and Chinese

Another important criterion for this study is the value of a rather psychological aspect of the phonetic resemblance of each kanji, as evaluated by the questionnaires (see Figure 8).

The survey method introduced by Kayamoto (1995) was applied in this study.\(^\text{14}\) However, there are several differences in the data collection methods between Kayamoto (1995) and this research.

Firstly, the number of points for the rating scale was reduced to five in order to allow the respondents to easily answer the questions, whereas a seven-point scale was employed in Kayamoto (1995). Another advantage of the application of a five-point scale was that it enabled us to calculate the rate of phonetic resemblance quickly.

Secondly, most of the survey respondents (a total of 12 people) in the current work had not only acquired N1 of JLPT but had also been teaching Chinese to Japanese students,\(^\text{15}\) although in Kayamoto (1995), the 11 respondents were postgraduate students and research students (kenkyû-sei) whose mother tongue was Chinese.

Thirdly, the target databases of kanji differ. In Kayamoto (1995), the database included 996 kanji that had been previously designated for instruction by the Ministry of Education from 1977 to 1989, whereas we analysed 1,078 kanji from the two volumes of the Japanese grammar textbook, *Minna no Nihongo*, the first volume of which was printed in 2013 and the second in 2015.

My database contains 72% of the kanji analysed in Kayamoto (1995) and 74% of the kanji currently designated for instruction since 1989 (a total of 1,006 kanji).

\(^{14}\) Like Kayamoto’s research, because the four tones in Chinese are not clearly differentiated in Japanese, respondents were formerly informed not to care too much about them before evaluating each case.

\(^{15}\) All 12 respondents were Chinese native speakers who are teaching Chinese to Japanese students; 9 respondents have acquired JLPT N1 and 3 have passed N2.
Figure 8: The first page of the questionnaire (total of six pages)

Figure 9: Example of the plural on-reading or Chinese pronunciations
Evaluation of 1,095 Pairs of Kanji for Phonetic Resemblance

The number of characters in the database is 1,078, but the number evaluated in this questionnaire was 1,095. This is because, in the case of multiple on-readings or Chinese pronunciations existing in one kanji, respondents evaluated them separately. As shown in Figure 9, a kanji, 人, has two ways of on-reading, /jin/ and /nin/, and one Chinese pronunciation, /ren2/. In this case, two comparative evaluations between (1) /jin/ and /ren2/ and (2) /nin/ and /ren2/ were conducted. Another example is 行, which has two ways of on-reading, /kou/ and /gyou/, and two Chinese pronunciations, /hang2/ and /xing2/. In this case, the respondents must have assessed the phonetic similarity in four pairs independently.

Findings of the Two Statistical Analyses

(1) Findings from the Frequency of the On-Reading Usage

- The rate of the frequency was 59.7%.
- 294 kanji (27% of the total number of the data) only have on-reading pronunciations.
- 152 kanji only have kun-reading pronunciations.

(2) Findings from the Survey on Phonetic Resemblance by Means of Questionnaires

- The mean value for the phonetic resemblance from the survey was 19.6%. Compared with the average of the survey outcome from Kayamoto (1995), the result was slightly lower (Kayamoto’s mean value was 34%, 2.38/716).
- Very few kanji had complete correspondence between Japanese and Chinese (only n = 9 kanji: 医, 伊, 信, 衣, 敷, 因, 他, 愛, and 膚). The rate of the ones with more than 90% resemblance was only 4% (n = 42)17.
- Most of the kanji with perfect correspondence were first and fourth tones.
- Nine per cent of kanji (n = 208) were considered to have no resemblance between Japanese and Chinese. Furthermore, the ones whose resemblance was less than 10% comprised more than half of the data (n = 572).

---

16 As a result, the average of the overall rating was 2.38 (SD: 1.32) (Gi, 2017, p.63). The average of 2.38 (/7) is 34% when converted to a percentage.
17 These 42 kanji are 医伊信衣敷因他愛膚.
Chart 1 shows the mean and standard deviation.

![Chart 1](image1.png)

**Chart 1:** Distribution of the mean value of the questionnaires for the phonetic resemblance of 1078 kanji and standard deviation

Chart 2 shows the distribution of the number of kanji in the range of the mean value, presenting the number of kanji on the vertical axis and the range of the mean value on the horizontal axis.

![Chart 2](image2.png)

**Chart 2:** Distribution of the number of averages

As illustrated in this chart, the majority of dissimilarity is distributed on the left and the similarity on the right part of the chart.

---

18 The kanji at the bottom is representative of each stage.
Findings from the Total Scores by Multiplying the Values of (1) and (2)

- The score obtained by multiplying the values from the analysis of (1) and (2) was 12.4%.
- Only five kanji (医, 伊, 信, 愛, and 膚) have 100% frequency of usage in on-reading as well as 100% phonetic similarity.

The distribution graph (Chart 3) indicates a high rate of dissimilarity between Japanese and Chinese.

![Chart 3: 'Real' phonetic resemblance degree: (1) × (2)\(^{19}\)](chart)

Chart 4, which clarifies this dissimilarity, shows the distribution of the number of kanji in the range of the mean value, presenting the number of kanji on the vertical axis and the range of the mean value on the horizontal axis. The number of kanji in the range of 0–10% is 697, whereas the number of kanji in the range of 0–50% are the majority (95%). This means that most of the pronunciations of kanji used in both languages differ drastically.

![Chart 4: 'Real' phonetic resemblance degree: the number of averages](chart)

\(^{19}\)The kanji at the bottom is representative of each stage.
Chart 5: The number of averages from 50% to 100%

Chart 5 shows a zoomed-in view, from 50% to 100%, of the range of the mean values in the Chart 4, showing the minority group of kanji that phonetically resemble each other.
Figure 10 shows the final results for the three fields.
Conclusion

To conclude, after quantifying the phonetic resemblance of 1,078 kanji between Japanese and Chinese, it was found that this resemblance was very low compared to the high resemblance in shape and meaning. Such clarification of the three values of shape, semantic, and phonetic resemblance will allow us to explain the ‘kanji paradox’, whose notion was applied in one of our previous works to explain that the simultaneous acquisition of Chinese and Japanese is, in fact, difficult, despite the image of easy cross-comprehension due to the common usage of Chinese ideograms (Berger, C., & Obataya, Y. (2014), pp.162-163).

Despite the findings of this study, it is not necessary to fully deny the attempt of simultaneous learning or cross-comprehension of Japanese and Chinese. In fact, there are considerable advantages and incentives in the resemblance of forms and meanings in the common kanji for motivating learners to study two languages simultaneously. By making students aware of the imbalanced rates of resemblance among the three components, sounds, forms, and meanings at the beginner’s stage of learning, it may be possible to make their studying process more effective.

This means that it is important for students to develop special learning strategies to study kanji for acquiring listening and speaking skills, which should differ from the strategies for developing reading and writing skills in learning Chinese and Japanese simultaneously.

Acknowledgement

I would like to express my special gratitude to Pierre François Souyri, Professor Emeritus at the University of Geneva, for allowing me to take charge of all the modules of kanji. Without his help, it would have been impossible for me to collect the data for this study. I am also immensely grateful to the teachers of 天天中文 (https://ttcn.co.jp/) for their cooperation with the questionnaire, Ayako Shibata for her comments on an earlier version of the manuscript for improvement, Ping Zeller for her corrections to the Chinese part of the questionnaire, and Ikuo Obataya (Quantum Design Japan) for his help with measurements and analysis. Finally, I would like to thank administrators of Jisho.org and Jonathan Waller, creator of the JLPT resources pages for deciding the JLPT levels in my database.
References


References for data

3A Network. (2013). Minna no Nihongo, Book 1 Translation & Grammatical Notes, French version (2nd ed.). Tokyo, Japan: 3A Network.


Contact email: Yuji.Obataya@unige.ch