An Implementation of Japanese Characters Cartography as a Learning Tool

Antoine Bossard[*]

Abstract

Foreign learners of the Japanese language invariably have to face the difficult task of characters memorisation. Effectively, due to the huge number of characters involved, memorisation is a critical obstacle on the path towards Japanese reading and writing proficiency. We have recently proposed a novel Japanese characters memorisation technique for foreign learners aiming at facilitating characters adoption. This technique is based on characters cartography, child-ancestor relations and learning paths. In this paper, we propose an actual implementation of that particular learning technique, providing the learner with a concrete tool supporting characters memorisation. We shall detail our design approach as well as software engineering and user interface aspects. In addition, we have conducted an experimental evaluation of our educational methodology and its implementation. The results obtained are extremely positive.

Keywords: education, kanji, map, Chinese, language

1 Introduction

There exist numerous characters in the Japanese language; three main types of characters can be distinguished: hiragana, katakana and kanji characters, the latter being the most challenging for the learner as there are thousands of kanji. Japanese people themselves regularly encounter unknown characters, or characters they have forgotten, which shows the real difficulty of the memorisation task for the foreign learner. Indeed, Richmond has shown that foreign learners of the Japanese language "need specialized methods" [1]. Kanji characters originate from Chinese and are used (with variants) in multiple languages and dialects such as Taiwanese, Korean (hanja) and Vietnamese (chữ nôm) amongst others.

Several previous works can be cited: phonetic analysis by Vaccari and Vaccari [2], morphological information by Mori et al. [3], and primitive decomposition [4, 5] are some examples. As discussed in [6], the conventional learning approaches for kanji characters most always involve a long, "flat" list of characters, possibly grouped according to some criteria (theme, pronunciation, etc), and is obviously neither attractive nor efficient for memorisation.

---

[*] Advanced Institute for Industrial Technology, Tokyo Metropolitan University, Tokyo, Japan
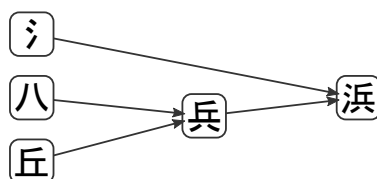
Figure 1: Possible learning paths for the character 浜.

In this paper, aiming at facilitating Japanese characters memorisation (for foreign learners first, applicability and relevance of this method to native Japanese speakers remains to be shown), we propose an implementation of the novel memorisation technique presented in [6], which is based on characters cartography and learning paths. Note that we describe here *one* possible implementation of this cartography method, and that other, different implementation solutions could be imagined as well.

The rest of this paper is organised as follows. First, the main results obtained in [6] are briefly recalled in Section 2. Then, details of the proposed method are described in Section 3. Data structure and software engineering related topics are addressed in Section 4. Next, user interface and control are discussed in Section 5. The proposed education methodology and its implementation are evaluated in Section 6 and discussed in Section 7. Finally, this paper is concluded in Section 8.

2   Previous work

We recall in this section part of the results obtained in [6]. First, characters have been decomposed into basic elements. For instance, the character 浜 can be decomposed into the three basic elements 氵, 八 and 丘.

Then, relations between characters have been defined. Several types of relations were introduced. For instance, bushu relations, that is the connection of one character with the character being its key (e.g. 清 linked to 水). The relation between a character and its basic elements is another type of relation. Most notably, ancestor-child relations have been introduced, bringing some hierarchy amongst characters. For example, the character 丘 is included inside the character 兵; we say that the character 丘 is an ancestor for the character 兵, and that the character 兵 is a child for the character 丘.

Next, from the definition of ancestor-child relations, through layerisation learning paths have been introduced. The purpose of learning paths is to facilitate characters memorisation by learners with a progressive approach: a succession of characters is provided so that a character can be memorised step by step. As an example, possible learning paths for the character 浜 are given in Figure 1.

The positive impact of this learning method has been shown through an experiment involving foreign learners with no or almost no knowledge of such characters, comparing the amount of correct answers to a common questionnaire submitted to two groups of subjects (a control group and an experimental group).

3   Method

As we are dealing with [Japanese characters] cartography, our main idea was to design a learning tool that would reuse the well known codes of geographical mapping, positioning
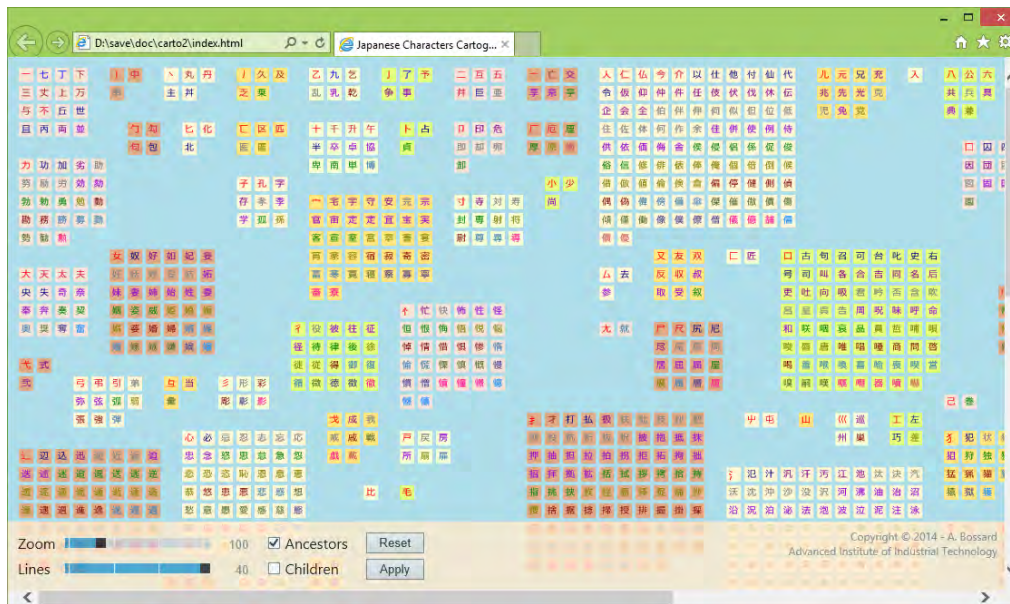
Figure 2: Full view of the cartography showing multiple characters islands and the control interface at the bottom.

blocks of characters side by side in a particular fashion, similar to the numerous isles of an archipelago. We have followed the design of a conventional geographical map: characters are grouped to form islands of various surface areas on a light blue background.

Characters grouping is performed as follows. To each character key (bushu), we create one island. Hence, each island can be uniquely identified by its character key. In addition to one unique character key, each island includes all the characters whose keys match that of the current island. For instance, the island corresponding to the character key 人 includes, amongst others, the characters 供 and 倍 whose respective keys are both 人.

Concretely, each island is populated as detailed below. First, the dimensions of one island are calculated based on the number of characters to be included onto that island, that is, the number of characters whose keys are that of the considered island. An island is a matrix of tiles, each tile containing one character. Practically, an island is a square of side $\lceil \sqrt{n} \rceil$ ($= k$) tiles with n the number of characters, including the key, to be contained on that particular island.

Then, characters are successively inserted from left to right, top to bottom, starting from the character key corresponding to the current island. So, even if an island is a square of side k tiles, the last tiles may be empty (precisely, the last $k - (n \bmod k)$ tiles are empty). Lastly, islands are packed up in order to avoid large unused areas on the map. A margin of at least one tile is always enforced between islands. An illustration is given in Figure 2.

Also, colour codes have been employed when rendering islands and characters. The character key (bushu) which identifies an island (one for each island) is painted in red. Then, characters are painted with a distinct colour depending on the number of lines they contain. For instance, the character 少 includes four lines and according to our character colour palette is painted in brown. Each island (i.e. the background of characters tiles that actually form the island) is painted in one colour. Islands are painted using a colour specified in the island colour palette which contains in total six different tones of soil-like colours. Islands are rendered sequentially using the background palette colours alternatively

Figure 3: Ancestor relations as displayed for the character 会. The three ancestors 人, 二 and ム are emphasized with arrows.

(by rotation).

In addition, child-ancestor relations are displayed similarly to airline flight connections between cities on traditional transportation maps so as to further enhance user friendliness with the reutilisation of well known geographical mapping codes. Concretely, related characters are connected by bent arrows originating from ancestor characters and leading to child characters. An illustration is given in Figure 3. In this figure, the character 会 has been selected by the user, and its three ancestors 人, 二 and ム are subsequently emphasized with arrows.

4 System Overview

In this section, we give some details regarding the software implementation of the cartography system. The front end of our system has been realised in HTML5 [8], while the back end has been implemented in Javascript. The map itself is rendered by using the versatile HTML5 canvas element [9]. The portability of our system is thus maximised.

First, we have manually assembled a database containing all the 2,136 regular usage characters (jōyō kanji) as defined by the Japanese Ministry of Education, Culture, Sports, Science and Technology [7]. In practice, this database consists of one Javascript file db.js including one single statement, the definition of the variable db which is a nested array.

In the database, characters are grouped depending on their keys (bushu), and keys are in turn grouped together depending on their numbers of lines. Concretely, the array db contains 14 elements, each one including all the character keys made of that particular number

of lines. Each character key consists in turn of several sets of characters, each set being uniquely identified by the number of lines of its characters (the number of lines of the key is subtracted). An overview of the database structure is given in Listing 1 and an example in Listing 2.

Listing 1: Structure of the nested array `db` used as database.

```
var db = [
 [key_lines,
  [key, [char_lines, char, char...], [char_lines, char,
      char...]...],
  [key, [char_lines, char, char...], [char_lines, char,
      char...]...],
  ...
 ],
 [key_lines,
  [key, [char_lines, char, char...], [char_lines, char,
      char...]...],
  ...
 ]...
];
```

Listing 2: Example of the `db` database.

```
var db = [
 [1, // keys of 1 line
  ['一', // the key
   [1, '七', '丁'], // first set
   [2, '下', '三', '丈', '上', '万', '与'],
   ... // other sets for the key 一
  ]... // other keys of 1 line
 ]... // keys of 2 and more lines
];
```

A second database is being used in order to define characters ancestor-child relations. This file `db_adjacency.js` is similar to the database `db.js`: it contains one single statement, the definition of the variable `relations` which is an associative array (similar to a JSON object) whose structure is detailed in Listing 3 and with an example in Listing 4.

Listing 3: Structure of the associative array `relations` used as ancestor-child relations database.

```
var relations = {
 char: { children: [char, char...], ancestors: [char, char...] },
 char: { children: [char, char...], ancestors: [char, char...] },
 ...
};
```

Listing 4: Ancestor-child relations database example.

```
var relations = {
 丙: { children: ['病'], ancestors: ['内'] },
 仙: { children: [], ancestors: ['人', '山'] },
 ... // other relations
};
```

Finally, the database `db` is iterated to render the map: for each character key, a new island is created and populated with all the characters stored under that particular key. The underlying structure of the map is a grid of tiles. Tiles are initially empty, and possibly filled with one character when islands are populated. In practice, the grid is a two dimensional array of fixed size; in our implementation, we are using 65 rows and 80 columns. Each element of this array is a reference to a `Character` object whose definition is given in Listing 5. In addition, a hash table `char2idx` (associative array) is also created so as to efficiently retrieve the grid (array) index for a particular character; characters thus act as the keys of this hash table. Additional details regarding colours are given in Section 3.

Listing 5: The `Character` object with its construction routine.

```
function Character(idx, val, lin, fg, bg) {
 this.value = val; // character itself
 this.lines = lin; // total number of lines
 this.foreground = fg; // foreground colour
 this.background = bg; // background colour
 char2idx[val] = idx; // hash table insertion
}
```

Regarding bent arrows as illustrated in Figure 3, we have added to the `CanvasRender ingContext2D` Javascript object a new method `arrow()` by modifying the `prototype` property as follows: `CanvasRenderingContext2D.prototype.arrow = function ( start_x, start_y, end_x, end_y, anchorx_offset, anchory_offset){}`. The body of this method includes calls to conventional canvas drawing functions such as `begin Path()`, `lineTo()` and `rotate()`.

5   Interface and control

In this section, we present the user interface of the cartography system. An illustration is given in Figure 2. As detailed in Section 4, we have been using the `canvas` HTML5 element to render our map. Hence, zooming is greatly facilitated: a simple call to the `scale()` method of the `CanvasRenderingContext2D` object corresponding to our canvas will take care of zooming issues. A slider control is used to scale up or down the view.

A second slider control is used to filter the characters to include in the map. Characters filtering is applied with respect to the number of lines included in a character. This feature aims at accommodating learners of different levels. Concretely, beginners will work on a restricted set of characters, most probably simple ones that are made of few lines, whereas more advanced learners may wish to work on the full set of the 2,136 jōyō kanji.

Two checkboxes allow users to display ancestors and/or children relations when selecting a character in the map. Character selection is triggered by the `pointerup` event as defined in the Pointer Events standard (Candidate Recommendation) of the W3C [10], thus enabling usage of this learning tool on modern touch interfaces such as tablets or smartphones in addition to classic mouse-keyboard interfaces. Fallback to the conventional `mouseup` event is provided for devices (browsers) not supporting Pointer Events.

As with any geographical map, usage of interactive large display solutions enables efficient and comfortable utilisation of the system. We give an illustration in Figures 4 and 5 of a user manipulating the cartography system through a Sharp BIG PAD PN-L702B featuring a 70in display and stylus input and control capabilities.

Figure 4: Global view of the characters map on the BIG PAD. The large display surface is very convenient for cartography purposes.

6 Experimentation and evaluation

In this section, we detail the experiment we conducted aiming at evaluating the proposed memorisation methodology and its implementation.

6.1 Experimental environment

We have gathered several volunteers to take part into this experiment. A total of nine individuals joined the experiment and answered our questionnaire. All participants were Europeans with no or almost no knowledge of Chinese characters, which we recall includes Japanese, Chinese traditional and simplified, Korean, and Vietnamese chữ nôm characters.

Practically, this experiment consisted in an online questionnaire made of multiple-choice questions. At the beginning of the survey, users are first asked to navigate with their Internet browser to the system URL. Then, they are given basic information on how to comfortably use this system, for instance how to zoom in and out on the characters map.

Each question of the survey was designed according to the same pattern. First, an explanation for users is given, detailing several steps using our the system. Users are asked to perform these steps. Then, a question (or a statement) is made to the user, question (or statement) to which users are asked to answer one of the five following values: *Definitely agree*, *Rather agree*, *Neutral*, *Not really agree*, and *Not agree at all*. Practically, users are asked to select (i.e. click with their mouse, or touch with their finger) one character in the cartography system; they are given precise information regarding the location of the character since the participants are novice to Japanese characters classification according to characters keys (bushu) as detailed in Section 3, and this is also the reason why characters used in the experiment were purposely located in an easily findable area on the map. Then ancestors of the selected character are displayed, showing how the selected character is actually formed. The detail of the questions are given in Table 1.

Figure 5: Actual user interaction with the cartography system on the BIG PAD using a stylus.
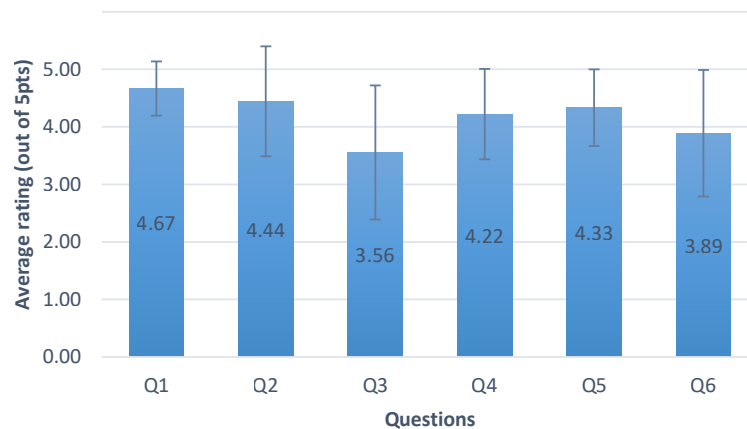

Figure 6: Experiment results: average user ratings, including the standard deviation. Five is the best possible value.

6.2   Results

We now present the results obtained during this experiment.

We have used a scale ranging from 1 to 5 in order to value the five possible answers by users at each question. The values repartition was done as follows: 5–Definitely agree, 4–Rather agree, 3–Neutral, 2–Not really agree, 1–Not agree at all. Hence, the closer a user ranking is to 5, the better. The results are given in Figure 6. This figure describes the average ratings by the users obtained at each of the six questions of Table 1 with the standard deviation.

7   Discussion

We discuss here the results obtained in Section 6. It first essential to notice that the cartography approach to Japanese characters memorisation is very highly ranked by users: we have an average value of 4.22 at the question Q4 asking users about the merit of such ap-

Table 1: Questions asked to the experiment participants.

Q1	Select the character 会 (3rd row, 6th from left). Its three "parent" (ancestors) characters are then displayed and thus help you understand how the character 会 is formed. Would you say that thanks to this system, learning of the character 会 ("meet") has been easy?
Q2	Similarly, select the character 京 (2nd row, north-west of the previous character 会, precisely 19th from left). Its three parents are displayed, thus showing you how the character 京 is formed. Would you say that thanks to this system, learning of the character 京 ("metropole") has been easy?
Q3	Lastly, select the character 仲 (2nd row, north-east of the character 会, precisely 24th from left). Its two parents are displayed. Select its parent the character 中. The two parents of 中 ("middle") are displayed (they are thus two grandparents of the character 仲), thus showing you how the character 仲 ("intermediary") is formed. Would you say that thanks to this system, notably this hierarchy ("genealogy") of characters, learning of the character 仲 has been easy?
Q4	This memorisation system based on "geographical cartography" and linking characters each others makes the memorisation of characters easier than a conventional book such as a dictionary simply listing characters one after the other.
Q5	For the memorisation of Japanese characters (their structure / composition), if you could choose, you would rather use this system based on cartography than a dictionary or a conventional book.
Q6	You found the system easy to use (user interface, etc.).

proach compared to the conventional books or dictionaries. In addition, we obtained an even better average value of 4.33 at the question Q5 inquiring users about their will to use such cartography system to learn Japanese characters instead of a conventional textbook or dictionary.

Very affirmative results at questions Q1 and Q2 (4.67 and 4.44, respectively) show that users found it (very) easy thanks to the ancestor-child relations to learn the two characters 会 and 京 as mentioned in the questionnaire. Again, this shows undoubtedly that introducing such relations between characters is really relevant methodology.

Lastly, the slightly lower result obtained at question Q3 (3.56) can be naturally explained by the fact that the steps to be performed by users were a little more advanced and thus some users may have been left behind. This is notably shown by the standard deviation value (1.17) which is higher at this question than others. We also obtained direct feedback from one user saying "I think this is a ludic approach but anyway efficient for memorisation.", which is worth mentioning.

Lastly, we mention the following remark. Starting from a character such as a basic element but not necessarily, the learner, by looking at the children characters, takes on a "construction" role: similarly to combining blocks or bricks together, here the learner combines characters together (by following a learning path). Giving this role to the learner is an interesting approach in this characters memorisation context. An interesting parallel is that with software development. Effectively, as software makers know, one masters his own work (source code in the case of programmers) much better than the work others. Applied to characters memorisation, enabling the learner to "build" his own characters is a good indication that characters are more likely to be acquired and remembered than looking at

"already built" ones. This is an interesting development of this research and is much worth further discussion.

8   Conclusions

Characters memorisation is an incredibly challenging task for foreign learners of the Japanese language, and even for Japanese people themselves. We have recently proposed a novel memorisation technique based on characters cartography, ancestor-child relations and learning paths. In this paper, we have proposed an actual implementation of the previously introduced methodology as an e-learning tool. We have limited our database to the 2,136 jōyō kanji, yet any Japanese character can be similarly inserted into the system. Application to other, related character sets such as those used in Chinese or hanja is possible. Then, we have conducted an evaluation experiment of the memorisation methodology and its system implementation. Surveyed users overwhelmingly found the cartography approach to characters memorisation and our system implementation very useful and of high merit compared to conventional approached such as textbooks and dictionaries.

Regarding future works, our characters relations database is not complete; expanding it is thus a primary objective. Then, a smartphone or tablet application may be a possibility to support the evaluation of the system. Also, evaluation by Japanese elementary school students would be interesting as they do have the required basic knowledge as mentioned above. Then, comparing the results of such experiment with the evaluation results of this paper would be very valuable.

References

[1] S. Richmond, "A Re-evaluation of Kanji Textbooks for Learners of Japanese as a Second Language," *Journal of the Faculty of Economics*, KGU, 15:43–71, July 2005.

[2] O. Vaccari and E. Vaccari, *Pictorial Chinese-Japanese Characters: A New and Fascinating Method to Learn Ideaographs*, O. Vaccari; sold by the publisher and C. E. Tuttle Co. (Tokyo), 1958.

[3] Y. Mori, K. Sato and H. Shimizu, "Japanese Language Students' Perception on Kanji Learning and their Relationship to Novel Kanji Word Learning Ability," *Language Learning*, 57(1):57–85, 2007.

[4] K. Henshall, *A Guide to Remembering Japanese Characters*, Tuttle Publishing, 1988.

[5] J. Heisig, *Remembering the Kanji, Volume 1: A Complete Course on How not to Forget the Meaning and Writing of Japanese Characters*, University of Hawaii Press, 2007.

[6] A. Bossard, *Japanese Characters Cartography for Efficient Memorization*, International Journal of Computers and Their Applications, 21(3):170–177, 2014.

[7] Jōyō Kanji Table (常用漢字表), Japanese Ministry of Education, Culture, Sports, Science and Technology, Cabinet Notice No. 2, November 30th, 2010. `http://www.bunka.go.jp/kokugo_nihongo/kokujikunrei_h221130.html` (in Japanese). Last visited in October 2014.

[8] Mark Pilgrim, *HTML5: Up and Running*, O'Reilly Media, 2010.

[9] Steve Fulton and Jeff Fulton, *HTML5 Canvas*, O'Reilly Media, 2011.

[10] W3C, *Pointer Events (Candidate Recommendation)*, 2013. `http://www.w3.org/TR/pointerevents/`. Last visited in October 2014.