Modelling the Orthographic Neighbourhood for Japanese Kanji

Lars Yencken and Timothy Baldwin

Computer Science and Software Engineering
University of Melbourne, Victoria 3010 Australia

NICTA Victoria Research Lab
University of Melbourne, Victoria 3010 Australia

{lljy,tim}@csse.unimelb.edu.au

Abstract. Japanese kanji recognition experiments are typically narrowly focused, and feature only native speakers as participants. It remains unclear how to apply their results to kanji similarity applications, especially when learners are much more likely to make similarity-based confusion errors. We describe an experiment to collect authentic human similarity judgements from participants of all levels of Japanese proficiency, from non-speaker to native. The data was used to construct simple similarity models for kanji based on pixel difference and radical cosine similarity, in order to work towards genuine confusability data. The latter model proved the best predictor of human responses.

1 Introduction

In everyday reading tasks, humans distinguish effortlessly between written words. This is despite languages often seeming ill-suited to error-free word recognition, through a combination of inter-*character* similarity (i.e. the existence of graphically-similar character pairs such as 士 [*shi*] and 土 [*tsuchi*]) and inter-*word* similarity (i.e. the existence of orthographically-similar word pairs such as *bottle* and *battle*). While native speakers of a language tend to be oblivious to such similarities, language learners are often forced to consciously adapt their mental model of a language in order to cope with the effects of similarity. Additionally, native speakers of a language may perceive the same character pair significantly differently to language learners, and there may be radical differences between language learners at different levels of proficiency or from different language backgrounds.

This paper is focused on the similarity and confusability of Japanese kanji characters. This research is novel in that it analyses the effects of kanji confusability across the full spectrum of Japanese proficiency, from complete kanji novices to native speakers of the language. Also, unlike conventional psycholinguistic research on kanji confusability, it draws on large-scale data to construct and validate computational models of similarity and confusability. This data set was collected for the purposes of this research via a web experiment, and consists of a selection of both

Yencken, Lars and Timothy Baldwin (2006) Modelling the Orthographic Neighbourhood for Japanese Kanji, In Proceedings of the 21st International Conference on the Computer Processing of Oriental Languages, Singapore.

control pairs aimed at targeted phenomena, and also random-selected character pairs. The research builds on psycholinguistic studies of the visual recognition of both Chinese hanzi and Japanese kanji.

The paper is structured as follows. We begin by discussing the background to this research (Section 2), then follow with a description of our web experiment and its basic results (Section 3). We construct some simple models of the similarity data (Section 4), then evaluate the models using the experimental data (Section 5). Finally, we lay out our plans for future work (Section 6).

2 Background

2.1 Types of similarity

This paper chiefly concerns itself with orthographic similarity of individual Japanese characters, that is, graphical similarity in the way the characters are written. Note that within the scope of this paper, we do not concern ourselves directly with the question of orthographic similarity of multi-character words. That is, we focus exclusively on inter-character similarity and confusability.

Other than simple orthography, character similarity can also be quantised *semantically* or *phonetically*; identically-pronounced characters are termed *homophones*. In Chinese and Japanese, radicals[1] are often semantic or phonetic cues of varying reliability in determining character similarity. When radicals are shared between kanji, more than orthographic similarity may be shared. If a radical is reliably semantic, then two kanji sharing it are likely to be semantically related in some way (e.g. kanji containing the radical 月, such as 胸 [*mune*] "chest" and 腕 [*ude*] "arm", are reliably body parts). If reliably phonetic (e.g. 同, as in 銅 [*dō*] "copper" and 胴 [*dō*] "body"), the two kanji will share a Chinese or *on* reading, and will hence be homophones. It may thus not be impossible for skilled readers to give purely orthographic similarity judgements in the presence of shared radicals, since evidence shows these cues are crucial to reading, as to be discussed in Section 2.2.

2.2 Lexical processing of kanji and hanzi

In considering potential effects to control for, and types of similarity effects, we draw on the many psycholinguistic studies of kanji or hanzi recognition, with the basic assumption that human reading processes for the two scripts are fundamentally similar. It is beyond the scope of this paper to give these studies full treatment, but we refer the interested reader to some pertinent results.

There is much support for a form of hierarchical activation model for the recognition of kanji (Taft and Zhu 1997, Taft, Zhu, and Peng 1999). In

[1] Here, we use the term *radical* liberally to refer to any consistent stroke group, rather than in its narrow sense as either the dictionary index stroke group or the main semantic stroke group.

Fig. 1. Example stimulus pair for the similarity experiment. This pair contains a shared radical on the left.

such a model, firstly strokes are visually activated, which in turn activate the radicals they form, which then activate entire kanji. Evidence for such a model includes experiments which showed stroke count effects (summarized by Taft and Zhu (1997)), and numerous radical effects, including radical frequency with semantic radical interference (Feldman and Siok 1997, Feldman and Siok 1999), and homophony effects which only occurred with shared phonetic radicals (Saito, Inoue, and Nomura 1995, Saito, Masuda, and Kawakami 1998). There is also evidence that structure may be important for orthographic similarity in general, and for radical-based effects (Taft and Zhu 1997, Yeh and Li 2002).

3 Similarity experiment

3.1 Experiment outline

A short web-based experiment was run to obtain a set of gold-standard orthographic similarity judgements. Participants were first asked to state their first-language background, and level of kanji knowledge, pegged to one of the levels of either the Japanese Kanji Aptitude Test (日本漢字能力検定試験)[2] or the Japanese Language Proficiency Test (日本語能力試験).[3] Participants were then exposed to pairs of kanji, in a manner shown in Figure 1, and asked to rate each pair on a five point graded similarity scale. The number of similarity grades chosen represents a trade-off between rater agreement, which is highest with only two grades, and discrimination, which is highest with a large number of grades.
Although participants included both first and second language readers of Chinese, only Japanese kanji were included in the stimulus. Chinese hanzi and Japanese hiragana and katakana were not used for stimulus, in order to avoid potential confounding effects of character variants and of differing scripts. The pairs were also shuffled for each participant, with the ordering of kanji within a pair also random, in order to reduce any effects caused by participants shifting their judgements part-way through the experiment.
Each participant was exposed to a common set of control pairs, to be discussed in Section 3.2 below. Further, a remaining 100 random kanji

[2] A Japanese government test which is tied to Japanese grade school levels initially, but culminates at a level well above that expected in high-school graduates. `http://www.kanken.or.jp/`

[3] The standard general-purpose Japanese aptitude test taken by non-native Japanese learners of all first-language backgrounds. `http://www.jees.or.jp/jlpt/en/index.htm`

Effect type	Example	Description
Frequency (independent)	会 店	Frequency of occurrence of each kanji individually. Both kanji in the example pair are high-frequency.
Co-occurrence	法 考	Both kanji occur with high frequency with some third kanji. For example, 法 [*hō*] "Act (law)" occurs in 法案 [*hōaN*] "bill (law)", and 考 [*kaNga(e)*] "thought" occurs in 考案 [*kōaN*] "plan, idea".
Homophones	弘 博	Both kanji share a reading. In the example, both 弘 [*hiro(i)*] "spacious" and 博 [*haku*] "doctor" share a reading [*hiro*]. For 博 this is a name reading.
Stroke overlap	策 英	Both kanji share many similar strokes, although no radicals are shared.
Shared graphemes	働 動	Both kanji share one or more graphical elements. These elements might occur in any position.
Shared structure	幣 哲	Both kanji share the same structural break-down into sub-components, although the sub-components differ.
Stroke count	奮 撃	Pairs comparing and contrasting stroke counts. Both examples here have a very high stroke count.
Part of speech/function	方 事	Both kanji have a common syntactic function in language. When added to a verb, it is converted to a noun.
Semantic similarity	千 万	Both kanji are semantically similar. In the example, they are both numbering units.

Fig. 2. Groups of control pairs used, with an example for each. Parts of readings in brackets indicate *okurigana*, necessary suffixes before the given kanji forms a word.

pairs were shown where both kanji were within the user's specified level of kanji knowledge (where possible), and 100 were shown where one or both kanji were outside the user's level of knowledge. This was in order to determine any effects caused by knowing a kanji's meaning, its frequency, its readings, or any other potentially confounding properties.

Web-based experiments are known to provide access to large numbers of participants and a high degree of voluntariness, at the cost of self-selection (Reips 2002). Although participants of all language backgrounds and all levels of kanji knowledge were solicited, the nature of the experiment and the lists advertised to biased participants to be mainly of an English, Chinese or Japanese first-language background.

3.2 Control pairs

There are many possible influences on orthographic similarity judgements which we hoped to detect in order to determine whether the data could be taken at face value. A sample pair and a description of each control effect is given in Figure 2. Since the number of potential effects considered was quite large, the aim was not statistical significance for the presence or absence of any effect, but rather guidance in similarity modelling should any individual effect seem strong. All frequency and co-occurrence counts were taken from 1990–1999 Nikkei Shinbun corpus data.

3.3 Results

The experiment had 236 participants, with a dropout rate of 24%. The participants who did not complete the experiment, and those who gave no positive responses, were filtered from the data set. The remaining 179

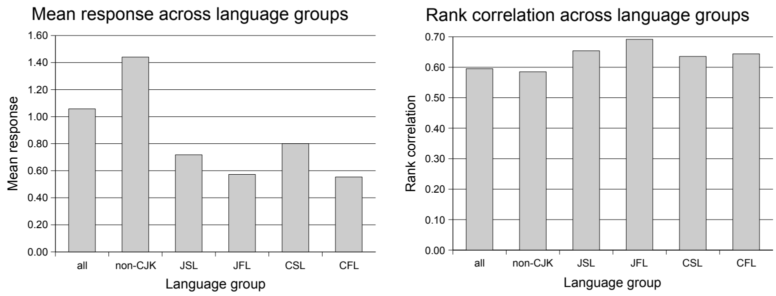

Fig. 3. Mean responses and rank correlation when broken up into participant groups, measured over the control set stimulus.

participants are spread across 20 different first languages. Mapping the responses from "Very different" as 0 to "Very similar" as 4, the mean response over the whole data set was 1.06, with an average standard deviation for each stimulus across raters of 0.98.

To measure the inter-rater agreement, we consider the mean rank-correlation across all pairs of raters. Although the kappa statistic is often used (Eugenio and Glass 2004), it underestimates agreement over data with graded responses. The mean rank correlation for all participants over the control set was strong at 0.60. However, it is still lower than that for many tasks, suggesting that many raters lack strong intuitions about what makes one kanji similar to another.

Since many of the first language backgrounds had too few raters to do significant analysis on, they were reduced to larger groupings of backgrounds, with the assumption that all alphabetic backgrounds were equivalent. Firstly, we group first-language speakers of Chinese (CFL) and Japanese (JFL). Secondly, we divide the remaining participants from alphabetic backgrounds into second language learners of Japanese (JSL), second language learners of Chinese (CSL), and the remainder (non-CJK). Participants who studied both languages were put into their dominant language based on their comments, or into the JSL group in borderline cases.[4]

Figure 3 shows mean responses and agreement data within these participant groups. This grouping of raters is validated by the marked difference in mean responses across these groups. The *non-CJK* group shows high mean responses, which are then halved for second language learners, and lowered still for first language speakers. Agreement is higher for the first-language groups (JFL and CFL) than the second-language groups (JSL and CSL), which in turn have higher agreement than the non-speakers. Both of these results together suggest that with increasing experience, participants were more discerning about what they found to be similar, and more consistent in their judgements.

[4] Many alternative groupings were considered. Here we restrict ourselves to the most interesting one.

3.4 Evaluating similarity models

Normally, with high levels of agreement, we would distill a gold standard data-set of similarity judgements, and evaluate any model of kanji similarity against our gold-standard judgements. Since agreement for the experiment was not sufficiently high, we instead evaluate a given model against all rater responses in a given rater group, measuring the mean rank-correlation between the model and all individual raters in that group.

We also have reference points to determine good levels of agreement, by measuring the performance of the *mean rating* and the *median rater response* this way. The mean rating for a stimulus pair is simply the average response across all raters to that pair. The median rater response is the response of the best performing rater within each stimulus set (i.e. the most "agreeable" rater for each ability level), calculated using the above measure.

4 Similarity models

4.1 Pixel difference model

In Section 2.2, we briefly discussed evidence for stroke level processing in visual character recognition. Indeed, confusability data for Japanese learners taken from the logs of the FOKS (Forgiving Online Kanji Search) error-correcting dictionary interface suggests that stroke-level similarity is a source of error for Japanese learners. The example 基 [*ki, moto*] "basis" and 墓 [*bo, haka*] "grave/tomb", was taken from FOKS dictionary error logs (Bilac, Baldwin, and Tanaka 2003), and is one of the pairs in our "Stroke overlap" control subgroup (Figure 2).

This example shows that learners mistake very similar looking kanji, even when there are no shared radicals, if there are sufficient similar looking strokes between the two kanji. Ideally, with a sufficiently rich data set for kanji strokes, we could model the stroke similarity directly. As an approximation, we instead attempt to measure the amount that strokes overlap by rendering both kanji to an image, and then determine the pixel difference d_{pixel} between the two rendered kanji. We can easily move from this distance metric to a similarity measure, as below:

$$d_{\text{pixel}}(k_a, k_b) = \frac{1}{L} \sum_{L}^{i=0} |\text{Image}(k_a)(i) - \text{Image}(k_b)(i)| \tag{1}$$

$$s_{\text{pixel}}(k_a, k_b) = 1 - d_{\text{pixel}}(k_a, k_b) \tag{2}$$

This calculation is potentially sensitive both to the size of the rendered images, and the font used for rendering. For our purposes, we considered an image size of 100×100 pixels to be sufficiently detailed, and used this in all experiments described here. To attempt to attain reasonable font independence, the same calculation was done using 5 commonly available fonts, then averaged between them. The fonts used were: Kochi Gothic (medium gothic), Kochi Mincho (thin mincho), Mikachan (handwriting), MS Gothic (thick gothic), and MS Mincho (thin mincho). The graphics program Inkscape[5] was used to render them non-interactively.

This method of calculating similarity is brittle. Suppose two characters share a significant number of similar strokes. If the font renders the characters in such a way that the similar strokes are unaligned or overly scaled, then they will count as differences rather than similarities in the calculation. Further robustness could be added by using more sophisticated algorithms for scale and translation invariant image similarity.

Consider the minimum pixel difference of a pair over all possible offsets. This defines a translation invariant similarity measure. Since the current method calculates only one alignment, it is an underestimate of the true translation invariant similarity. Since characters are rendered in equal size square blocks, and radical frequency is position-dependent, the best alignment usually features a low offset between images. The current approximation is thus a close estimate on average, and is considerably less expensive to compute.

Pixel difference is also likely to underestimate the perceptual salience that repeated stroke units (i.e. radicals) have, and thus underestimate radical-level similarity, except where identical radicals are aligned. Nevertheless, we expect it to correlate well with human responses where stroke-level similarity is present. Pairs scored as highly similar by this method should thus also be rated as highly similar by human judgements.

4.2 Bag of radicals model

Just as the pixel model aimed to capture similarity effects at the stroke-level, we now try to capture them at the radical level. Fortunately, at the radical-level there is an existing data-set which indexes kanji by all of their contained radicals, the *radkfile*.[6] It was designed to aid dictionary look-up, and serves as a simple method of determining all the unique radicals used by a kanji.

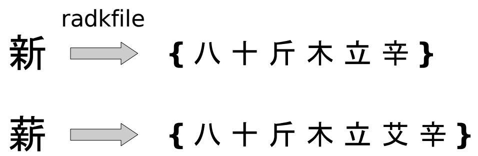

Fig. 4. Kanji are decomposed into their radicals using the radkfile. Each set of radicals can be considered a boolean vector over all radicals, where only the radicals which are present are stored. Note that the character used to represent each single radical is not always identical to that radical.

Using all the potential radicals as dimensions, we can map each kanji onto a vector space of radicals, giving it a boolean entry in each dimension determining whether or not the kanji contains that radical. On

[5] `http://www.inkscape.org`

[6] `http://ftp.monash.edu.au/pub/nihongo/radkfile.gz`

Predictor	Example	Similarity
Mean rating	a. 漢 菓	0.938
	b. 飫 餘	0.833
	c. 基 墓	0.830
Pixel	d. 土 士	0.878
	e. 人 入	0.844
	f. 官 宮	0.771
Bag of radicals	g. 火 炎	1.000
	h. 累 細	1.000
	i. 馬 駅	0.500

Fig. 5. Examples of high similarity pairs according to mean rating, the pixel model and the bag of radicals model. The mean rating pairs were taken from the experimental data, whereas the other pairs were taken from general use kanji.

this vector space, we can calculate the cosine similarity between the two kanji vectors, to achieve a simple similarity measure based on radical composition:

$$s_{\text{radical}}(k_a, k_b) = \frac{\text{radicals}(k_a) \bullet \text{radicals}(k_b)}{|\text{radicals}(k_a)||\text{radicals}(k_b)|} \tag{3}$$

Comparing high-similarity examples from the different methods (Figure **??**), we can immediately see some drawbacks to this model. Example (g) shows that the number of each radical present is discarded, hence 火 and 炎 are considered identical with this method. Example (h) shows that position is also discarded, yet there is evidence that radical effects are position specific (Taft and Zhu 1997, 1999). This model also ignores similarity due to stroke data, yet the existence of high-similarity examples such as (d) and (e) which do not share radicals indicates that stroke overlap can also be a significant contributor to similarity. Larger structure such as layout may also be important for similarity (Yeh and Li 2002), and it too is discarded here.

Nonetheless, radicals are clearly significant in the perception of kanji. If the presence or absence of shared radicals is the main way that individuals perceive similarity, then this model should agree well with human judgements, whether or not they make use of the additional semantic or phonetic information these radicals can encode.

5 Model evaluation

The pixel and radical models were evaluated against human judgements in various participant groups, as shown in Figure 6, and can be compared to the mean rating and median raters. The pixel based similarity method exhibits weak rank correlation across the board, but increasing in correlation with increasing kanji knowledge. The radical model however shows strong rank correlation for all groups but the non-CJK, and better improvements in the other groups.

These results match our predictions with the pixel-based approach, in that it performs reasonably, but remains only an approximation. The

Group	*Mean*	*Median*	*Pixel*	*Radical*
Non-CJK	0.69	0.55	0.34	0.47
CSL	0.60	0.65	0.38	0.56
CFL	0.51	0.62	0.44	0.66
JSL	0.64	0.70	0.43	0.59
JFL	0.56	0.69	0.46	0.68
All	0.65	0.62	0.39	0.54

Fig. 6. Rank correlation of pixel and radical models against raters in given participant groups. Mean and median raters provided as reference scores.

Band	*Mean*	*Median*	*Pixel*	*Radical*
$[0, 1)$	0.69	0.55	0.34	0.47
$[1, 200)$	0.62	0.60	0.38	0.53
$[200, 600)$	0.64	0.69	0.41	0.61
$[600, 1000)$	0.69	0.72	0.46	0.52
$[1000, 2000)$	0.56	0.70	0.46	0.65
$[2000, \ldots)$	0.58	0.73	0.48	0.70

Fig. 7. Rank correlation of pixel and radical models against raters in across bands of kanji knowledge. Each band contains raters whose number of known kanji falls within that band's range.

radical method's results however are of a comparable level of agreement within the CFL and JFL groups to the median rater, a very strong result. It suggests that native speakers, when asked to assess the similarity of two characters, make their judgements primarily based either on the radicals which are shared between the two characters, or on some other measure which correlates well to identification of shared radicals. Intuitively, this makes sense. Native speakers have a great knowledge of the radicals, their meaning and their semantic or phonetic reliability. They also have the most experience in decomposing kanji into radicals for learning, writing and dictionary lookup.

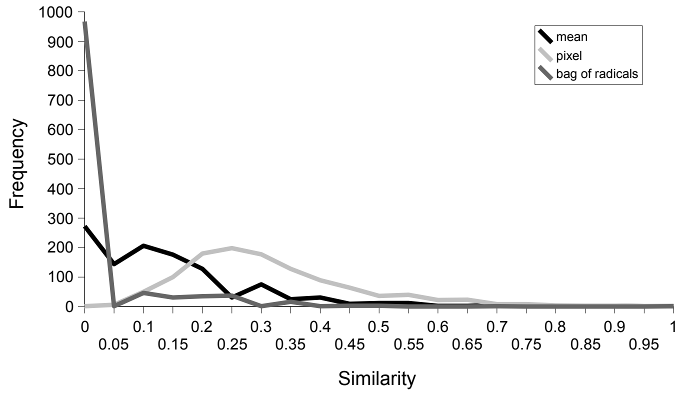

Fig. 8. Histograms of scaled responses across all experimental stimulus pairs, taken from mean rating, pixel and bag of radical models. Responses were scaled into the range $[0, 1]$.

The radical model still has poor correlation with the non-CJK group, but this is not an issue for applications, since similarity applications primarily target either native speakers or learners, who either already have or will pick up the skill of decomposing characters into radicals. To attempt to determine when such a skill gets picked up, Figure 7 shows agreement when raters are instead grouped by the number of kanji they claimed to know, based on their proficiency level. Aside from the $[600, 1000)$ band, there are consistent increases in agreement with the radical method as more kanji are learned, suggesting that the change is gradual, rather than sudden. Indeed, learners may start by focusing on strokes, only to shift towards using radicals more as their knowledge of radicals improves.

If we compare the histograms of the responses in Figure 8, we can see stark differences between human responses and the two models. The radical model considers the majority of stimuli to be completely dissimilar. Once it reaches stimulus pairs with at least one shared radical, its responses are highly quantised. The pixel model in comparison always finds some similarities and some differences, and exhibits a normal style bell curve. Human responses lie somewhere in between the pixel and radical models, featuring a much smaller number of stimuli which are completely dissimilar, and a shorter tail of high similarity than found with the pixel model.

6 Applications and future research

6.1 Similarity

We have suggested several potential improvements to our similarity modelling. In particular, a translation invariant version of pixel similarity could be easily constructed and tested. On the other hand, the data-set created by Apel and Quint (2004) provides rich stroke data, which would allow a holistic model combining strokes, radicals and layout into a unified similarity metric. This should be superior to both the pixel model, which only approximates stroke-level similarity, and the radical model, which discards position and stroke information. The data-set created here allows fast and simple evaluation of any new similarity models, which should help foster further experimentation.

Kanji similarity metrics have many potential uses. A similarity or difference metric defines an orthographic space across kanji, which we can in turn use in novel ways. Our interest lies in dictionary lookup, and indeed a user could browse across this space from some seed point to quickly and intuitively arrive at a target kanji whose pronunciation is unknown. Particularly dense regions of this space will yield easily confusable pairs or clusters of high similarity. Presenting these to learners during study or testing could help these learners to differentiate between similar characters, but also to better structure their mental lexicon. Depending on the level of similarity the application is concerned with, the high amount of quantisation of responses may be a disadvantage for thresholding to only high-similarity responses. This remains one advantage of the pixel model over the radical model.

6.2 Confusability

From a similarity metric, we can then construct a confusability probability across pairs of kanji. Since confusability need not be symmetric, there may be other effects such as frequency which also play a role. Several studies of character perception at the word-level have found evidence of

asymmetric interference effects for low frequency words with high frequency neighbours (van Heuven, Dijkstra, and Grainger 1998).
Similarity provides a means to bootstrap collection of confusability data, useful since authentic confusability data is difficult to find or construct. The available data mainly comes from controlled experiments in artificial environments, for example in explorations of illusory conjunctions (Fang and Wu 1989). Hand analysed logs for the FOKS dictionary detected a few accidentally corrected orthographic confusability examples, suggesting genuine occurrence of these errors (Bilac, Baldwin, and Tanaka 2004). The FOKS system also provides a method of turning a basic confusability model into a source of genuine confusability data. By adding the confusability model to the FOKS error model, any errors successfully corrected using the model will indicate genuine confusion pairs. We thus can create an informed confusability model, which bootstraps a cycle of confusability data collection and model validation.

6.3 Perception

There remain many open questions in orthographic similarity effects. Since the control pairs were not numerous enough to statistically determine similarity effects from the various effect types, further experimentation in this area is needed. In particular, it unconfirmed as to whether semantic or phonetic similarity contributed to the similarity judgements analysed here. It could be tested by comparing pairs that share the same number of radicals, where the shared radicals for one pair were reliable semantic or phonetic cues, but the shared radicals for the other pair were not. We have discussed positional specificity of shared radicals as shown by Taft and Zhu (1997, 1999); the same specificity may also occur in radical-based similarity effects, and should be further investigated, as should stroke level effects.

7 Conclusion

We carried out an experiment seeking graphical similarity judgements, intending to form a similarity dataset to use for modelling. Since agreement within raters was moderate, we instead used the human judgements directly as our dataset, evaluating models against it. Two models were proposed, a stroke-based model and a radical based model. The stroke-based model was approximated using pixel differencing, rather than created directly. The pixel model showed medium agreement, but the radical model showed agreement as strong as the best individual rater for native speakers.
Although the radical-based model's performance may be adequate for some applications, there is much promise for a holistic model taking into account stroke, radical and positional effects. The data-set created here provides a means for quick and effective evaluation of new similarity models for kanji, thus allowing much experimentation. As well as seeding new dictionary lookup methods, the similarity models considered provide

a basis for confusability models, which are themselves useful for error-correcting lookup, and in turn generating confusion data. Such confusion data, along with the similarity judgments collected here, will provide important evidence for understanding the perceptual process for kanji.

References

Apel, U. and Quint, J.: Building a graphetic dictionary for Japanese kanji - character look up based on brush strokes or stroke groups, and the display of kanji as path data, in Proceedings of the 20th International Conference on Computational Linguistics

Bilac, S., Baldwin, T., and Tanaka, H.: Modeling learners' cognitive processes for improved dictionary accessibility, in Proceedings of the 10th International Conference of the European Association for Japanese Studies, Warsaw, Poland

Bilac, S., Baldwin, T., and Tanaka, H.: Evaluating the FOKS error model, in Proc. of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, 2119–2122

Eugenio, B. D. and Glass, M.: The kappa statistic: A second look. Computational Linguistics **30** (2004)(1) 95–101

Fang, S.-P. and Wu, P.: Illusory conjunctions in the preception of Chinese characters. Journal of Experimental Psychology: Human Perception and Performance **15** (1989) 434–447

Feldman, L. and Siok, W.: Semantic radicals contribute to the visual identification of chinese characters. Journal of Memory and Language **40** (1999) 559–576

Feldman, L. B. and Siok, W. W. T.: The role of component function in visual recognition of Chinese characters. Journal of Experimental Psychology: Learning, Memory and Cognition **23** (1997) 776–781

Reips, U.-D.: Standards for internet-based experimenting. Experimental Psychology **49** (2002)(4) 243–256

Saito, H., Inoue, M., and Nomura, Y.: Information processing of Kanji (Chinese characters) and Kana (Japanese characters). Psychologia **22** (1995) 195–206

Saito, H., Masuda, H., and Kawakami, M.: Form and sound similarity effects in kanji recognition. Reading and Writing **10** (1998)(3 - 5) 323–357

Taft, M. and Zhu, X.: Submorphemic processing in reading Chinese. Journal of Experimental Psychology: Learning, Memory and Cognition **23** (1997) 761–775

Taft, M., Zhu, X., and Peng, D.: Positional specificity of radicals in chinese character recognition. Journal of Memory and Language **40** (1999) 498–519

van Heuven, W. J. B., Dijkstra, T., and Grainger, J.: Orthographic neighborhood effects in bilingual word recognition. Journal of Memory and Language **39** (1998) 458–483

Yeh, S.-L. and Li, J.-L.: Role of structure and component in judgments of visual similarity of chinese characters. Journal of Experimental Psychology: Human Perception and Performance **28** (2002)(4) 933–947