

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Parallel vs Serial Issues in Reading Aloud: Evidence for Parallel Processing from a Computational Model of Japanese Kanji & Kana Nonword Reading

Permalink

<https://escholarship.org/uc/item/40f7h9x7>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN

1069-7977

Authors

Ueno, Taiji
Ikeda, Kenji
Ito, Yuichi
[et al.](#)

Publication Date

2014

Peer reviewed

Parallel vs Serial Issues in Reading Aloud: Evidence for Parallel Processing from a Computational Model of Japanese Kanji & Kana Nonword Reading

Taiji Ueno^{1,2} (taijiueno7@gmail.com)

Kenji Ikeda^{2,3} (ikeda.kenji@f.mbox.nagoya-u.ac.jp)

Yuichi Ito^{2,3} (ito.yuichi@nagoya-u.jp)

Shinji Kitagami³ (kitagami@cc.nagoya-u.ac.jp)

Jun Kawaguchi³ (kawaguchijun@nagoya-u.jp)

1. Department of Psychology, University of York, Heslington, York, YO10 5DD, UK. 2. Japan Society for the Promotion of Science; 3. Graduate School of Environmental Studies, Nagoya University, Nagoya, 4648601, JAPAN.

Abstract

Lower nonword reading performance in the old version of parallel-distributed processing models in English was taken as evidence for the necessity of sequential components in reading aloud, but the later updates have resolved this issue without discarding the parallel processing principle. A model's validity can be tested by cross-script extensions. On this point, an extension to Japanese kanji reading has posed a question. Specifically, the latest Japanese connectionist model has incorporated a sequential component in order to improve kanji nonword reading performance. The present study, however, proposed an alternative, fully parallel-distributed processing model to map distributed kanji and kana visual representations to phonemic/moraic representations, with satisfactory nonword reading performance. First cross-linguistic evidence from Japanese is provided to support a single-mechanism theory that postulates parallel-processing for both words and nonwords reading.

Keywords: parallel-distributed processing model; sequential processing; reading, Japanese kanji; nonword reading

Introduction

The processes and mechanisms for reading aloud is one of the controversial issues in cognitive science. Implemented computational models have contributed to this issue by explicitly demonstrating the mechanisms and processes that can reproduce the behavioral data (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007; Plaut, McClelland, Seidenberg, & Patterson, 1996). These models differ in whether a sequential component is permitted as well as parallel processing component, mainly in order to simulate nonword reading accuracy (Besner & Smith, 1992) and the word length effect (Weekes, 1997), and so on. The present study contributes to this issue by providing the first example of a parallel-distributed processing (PDP, without a sequential component) connectionist model which can read Japanese kanji and kana words/nonwords.

A historical overview of the parallel vs serial issues in English models should be helpful here. The PDP model by Seidenberg and McClelland (1989) incorporated only a single mechanism with parallel processing, and was able to read both *regular* (e.g., *mint*) and *irregular* (i.e., *pint*) words.

However, this model (hereafter, SM89) was unable to read nonwords (i.e., untrained input pattern) as satisfactorily as humans (Glushko, 1979). Advocates of dual-route models argued that SM89 was only an approximation of a lexical look-up process, and this was taken as evidence for the necessity of sequential components (Coltheart et al., 2001). Specifically, they expected that sequentially applying a grapheme-phoneme correspondent rule per a grapheme would allow a model to pronounce a nonword satisfactorily. In contrast, Plaut et al. (Plaut et al., 1996) explained it was only due to the size of the training set and/or to the input coding schemes. After sorting these issues out, Plaut et al. (hereafter, PMSP96) successfully simulated nonword reading performance without a sequential component.

Interestingly, the connectionist models for Japanese kanji reading went part of the same way as English, but arrived at the opposite end. It is widely accepted that a cross-script extension is a useful paradigm to glean further evidence for the validity of computational models (Perry et al., 2007). Given the success of PMSP96, it was natural that one of these authors, Patterson and her Japanese colleagues embarked on a Japanese PDP project. The outcome of their initial attempt was successful and similar to the situation of SM89, such that the model was able to read both Japanese kanji consistent/regular words (later in details for the operational definition) and irregular words within a single mechanism (Ijuin, Fushimi, Patterson, & Tatsumi, 1999). However, this model was unable to read kanji nonwords (2-kanji compounds, but the combination of these 2 kanji characters does not exist in real Japanese) as correctly as humans (Fushimi, Ijuin, Patterson, & Tatsumi, 1999). Given the history of English models, they naturally expected that poor nonword reading would be attributed to extrinsic factors such as training corpus and coding schemes, rather than intrinsic factors such as its lack of indispensable mechanisms. However, the follow-up studies conducted by Japanese researchers in this group (Ijuin, Fushimi, & Tatsumi, 2001, 2002) have actually reached the different conclusion. Specifically, Ijuin et al. (2001) explicitly mentioned as follows (translated into English by us):

Our network was able to read only 7%~20% of nonword sets. There are possible factors to affect network's nonword reading, including (i) the size of the training corpus, (ii) input coding scheme, & (iii) the architecture of the network ('architecture' is the direct translation of their Japanese, but as will be shown later, what Ijuin et al. actually referred here was 'mechanism', not 'architecture'). We have already tested the effects of (i) and (ii), but we know that modifying these parameters do not improve the performance equivalent to humans

(The note in the parenthesis is added by us)

Thus, Ijuin et al. (2001) concluded that manipulating the extrinsic factors was not enough to improve nonword reading accuracy, and instead dealt with the 'architecture'. For this purpose, Ijuin et al. (2001, 2002) hardwired a *position layer*, which contained two task units, each of which sent a *time-/position-specific signal sequentially per a pronunciation of one character* to the network (Figure 1), and the network generated the correct pronunciation of each character sequentially. As a result, the network improved its nonword reading equivalent to human adults (Fushimi et al., 1999). They concluded "this model is clearly different from dual-route models because the 2 kanji characters were presented in parallel (Ijuin et al., 2001, 2002)", and assumed it was a so-called triangle model.

In this study, we report two simulations. In Simulation 1, we replicated Ijuin et al. (2001, 2002)'s sequential output model with a position layer, and our reanalysis of the model showed that the model was actually hardwired dual mechanisms, one of which was specialized for a sequential processing (parsing) mechanism. In Simulation 2, we proposed a PDP model with fully-parallel outputs. This model was able to read both kanji words and nonwords satisfactorily without a sequential component.

Simulation 1 (Replication and Reanalysis)

Methods

Reading Japanese Kanji Words and Nonwords

To facilitate understanding of the implementation details, an explanation for the nature of mapping in Japanese kanji reading would be helpful. Correct reading of a Japanese 2-kanji ideogram is an example of a quasiregular domain (Seidenberg & McClelland, 1989). One-third of a single kanji character has only one possible pronunciation (thus *consistent*). In contrast, the remaining two-thirds have several legitimate pronunciations (*inconsistent*). In addition, the level of *typicality* in each correct pronunciation has been operationalized by Fushimi et al. (1999) as the type-frequency of the phonemic-/moraic-friends among the orthographic-neighbors (i.e., the entire Japanese 2-kanji ideogram cohort sharing the same kanji character in the same position). For example, when 食 (meaning *eat*) appears in the 2nd position of a 2-kanji ideogram, then in most cases (thus *typical*) the target pronunciation is [syo-ku]

(e.g., 和食 [wa-syo-ku], meaning *Japanese food*; 洋食 [yo-u-syo-ku], meaning *Western food*). However, there are a few words (thus *atypical*) whose pronunciation is [ji-ki] (e.g., 断食 [da-N-ji-ki], meaning *fasting*). In other words, reading these inconsistent-atypical words cannot rely on the statistical bias (or *rule*), and instead should refer to the adjacent character to compute word-specific information, such as lexical/semantic knowledge (Fushimi et al., 1999).

Fushimi et al. (1999) also created a 2-kanji nonword by randomly combining 2 kanji characters. These combinations do not exist in real Japanese, such that there is no way to decide a single correct pronunciation for each character. Therefore, a human output in nonword kanji reading is scored as correct if any of the possible/legitimate pronunciations is generated in each character (mean accuracy = 89% in Fushimi et al., 1999). Because of this scoring method, letter-by-letter reading without a reference to the adjacent characters always leads to correct nonword reading. We will return to this issue later.

Finally, there is another script, kana in Japanese. Each kana character has almost perfect transparency to its mora phonology (thus consistent). Because of this nature, every kanji word can be represented in kana form by translating a kanji orthography into its pronunciation (morae), and then converting that morae sequence to the one-to-one corresponding kana sequence (thus, kana-length of a kanji word = its mora-length). The reading models here were trained for mapping from a kanji representation to its mora representation as well as mapping from its kana form to the same mora representation for all the lexical items.

Network: Sequential Outputs & Position Layer

Figure 1 shows the architecture of the network, adapted from Ijuin et al. (2002). There was a reason why we did not use exactly the same network in a strict sense reported in Ijuin et al. (2002). Specifically, their original model implemented the so-called semantic layer (Plaut et al., 1996), whose input moved the activity of the output layer units towards the target value. When this technique was *combined* with the localist representations and with the max-criterion scoring method, then high kanji nonword reading accuracy could be an artefact (NB. Each of these techniques has no issue when used in isolation, see Simulation 2). Specifically, during training of words, the direct orthography-phonology route of such a model becomes a *system which knows* all the legitimate pronunciations in each character (one may call this as a rule system) rather than a phonological system, such that it activates all the localist units for all the legitimate pronunciations in each character. In contrast, the semantic layer becomes a *system which selects* the correct one of these localist units by boosting its activity towards 1 whereas inhibiting the other legitimate localist units. As a result of this additional positive/negative inputs from the semantic layer, the output is scored as a correct word reading in the max-criterion. Then, shutting off the inputs from this selecting system (a procedure to test nonwords) leaves one of the legitimate localist units to be more active

than the rest of non-legitimate localist units, resulting in correct nonword reading (as explained above, kanji nonword reading is scored correct if each character is pronounced with any of the legitimate pronunciations). Taken together, when *combining* these techniques, a higher Japanese nonword reading would be an artifact, not due to the acquisition of the statistical structure underlining kanji reading. So, we removed the semantic layer and focused on the demonstration of the sequential nature of that model.

When the network in Figure 1 was trained for reading a 2-kanji ideogram, then two units in the kanji input layer were activated, each of which represented each of the two kanji characters, respectively. These two ‘on’ units were maintained during the two time steps in a trial. In the first time step, the network activated the output localist unit for the correct pronunciation of the 1st kanji character (single/double morae), and then in the 2nd time step, it activated another localist unit for the 2nd kanji character pronunciation. When the same word was trained in kana form, then the position-specific localist units for kana characters were activated during the two time steps, and the network was trained to produce the same output patterns as those of the corresponding kanji word. When the kana-length (thus mora length) was 2 or 3, it was centred in the kana input layer in order to minimize the dispersion problem (Plaut et al., 1996). Thus, importantly, the model had to inhibit the mora(e) localist unit for the 2nd character during activating the unit for the 1st character in time 1, and vice versa in time 2.

Given the input pattern was stable during the two time steps in a trial, the only mechanism that allowed this sequential on/off shift in the output layer was the time-specific signal from the position layer. The position layer

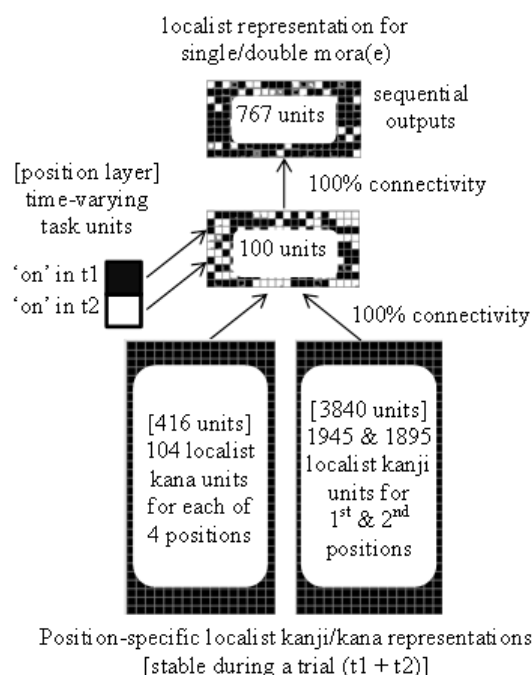


Figure 1. Architecture of the sequential output model with a position layer, adapted from Ijuin et al. (2002).

consisted of two *task units*, which alternated its ‘on/off’ status in time 1 and 2. The connections from these two units to the hidden layer were trainable. Therefore, it is crucial to analyse these connections in order to understand the nature of computations, which we did in the present study.

The 2-kanji ideograms for training were extracted from NTT database (Amano & Kondo, 2000). We first excluded both the archaic kanji characters and the daily characters yet with an archaic pronunciation. *Jouyou-list by Ministry of Education, Culture, Sports and Technology* specifies the daily kanji characters and daily pronunciations for teaching in school. The training words were made up of two kanji characters with daily pronunciations in the Jouyou-list. The words were selected only if the word frequency and familiarity were available. Words with a loan sound were removed. Also, mora-length in each kanji character was limited to one/two to minimize the dispersion problem. In other words, the maximum mora-length (kana-length) of a word was 4, and kanji-length was always 2. The resultant 41,215 kanji 2-ideograms and their kana-forms (i.e., in total 82,430 input patterns) were used for training.

Lens (<http://tedlab.mit.edu/~dr/Lens/>) was used for training. Learning rate was set to 0.005. The root-squared word frequency count per million was used to scale the error derivatives. Weight decay was not used in Ijuin et al. (2002). Cross-entropy was used to estimate the error, and the connection strength was adjusted through the back-propagation algorithm after every trial. No adjustment was made if the target-output difference was within 0.2. Momentum parameter was set to 0.9. An output was scored as correct if the most active localist unit was that for the target pronunciation for real words (max-criterion). Kanji nonword reading was scored correct if the most active unit was one of the legitimate pronunciations in each kanji character. During testing, 120 kanji words and 120 kanji nonwords (Fushimi et al., 1999) were used. Their corresponding kana input patterns were used for assessing kana words and kana nonwords accuracies. The activity of each unit was a sigmoid function of the summed weighted input from other units.

Results and Interim Discussions

After 250 epochs of training (1 epoch = 82,430-word presentation), the network accuracy was equivalent to humans (Fushimi et al., 1999). Accuracies were 91.66% for 120 kanji words (93.66% for humans), 93.33% for 120 kanji nonwords (88.79% for humans), 100% for 120 kana words, and 92.5% for kana nonwords (there is no human kana reading data, measured with the same item set).

As explained above, the only mechanism that allowed sequential outputs from a stable 2-kanji input was the signal from the position layer (two task units). The amount of the inputs from these task units in each time step is mathematically equal to the connection strengths from each of the task units to the 100 hidden units. Left half of Figure 2 visualizes these. We plotted the connection weights from

the 2nd task unit (i.e., task unit input in time 2) to each of the 100 hidden units against the weights from the 1st task unit (i.e., task unit input in time 1). The resultant 100 plots in Figure 2a are color-/shape-coded in terms of the spatial position they occupy in Figure 2b (later in detail). In Figure 2a, there was a clear negative correlation between the task unit inputs from the position layer in Time 1 and those in Time 2 [$r_{(98)} = -.74$, $p < .05$, 95% $CI = -82$:-64]. In other words, this hardwired position layer separated the hidden units at least into two (/more) groups in terms of their ‘on/off’ status in Time 1 and 2. The functions of these two groups are more clearly visible if we add the bias input values to these connection weights (Figure 2b) because the resultant scatterplot means the default activity of hidden units without an orthographic input in Time 1 and 2, respectively. As a consequence, the hidden units were clustered into five. Blue asterisk markers denoted the hidden units that were “allowed” to be active in Time 1 upon orthographic inputs but were “forced” to be inactive in Time 2 by the negative input from the task unit (position unit 2). In contrast, red triangle makers denoted the opposite. It is natural to expect that the first group is specialized (i.e., modular) only for the pronunciation of a 1st kanji character (and corresponding kana characters) whereas the second group is for a 2nd kanji character. This can be tested by measuring character reading accuracy after ‘lesioning’ each unit group selectively. As a result, Figure 3 shows double dissociations of the 1st kanji character (and corresponding kana) specific deficit and the 2nd kanji character specific deficit by lesioning each group selectively. The other hidden units seemed to be involved in both characters.

In summary, this model separated two distinct mechanisms as DRC/CDP models (Coltheart et al., 2001; Perry et al., 2007). One (asterisks & triangle markers) was specialized for a character-position-specific (thus script-specific) sequential parsing mechanism, and the other (remaining markers) involved for whole-word processing. The second one should be helpful for the network to read inconsistent-atypical words. In Simulation 2, we provide an alternative PDP model for kanji and kana reading without a sequential component.

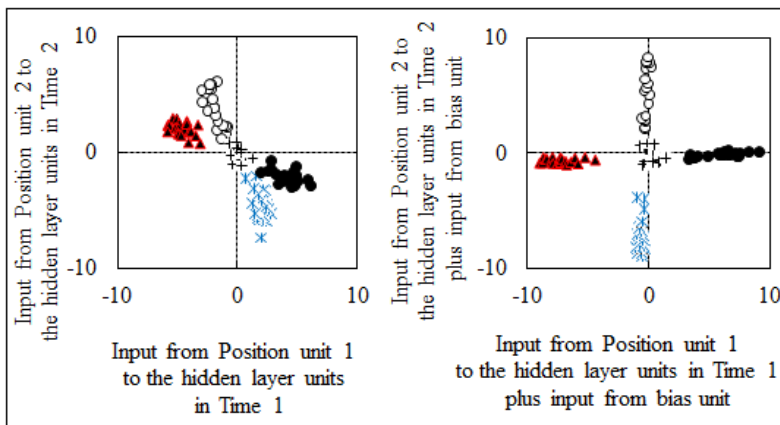


Figure 2. Input from the position units (left) plus bias unit (right)

Simulation 2 (Fully Parallel Output Model)

Methods

Figure 4 shows the architecture of the network. The same 82,430 words as Simulation 1 were used for training. This time, the input/output patterns were distributed representations for kanji/kana characters and for phonetic distinctive features of morae (Halle & Clements, 1983), respectively (Figure 5). The input patterns for kana words with 2 or 3 morae in length were centred to minimize the dispersion problem. Specifically, the kana character that corresponded to the first pronunciation of the 2nd kanji character (when it is written in kanji) was always placed at the 3rd character position of the kana layer (Figure 5). The output patterns were also centred in the same way.

The output of the network was fully parallel, such that there was only one time step in a trial. The distributed inputs for 2 kanji characters (or corresponding kana inputs) were presented simultaneously, and the correct distributed representations for all the morae patterns were produced in the output layer. The semantic system (Plaut et al., 1996) sent an additional input (S , below) to move the output layer activations towards the correct target value (0 or 1).

$$S = \frac{3.2 * \log(\text{frequency per million} + 2) * \text{training epochs}}{\log(\text{frequency per million} + 2) + 50}$$

S parameter was set to 0 when testing for nonword reading, following the past studies (Ijuin et al., 2002). One might say this is not a validated procedure, but our pilot study confirmed this model was able to read nonwords satisfactorily even when trained without this semantic input. The 20-unit output pattern per mora (one row of the output layer in Figures 4-5) was compared to all the possible 104 morae patterns, and the closest mora pattern was selected as its output in that mora position. Momentum was not used. The error derivative was set to zero when the target-output difference was below 0.1. Learning rate and weight decay started from 0.5 & 5-E07, respectively, and were reduced by 0.1 & 1-E01, per 5 epochs from the 31st epoch. After the 50th epoch, each parameter was set to 0.01 & 0, respectively

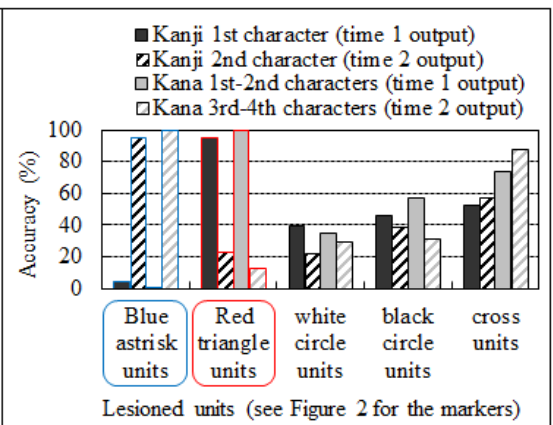


Figure 3. Accuracy upon selective lesioning

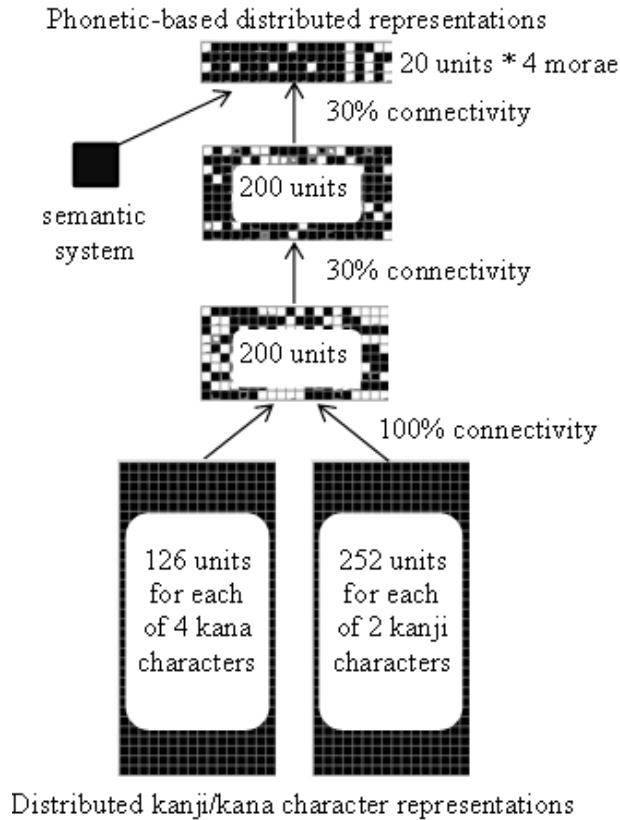


Figure 4. Architecture of a fully-parallel output model

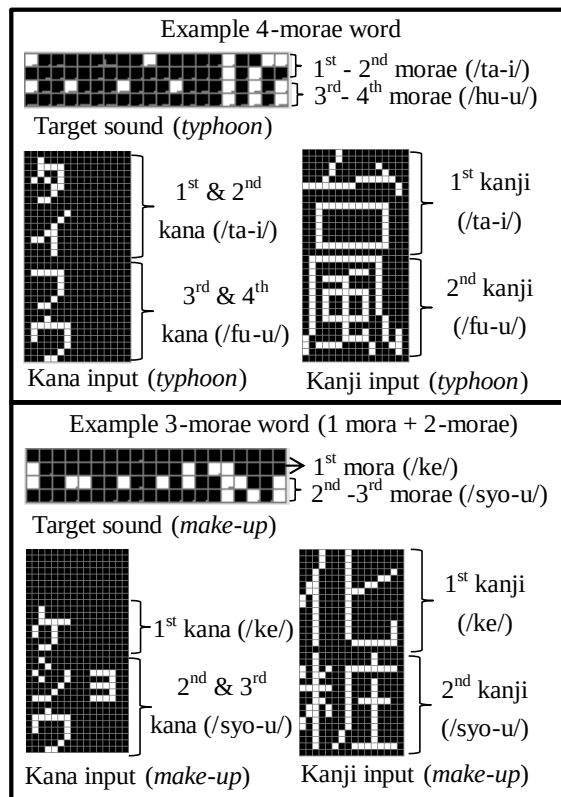


Figure 5. Examples of input-output mapping

for further 5 epochs training. The other parameters were the same as Simulation 1.

Results and Interim Discussions

After 55 epochs of training, accuracies were 100% for 120 kana words, 98.33% for kana nonwords, 95% for 120 kanji words, and 81.65% for 120 kanji nonwords. Kanji nonword accuracy was lower than humans (89%) in Fushimi et al. (1999). However, as explained above, human kanji nonword reading accuracy can be exaggerated if a letter-by-letter reading strategy is taken. Importantly, Fushimi et al. (1999) used a pure kanji nonword list. Testing with a pure kanji nonword list allows participants to know they do not need to refer to the adjacent character to generate a legitimate pronunciation for each character, leading to letter-by-letter reading. In fact, our replication of Fushimi et al. (1999) showed that testing kanji nonword in a pure list overestimated accuracy by 10% (Ikeda, Ueno, Ito, Kitagami, & Kawaguchi, In preparation). Testing with a mixed list of kanji words (necessary to refer to the adjacent character for computing word-specific information) and nonwords in a block yielded a more modest nonword reading (81.57%, 95%CI = 77:86). Note that the network also had no way to know if the next stimulus was a word or nonword. Therefore, our network accuracy was equivalent to humans if compared in a fair manner.

In summary, the current connectionist model had no sequential mechanisms yet acquired the statistical structure underlying kanji/kana reading, and generalized it to nonwords. Thus, there is no necessity to incorporate dual mechanisms for simulating kanji nonword accuracy. The first attempt to generalize PMSP96 into Japanese was as successful in simulating word reading (Ijuin et al., 1999) as SM89. However, one could argue it only approximated lexical lookup process (Coltheart et al., 2001). The authors naturally attributed it to the coding scheme and the corpus size (Ijuin et al., 1999), but they later concluded these factors did not improve performance, and attempted to incorporate a sequential mechanism (Ijuin et al., 2001, 2002). In this study, however, we only improved some parameters (weight decay, corpus size, and generalizable input/output coding schemes) compared to Ijuin et al. (2001, 2002), and demonstrated successful kanji nonword reading even without incorporating a sequential mechanism.

General Discussion

Nonword reading accuracy has been one of the test cases for the necessity of a grapheme/character-based sequential component (Coltheart et al., 2001; Perry et al., 2007; Plaut et al., 1996). Plaut et al. (1996) has shown such a component is not necessary to simulate English nonword reading. Further insights can be gleaned by cross-linguistic extension (Perry et al., 2007). However, the Japanese pioneering works were successful to simulate kanji nonword reading by implementing a character-based sequential processing (parsing) mechanism (Ijuin et al., 2001, 2002),

like the dual-mechanism models (Coltheart et al., 2001; Perry et al., 2007). In contrast, our model did not implement a sequential component, but nonetheless acquired the statistical knowledge underlying Japanese kanji reading, rather than approximating lexical lookup process, and generalized this knowledge to nonword reading. Thus, we provide first cross-linguistic evidence from Japanese to support a single-mechanism theory that postulates parallel-processing in converting kanji/kana orthography to morae representations of words/nonwords.

Note that the position layer in Ijuin et al. (2001, 2002) is not the same as that in Plaut (1999). Plaut (1999) has extended PMSP96 to sequential processing, but the signal from his 'position layer' was an approximation of an attentional fixation mechanism (i.e., domain-general). Specifically, that signal was not sequential from left to right per a character/grapheme (i.e., reading-specific).

Finally, other behavioural phenomena, such as word-length effect (Weekes, 1997) have been taken as evidence for sequential mechanisms (Coltheart et al., 2001; Rastle, Havelka, Wydell, Coltheart, & Besner, 2009), and a similar argument was made in Japanese literature (Tamaoka, 2005) to support the kanji-character based sequential processing (parsing) mechanism (Ijuin et al., 2002). However, Chang, Furber, and Welbourne (2012) successfully simulated the length effect by a PDP model, and attributed this effect to the peripheral (visual/articulatory) processing, rather than to the translation process itself from orthography to phonology. Our ongoing modelling project (Ikeda et al., In preparation) has already simulated the length-by-lexicality interaction in Japanese kana reading, reported by Rastle et al. (2009), and other crucial phenomena, such as frequency-regularity interaction in kanji words/nonwords (Fushimi et al., 1999), and surface dyslexic kanji reading upon damage in the semantic system (Plaut et al., 1996). Taken together, like English (Plaut et al., 1996), evidence from Japanese reading literatures supports parallel-distributed processing for reading aloud despite of the clear cross-linguistic differences in orthography and phonology.

Acknowledgments

This study was supported by Grant-in-Aid for JSPS (Japan Society for the Promotion of Science) Fellows to T. Ueno.

References

- Amano, S., & Kondo, T. (2000). *Japanese NTT database series: Lexical properties of Japanese, word-frequency*. Tokyo: Sanseido.
- Besner, D., & Smith, M. C. (1992). Models of visual recognition - When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18(3), 468-482.
- Chang, Y.-N., Furber, S., & Welbourne, S. (2012). "Serial" effects in parallel models of reading. *Cognitive Psychology*, 64(4), 267-291.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204-256.
- Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. F. (1999). Consistency, frequency, and lexicality effects in naming Japanese Kanji. *Journal of Experimental Psychology-Human Perception and Performance*, 25(2), 382-407.
- Glushko, R. J. (1979). Organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology-Human Perception and Performance*, 5(4), 674-691.
- Halle, M., & Clements, G. (1983). *Problem book in phonology*. Cambridge, MA: MIT Press.
- Ijuin, M., Fushimi, T., Patterson, K., & Tatsumi, I. (1999). A connectionist approach to Japanese kanji word naming. *Psychologia*, 42(4), 267-280.
- Ijuin, M., Fushimi, T., & Tatsumi, I. (2001). *Kanji nonword reading by a connectionist model [written in Japanese]*. Paper presented at the The 4th Annual Meeting of Japanese Cognitive Neuropsychology [dai 4 kai ninchi shinkei shinrigaku kenkyuukai].
- Ijuin, M., Fushimi, T., & Tatsumi, I. (2002). Surface dyslexia in Japanese - A computational account from simulation study [written in Japanese]. *Japanese Journal of Neuropsychology*, 18, 101-110.
- Ikeda, K., Ueno, T., Ito, Y., Kitagami, S., & Kawaguchi, J. (In preparation). Parallel vs serial issues in reading aloud: Behavioral and computational approach to explain nonword reading and lexicality by length interaction in Japanese.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114(2), 273-315.
- Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, 23(4), 543-568.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol Rev*, 103(1), 56-115.
- Rastle, K., Havelka, J., Wydell, T. N., Coltheart, M., & Besner, D. (2009). The Cross-Script Length Effect: Further Evidence Challenging PDP Models of Reading Aloud. *Journal of Experimental Psychology-Learning Memory and Cognition*, 35(1), 238-246.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568.
- Tamaoka, K. (2005). Is an orthographic unit of a single Japanese kanji equivalent to a kanji phonological unit in the naming task? [written in Japanese]. *Cognitive Studies*, 12(2), 47-73.
- Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 50(2), 439-456.