A CONNECTIONIST APPROACH TO NAMING DISORDERS OF JAPANESE IN DYSLEXY PATIENTS

Mutsuo Ijuin*, Takao Fushimi*, Karalyn Patterson**, Naoko Sakuma*, Masayuki Tanaka***, Itaru Tatsumi*, Tadahisa Kondo**** and Shigeaki Amano*****

* Department of Language and Cognition, Tokyo Metropolitan Institute of Gerontology
35-2 Sakae-cho, Itabashi-ku, Tokyo, 173-0015, JAPAN

** Medical Research Council, Cognition and Brain Sciences Unit
15 Chaucer Road, Cambridge, CB2 2EF, England, U.K.

*** Primate Research Institute, Kyoto University
Inuyama, Aichi, 484-8506, JAPAN

**** NTT Communication Science Laboratories
3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 243-0198, JAPAN

ABSTRACT

The triangle model is a computational model for lexical processing which computes word orthography, phonology and semantics using the architecture of a parallel distributed processing network. The computation takes the form of interactions among neuron-like processing units. In the present research, a Japanese triangle model computed phonology directly from orthography for both Kanji and Kana strings, with additional input to phonology from a component representing putative word semantics. This model successfully simulated certain effects seen in the performance of Japanese skilled readers. Moreover, different types of damage to the model reproduced data on both the surface and phonological forms of acquired dyslexia.

After damage to the semantic component, the model’s reading performance remained good for Kana words, Kana nonwords, and Kanji words with consistent character-sound correspondences, but was significantly impaired on Kanji words with atypical correspondences: this simulates surface dyslexia. After damage to the phonological component of the model, the network’s performance remained good for both Kanji and Kana words but was impaired on Kana nonwords: this simulates phonological dyslexia.

These results are basically comparable to those of previous models developed for English, and thus demonstrate that the same computational principles of the triangle model can be applied to alphabetic and non-alphabetic writing systems. Mechanisms and properties of the model for Japanese are discussed.

1. INTRODUCTION

In the domain of single-word processing, theoretical models have an architecture composed of several relatively independent subsystems and their interconnections, which attempt to describe how one code or representation is computed from another. In models of reading, there has been particular focus on the computation of phonology (the word’s pronunciation) from orthography. Amongst such approaches, the dual-route model (e.g., Coltheart et al., 1993) has been particularly influential. The dual-route model employs two different procedures to produce pronunciations from written letter strings. One (the sublexical procedure) assembles typical pronunciations by the application of a set of grapheme-phoneme correspondence rules. The other (the lexical procedure) is a routine in which familiar orthographic strings activate whole-word pronunciations in the lexicon. The two procedures are adept at processing different types of letter strings. The sublexical procedure produces correct pronunciations for regular words with typical pronunciations of the word’s components (e.g., MINT, cf. HINT, LINT, PRINT) and reasonable pronunciations of nonwords, but incorrect pronunciations for irregular or exception words that break standard grapheme-phoneme correspondence rules (e.g., PINT). On the other hand, the lexical procedure produces correct pronunciations for all words but has difficulty with any letter string not represented in the lexicon, i.e. nonwords. This model has been extensively explored in interpreting both reading processes of normal adults and patients with different types of acquired dyslexia (e.g., Berndt et al., 1996; Coltheart et al., 1993).

The triangle model, based on Seidenberg & McClelland (1989) and Plaut et al. (1996)

An alternative to the dual-route model views lexical processing from a connectionist (or PDP: parallel distributed processing) approach (e.g., Seidenberg & McClelland, 1989). Figure 1 shows an example of this approach, called “the triangle model”. This is an artificial neural network model designed to compute word orthography, phonology, and semantics. The
computation takes the form of interactions among neuron-like processing units. Unlike the dual-route model, there are no separate mechanisms or procedures that operate on fundamentally different principles: all letter strings are processed in the same architecture by the same common principles of computation. English versions of the triangle model have successfully simulated the oral reading performance of normal adults (e.g., Seidenberg & McClelland, 1989; Plaut et al., 1996) and also, when the model is damaged after training, of dyslexic patients (surface dyslexia: Plaut et al., 1996, phonological dyslexia: Harm & Seidenberg, 1999, deep dyslexia: Hinton & Shallice, 1991; Plaut & Shallice, 1993).

It is important to determine whether a similar approach can be successfully applied to a non-alphabetic writing system. The present research was designed to simulate skilled word naming and two different patterns of acquired dyslexia in Japanese using the connectionist approach. Japanese orthography comprises two scripts: Kanji and Kana. Kanji are morphographic characters of Chinese origin; the majority have two or more completely different pronunciations, with the correct pronunciation for a character in any particular word determined by intra-word context (i.e., the specific combination of characters in that word). Kana, on the other hand, are phonographic characters for morae (the major units of spoken Japanese) with highly consistent orthography-to-phonology mappings. To apply the connectionist approach to the naming of Japanese Kanji and Kana words, we built a triangle network based on Plaut et al. (1996). This network contains no specific procedures depend on either script type (Kanji or Kana) or lexical status (word or nonword strings): all orthographic input is processed in the network by the same common computational principles. First of all, we will determine whether the intact trained network can simulate the performance of normal adult readers to name both words and nonwords. Then, we will focus on simulating (i) the pattern of surface dyslexia in Japanese by damaging the semantic component of the network, and (ii) the pattern of Japanese phonological dyslexia by damaging the phonological processing of the network.

2. THE MODEL

2.1. Training Corpus

As a training corpus, we used 5,078 two-character Kanji words and 2,013 one- to four-character Kana (Katakana) words. In this word corpus, each Kanji word corresponded to 4 morae of spoken Japanese; each Kana word comprised 1 to 4 morae. All words had familiarity values ranging from 3.03 to 6.75 on a 7-point scale (Amano & Kondo, 1999).

2.2. Representations

In the current implementation, the representation of orthography was a 16 x 16 grid matrix pattern for each Kanji character and a 8 x 16 grid for each Kana character (see Figure 2). A total of 512 units (Kanji : 16 x 16 grid x 2 characters; Kana : 8 x 16 grid x 4 characters) was required to represent the target word.

In phonology, the pronunciation for each word was represented by an ordered sequence of phonemes. Morae comprise 108 distinct examples and can be classified into 5 types: CV, CCV, V, nasal coda /N/ and geminate consonant /Q/ (Otake et al., 1993). Since the majority of morae correspond to CV pairings (e.g., /ka/,/sa/,/ta/), consonant units (N=26) and vowel units (N=5) were prepared for representing a single mora: the network's response /ka/ can thus be represented as the activation of /k/ and /a/ units without order constraints because VC or VCC pairings do not occur in Japanese morae. Each consonant unit represents either C (e.g., /k/, /s/, /t/) or CC (e.g., /kj/, /tj/, /zj/) or an empty phoneme (for which we used @; this is required for representing a single-vowel mora, e.g., /@a/ for /a/). In addition, one geminate consonant unit and one nasal coda unit were prepared. Thus, a total of 33 units was required to represent a single mora slot. When all 33 units were in the state 0.0, the mora slot was regarded as empty. Table 1 illustrates the phonological representation of the word /reNsai/. A total of 132 (33 phonemes x 4) units was required to represent the pronunciation of both Kanji and Kana words.

| Table 1: An example of phonological representation for the word /reNsai/ |
|---|---|---|---|
| 26 consonants | 5 vowels | geminate nasal coda |
| /n/ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| /s/ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| /t/ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| /r/ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| /a/ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |

2.3. Network Architecture

The network’s task is to compute the pronunciations of Kanji and Kana words directly from their written forms, supplemented by an additional input intended to stand for word meaning (semantics). Figure 2 shows the architecture of the network. When the orthographic pattern of a word is presented, the network produces its pronunciation as an activity pattern over the output layer. To produce the phonological output pattern, the network needs three pathways: (i) O → P (Orthography to Phonology), (ii) P → P (Phonology to Phonology), and (iii) S → P (Semantics to Phonology).

![Figure 2: The network architecture. Each oval in bold represents a group of units (layer) and each circle represents an individual processing unit. Each arrow represents a set of connections from one layer to another. The parts with dashed lines are not implemented in this simulation.](image-url)
In the P \( \rightarrow \) P pathway, all of the phonological units were connected to each other directly and to a set of 50 cleanup units. These cleanup units receive activation from and send activation back to the phonological units; they permit the encoding of higher-order phonological dependencies than those achieved by direct P \( \rightarrow \) P connections alone (Hinton & Shallice, 1991, Plaut & Shallice, 1993).

For the S \( \rightarrow \) P pathway, rather than attempting the major enterprise of implementing a real semantic layer, we followed the “putative semantics” procedure of Plaut et al. (1996). S \( \rightarrow \) P is the key component in the processes of O \( \rightarrow \) S \( \rightarrow \) P or O \( \rightarrow \) P \( \rightarrow \) S \( \rightarrow \) P. Plaut et al. characterized the process of S \( \rightarrow \) P in terms of an additional source of activation to the phonological response of the network. During training, the putative semantics procedure was used to provide additional input — increasing as a function of both word frequency and training epoch — that pushed the phonological units toward their correct activations.

### 2.4. Training Procedures

We trained the network to name Japanese Kanji and Kana words by the following three phases. Like a real human child, the network first learned about the phonological characteristics of words before any orthographic information was introduced. In the P \( \rightarrow \) P pathway, for all words in the training corpus, phonological patterns were trained to reproduce themselves. After this mapping had been learned, the connection weights in the P \( \rightarrow \) P pathway were frozen and not allowed to change during the following training phases. In the second phase, the model was trained to read aloud only Kanji words; this represents the real experience of Japanese children who learn Kana before Kanji. The third phase included training for the naming of both the Kanji and Kana words in the corpus. Once again based on the experience of real Japanese children, if the network produced an incorrect pronunciation to a Kanji word, it was given training on the Kana transcription of the Kanji word. Putative semantic input operated during all training phases.

In the network, each unit has an activity level or state and the states of units change gradually over time. The state of unit \( j \) at time \( t \), \( x_j(t) \), is a weighted average of its state at time \( t-1 \) and the state dictated by its summed input:

\[
x_j(t) = \sum_i x_j(t-1)w_{ij} + b_j = \sum_i x_i(t-1)w_{ij} + b_j = \sum_i w_{ij}x_i(t-1) + \beta_j x_j(t-1) + (1 - \tau)x_j(t-1)
\]

where \( w_{ij} \) is the weight of the connection to unit \( j \) from unit \( i \), \( b_j \) is the real-valued bias of unit \( j \), \( \beta \) is the weighting proportion by which the states of units change (Seidenberg et al., 1994), and \( \beta(t) \) is a nonlinear logistic function.

The network was trained by back-propagation through time. This procedure calculated the appropriate amount of weight change, resulting in a gradual reduction in the discrepancy between the phonological output pattern generated by the network and the “desired” (correct) pattern for the target word. We used cross entropy (CE: Hinton, 1989) as the measure of this discrepancy:

\[
CE[t] = -\sum_j t_j \log s_j(t) + (1 - t_j) \log(1 - s_j(t))
\]

where \( t_j \) is the desired state of phonological unit \( j \).

There are two ways to modulate weight change according to the frequency of the words in the training corpus. Seidenberg & McClelland (1989) presented each word to the network with a probability proportional to its frequency (transformed by a logarithmic function) and the weights were updated after each sweep. By contrast, Plaut et al. (1996) used frequency values to scale the amount of weight change. Thus, error derivatives calculated by back-propagation were modulated in proportion to word frequency. This means that more frequent words engendered greater change in the connection weights during training (Plaut et al., 1996). We adopted the latter procedure in our training. The scaling value for each word was determined by dividing its familiarity (transformed by an exponential function) by the highest value in the training corpus. The values ranged from 1.0 to 0.02, with a mean across all the training words of 0.29.

### 2.5. Network Performance

After 200 training epochs, the network correctly pronounced 5,074 of the Kanji words (99.9% of 5,078 words) and 2,006 of the Kana words (99.7% of 2,013 words) in the training corpus. The network misread 11 words, nine of which were of low familiarity.

We evaluated the network’s performance in naming nonwords: these were created from 80 words that were selected among the training corpus. Kanji nonwords were created by randomly recombinating the character at the first position of one word with the character at the second position of a different word (e.g., 瑠分 /oya bu/: meaning a boss, 物音 /mono oto/: meaning a sound \( \rightarrow \) 音無: meaningless nonword). Each Kanji nonword is pronounceable, but most have several different plausible pronunciations. We considered the response of the network to a Kanji nonword as correct if the pronunciation of its component characters matched any legitimate pronunciation in the training corpus. The network achieved very poor nonword performance, scoring only 4%.

Kana nonwords were created by randomly recombinating the two characters at the first and second positions of a four-character word with the two characters at the third and fourth positions of a different four-character word (e.g., 酸ハプ /pu re ha bu/: meaning a prefab, ウィルプリ /i ru su/: meaning a virus \( \rightarrow \) プルウィ: meaningless nonword). Each Kana nonword is not only easily pronounceable but has a unique correct pronunciation. The network produced correct responses to 80% of the nonwords.

One possible reason for the poor performance in naming Kanji nonwords is the size of the training corpus. Seidenberg & McClelland (1990) pointed out that the network’s performance on a particular nonword (e.g., MAVε) is highly dependent on the number of orthographically similar words in the training corpus (e.g., SAVE, GA VE, MAKE). Because the network has not been trained on the whole nonword, the computation of its pronunciation should be based on knowledge of words sharing sub-word sized orthographic components with the nonword (e.g., -AVE). In Kanji, there are a large number of characters: at least 3,000 characters appear in daily newspapers or magazines (Sasanuma, 1986). Our training corpus comprising 5,078 words contained over 1,000 characters, and therefore, the number of other characters co-occurring with a particular character is small. The small neighbourhood on which the network can rely in pronouncing nonwords might be responsible for its poor Kanji nonword performance (Ijuin et al., 1999). By contrast, Kana comprises only about 70 characters and the number of characters co-occurring with a particular character is large. Thus, Kana nonwords have many neighbours in the training corpus and the performance on Kana nonwords is much better than that.
of Kanji. In the following section, we focus on the network’s performance in naming Kanji words, Kana words, and Kana nonwords (including pseudohomophones), but not Kanji nonwords.

3. SIMULATION 1

Normal Adults

Before turning to simulations of the two types of dyslexia, it is important to establish that the network can reproduce more detailed characteristics of the performance of normal adult Japanese readers in naming words and nonwords. In this section, we will demonstrate that the network simulated familiarity and consistency effects in naming Kanji words, and lexicality and pseudohomophone effect for naming Kana strings.

Our training corpus contains Kanji and Kana words used in the word naming experiment by Fushimi et al. (2000). Figure 3(a) shows the response times of the network in pronouncing Kanji stimulus words with two different levels of character-sound consistency and two levels of word familiarity. In terms of consistency, Fushimi et al. classified Kanji words into two different types according to the consistency of the character-sound correspondences of the words’ constituent characters. A word was classified as consistent if each constituent character takes the identical pronunciation across all words with that character at the same position (e.g., 動送 /nN sou : meaning transportation). A word was called atypical if one or both characters have an uncommon or statistically atypical pronunciation in the target word (e.g., 書留 /kaki tome/ : meaning registration).

When computing the pronunciations for these words, the network (i) was faster at high familiarity words than low familiarity words, (ii) was faster at consistent words than words with atypical character-sound correspondences, and (iii) showed a significant consistency effect only for low familiarity words. This pattern of a familiarity-by-consistency interaction in the response time of the network to Kanji words is largely comparable to the naming latencies of Japanese skilled readers (Figure 4(b)).

Figure 3: Mean Kanji word naming latencies of (a) the network and (b) normal adults from Fushimi et al. (2000) for four word conditions: high-familiarity (High-Fam.) consistent, low-familiarity (Low-Fam.) consistent, high-familiarity atypical, low-familiarity atypical.

Figure 4: Mean Kana word and nonword naming latencies of (a) the network and (b) normal adults from Fushimi et al. (2000) for three conditions: words, pseudohomophones (PH), and nonwords.

In the O P pathway, words in the training corpus are faster than pseudohomophones and nonwords which were not trained. The S P pathway (putative semantics) also helps to compute the phonology of words and pseudohomophones, but not nonwords. These are the reasons why the network simulates lexicality and pseudohomophone effects in Kana.

4. SIMULATION 2

Surface Dyslexia

Surface dyslexic patients produce good (often normal) levels of performance when asked to read aloud regular (or consistent) words and nonwords, but their performance suffers in the reading of low frequency exception words, to which the patients often produce regularized pronunciations. Patterson & Hodges (1992) and Graham et al. (1994) reported this surface pattern of reading performance in a series of patients with semantic dementia and progressive fluent aphasia. The degree
of impairment to exception word reading was related to the severity of the patient’s comprehension loss. These authors concluded that deterioration of word meaning leads to surface dyslexia.

In keeping with this finding, damaging the $S \times P$ pathway and thus forcing the triangle network to rely on the intact $O \times P$ (and $P \times P$) pathway should simulate surface dyslexia. This outcome was confirmed for the English triangle model by Plaut et al. (1996), and we tested the same prediction for Japanese in this study.

After damage to the $S \times P$ pathway, the network’s performance in naming Kana was relatively well preserved (words: 100% $\times$ 100%, pseudohomophones: 86% $\times$ 58%, nonwords: 80% $\times$ 80%). However, the network lesioned in this way misread a substantial number of the Kanji words on which it had been trained (100% $\times$ 39%). Figure 5(a) shows the percent error of the semantically damaged network in pronouncing Kanji stimulus words with two different levels of character-sound consistency and two levels of word familiarity (using the same Kanji words in SIMULATION 1). The network especially mispronounced low-familiarity atypical words. Figure 5(b) indicates the percent error in Kanji reading by a Japanese surface dyslexic patient with progressive aphasia and impaired comprehension (Case NK: Patterson et al., 1995). Although the stimuli that were used in the experiment are different from ours (in Figure 5(b), high or low-frequency (High-Freq. or Low-Freq.) consistent, inconsistent-regular (Inc-Reg.), inconsistent-irregular (Inc-Irreg.), and exception words), the error patterns for Kanji words by the network and the patient reveal the same significant familiarity (frequency)-by-consistency interaction.

Figure 5: Mean results (% error) for (a) network and (b) surface dyslexia patient (Case NK from Patterson et al., 1995) in reading Kanji words with various degrees of consistency as a function of frequency.

A key feature of the connectionist account of surface dyslexia is a division of labor between the $O \times P$ and $S \times P$ pathways (Plaut et al., 1996). As described above, the extra source of input to the phonological units from putative semantics was gradually introduced to the network during the training epoch. The contribution from word meaning is particularly critical for processing of the words that generate the most error in the $O \times P$ pathway, namely low-familiarity atypical words. The error for these words is reduced by the additional boost to the correct pronunciation from putative semantics. By contrast, high familiarity words, Kanji words with regular or consistent character-sound correspondences, and Kana words – even though they also receive additional input from the semantic component – do not depend it because these words are learned quickly and easily by the $O \times P$ pathway alone. Therefore naming success on these words, and also on nonwords, is little affected by a lesion to semantics. This division of labor between the $O \times P$ and $S \times P$ pathways caused the lesioned network to replicate the basic pattern of surface dyslexia in Japanese.

5. SIMULATION 3

Phonological Dyslexia

Phonological dyslexic patients are severely impaired at reading nonwords aloud, but they succeed well in reading most real words, whether these have typical or atypical spelling-sound correspondences. These patients are also almost invariably impaired at various phonological tasks (such as phoneme segmentation or blending) that involve no orthographic processing (Coltheart, 1996). As a result, one hypothesis about phonological dyslexia is that it is not a specific reading disorder but rather results from disruption to phonological processing: with reference to the triangle model, an impairment within phonology itself (e.g., Patterson et al., 1996; Plaut et al., 1996).

One way in which the triangle network may simulate phonological dyslexia is the severing of connections between the phonological units in the $P \times P$ pathway (Harm & Seidenberg, 1999). The network’s performance was evaluated after 20 instances of this kind of damage, and the data were averaged across the effects of individual lesion. In terms of Kanji words, there was little effect of this damage to the network (100% $\times$ 91%). Figure 6(a) shows the lesioned network’s error performance on naming Kana strings (the same stimuli as in SIMULATION 1). Average performances on both Kana words and pseudohomophones were comparable to the performance of the intact network, whereas the lesioned network’s success on nonwords was reduced. Figure 6(b) indicates the percent error of a Japanese phonological dyslexic patient in naming Kana (Hiragana) strings with the same three attributes (words, pseudohomophones, nonwords) as our stimuli (Case KT: Patterson et al., 1996). The general patterns of performance by the phonologically damaged network and the Japanese phonological dyslexic patient are the same.

The computational processes of phonology in the current network consist of (i) $O \times P$, (ii) $S \times P$, and (iii) $P \times P$. If phonological dyslexia were caused by an impairment of (i), it would be difficult to explain the significant pseudohomophone
effect that characterizes the nonword naming of both the patient KT and our network. We hypothesize that this effect arises from the contribution of O ⊕ P ⊕ S ⊕ P, a proposal that is reinforced by the demonstration that patient KT’s advantage for pseudohomophones was significantly greater if these nonwords shared their pronunciations with imageable rather than abstract words (Patterson et al., 1996). In SIMULATION 2, damage to (ii) produced the misreading pattern of surface, not phonological dyslexia. The only remaining possibility is disruption to (iii), and this simulation demonstrates that poor nonword performance can be reproduced by damage entirely within the phonological system, by eliminating connections in the P ⊕ P pathway.

6. CONCLUSION

In the dual-route model, there are two procedures to pronounce letter strings: lexical and sublexical. The lexical procedure can produce correct pronunciations for all familiar words, but not for nonwords. The sublexical procedure can produce inappropriate pronunciations for regular or consistent words and nonwords, but mispronounces exception (atypical) words. In the dual-route account, surface dyslexia results when the sublexical procedure is intact but the lexical procedure is damaged, while phonological dyslexia is the outcome of an intact lexical procedure with a damaged sublexical route.

The triangle model has a different architecture from the dual-route model: all letter strings are processed in the same architecture by common computational principles. There are no subsystems specially designed for either script type (Kanji or Kana in Japanese) or lexical status (word vs nonword). Nevertheless, our model successfully simulated the reading performance (response times) of normal readers in naming Kanji words of varying familiarity and consistency, and also of Kana words, pseudohomophones and nonwords. Moreover, the model with lesions to specific connections (pathways) reproduced some of the principal characteristics of different types of acquired reading disorders in Japanese: surface dyslexia after damage to the S ⊕ P pathway and phonological dyslexia after damage to the P ⊕ P pathway.

In the testing phase, putative semantic input was included only for letter strings with the phonology of a familiar word (words and pseudohomophones), and not for nonwords. In a strict sense, this procedure violates the claim that all letter strings are processed with common computational principles. In practice, however, any input from the semantic layer (for example from semantic representations of words that resemble a target nonword orthographically or phonologically) seems unlikely to have much impact on the network's response to nonwords. This issue, plus the more important one of poor Kanji nonword naming by the current network, both remain to be addressed in future simulations. In other regards, however, the results from this study are comparable to those of previous connectionist models of reading developed for English (e.g., Plaut et al., 1996; Harm & Seidenberg, 1999), and demonstrate that the connectionist approach can be successfully applied to both alphabetic and non-alphabetic writing systems.

REFERENCES