

## Semantic effects in word naming: Evidence from English and Japanese Kanji

Naoki Shibahara

*University College London, London, UK and Otemon Gakuin University, Osaka, Japan*

Marco Zorzi

*University College London, London, UK and University of Padova, Padova, Italy*

Martin P. Hill

*Middlesex University, London, UK*

Taeko Wydell

*Brunel University, Uxbridge, UK*

Brian Butterworth

*University College London, London, UK*

Three experiments investigated whether reading aloud is affected by a semantic variable, imageability. The first two experiments used English, and the third experiment used Japanese Kanji as a way of testing the generality of the findings across orthographies. The results replicated the earlier findings that readers were slower and more error prone in reading low-frequency exception words when they were low in imageability than when they were high in imageability (Strain, Patterson, & Seidenberg, 1995). This result held for both English and Kanji even when age of acquisition was taken into account as a possible confounding variable, and the imageability effect was stronger in Kanji compared to English.

Most current models of visual word recognition and naming assume the existence of at least two processing routes for the pronunciation of written words. That is, upon presentation of a printed word, phonology is retrieved through a lexical–semantic pathway (or network) as well as assembled through a spelling–sound mapping process (see, e.g., the computational models of Coltheart, Curtis, Atkins, & Haller, 1993; Jacobs, Rey, Ziegler, & Grainger, 1998; Plaut, McClelland, Seidenberg, & Patterson, 1996; Zorzi, Houghton, & Butterworth, 1998b). Although the specific architectures and processing assumptions of these models may vary to a

---

Requests for reprints should be sent to Naoki Shibahara, Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK. Email: n.shibahara@ucl.ac.uk

We thank Chris Barry, Karalyn Patterson, Eamon Strain, and one anonymous reviewer for their helpful comments on an early version of this paper. The research was supported by a project grant (G9015838N) from the Medical Research Council to Brian Butterworth and a fellowship (042023) from the Wellcome Trust to Taeko Wydell.

large degree, there is a consensus view that the interaction between the two different sources of phonological information must be assumed to explain evidence from both normal and impaired readers (see Zorzi, in press, for a review).

One source of controversy, however, concerns the role of semantics in oral reading of single words. The dual-route model of reading (e.g., Coltheart et al., 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) assumes that the lexical route can be further divided into two processing routes (note that this distinction means that the dual-route model is in fact a three-route model). One, named the *direct lexical route*, is conceptualized as a direct link between orthographic and phonological word forms; the other, named the *lexical-semantic route*, is thought to be mediated by word meanings. In this lexical-semantic route the printed form accesses the semantic representation of that word, which in turn activates the corresponding phonology; therefore, this procedure has been usually referred to as *reading via meaning*. The distinction between a lexico-semantic route and a direct lexical (i.e., nonsemantic) route was first suggested by Schwartz, Saffran, and Marin (1980) in their case study of the acquired dyslexic patient WLP, who could read aloud words (including exception words) that she could not understand.

Most theorists have long assumed that the semantic route contributes little to word naming in skilled readers of alphabetic scripts (but see Patterson, Graham, & Hodges, 1994; Patterson & Hodges, 1992; Plaut et al., 1996), as, being indirect, it is slow in delivering a word pronunciation. Moreover, some theorists argued that reading is fundamentally phonological (e.g., Carello, Turvey, & Lukatela, 1992; Van Orden, Pennington, & Stone, 1990), because even tasks requiring access to meaning but where phonology was task irrelevant, such as semantic categorization, were found to be affected by the phonological characteristics of the stimuli (e.g., Van Orden, 1987; Van Orden, Johnston, & Hale, 1988; Wydell, Patterson, & Humphreys, 1993; see Frost, 1998, for a comprehensive review). However, Strain, Patterson, and Seidenberg (1995) have recently demonstrated that a semantic variable, imageability, can have an impact on naming of isolated words. In particular, they found that the imageability variable affected naming of low-frequency exception words (i.e., the words that usually yield the longest naming latencies), with low-imageability words yielding slower reaction times and more errors than high-imageability words.

The imageability effect is generally assumed to reflect semantic processing. For instance, imageability has been found to affect the reading performance of patients presenting with the deep dyslexic syndrome (e.g., Coltheart, Patterson, & Marshall, 1980). Most of these patients are more successful in reading concrete than abstract words (see Denes, Cipolotti, & Zorzi, 1999; McCarthy & Warrington, 1990, for reviews of acquired dyslexias). The prediction of Strain et al. (1995) that imageability would affect low-frequency exception words in the reading performance of *normal* readers was based on both neuropsychological and computational considerations. First, the hypothesis that correct exception word reading is dependent upon semantic representations was proposed by Patterson and Hodges (1992) to explain the neuropsychological association between semantic dementia and surface dyslexia. Many patients presenting with semantic impairments are also surface dyslexic (e.g., Funnell, 1996; Patterson et al., 1994; Patterson & Hodges, 1992), although the corresponding dissociation (i.e., intact reading in the presence of semantic deficits) has also been observed (e.g., Cipolotti & Warrington, 1995; Lambon Ralph, Ellis, & Franklin, 1995). Second, the connectionist (PDP) model of reading developed by Seidenberg and McClelland (1989) showed that the

computation of phonology from orthography (i.e., the phonological route) is least competent on exception words, in particular when they are of very low frequency. Thus, the pronunciation of low-frequency exception words would rely much more on the semantic route than would that of any other word types. This assumption was incorporated in a more recent version of the PDP model (Plaut et al., 1996, Simulation 4). Plaut et al. trained the orthography-to-phonology network with an additional external input that represented the contribution of a (putative) semantic pathway; furthermore, the magnitude of this external input increased in the course of training, to simulate an increased competence of the putative semantic pathway. Plaut et al. showed that in this version of the model the orthography-to-phonology network is relieved from mastering low-frequency exception words, because the correct output for these words is provided by the semantic pathway. They argued that, as the semantic pathway's competence improves, the phonological pathway becomes specialized for regular words; this results in a redistribution of labour between the semantic and the phonological pathways.

One aim of the present study was to establish whether the imageability effect found by Strain and colleagues (1995) could be the result of a confounding variable, age of acquisition (AoA). AoA has been shown to be a major factor affecting word-naming and lexical-decision latencies in skilled readers (Gerhand & Barry, 1998, 1999; Morrison & Ellis, 1995), as well as object-naming speed (Barry, Morrison, & Ellis, 1997; Ellis & Morrison, 1998; Morrison, Chappell, & Ellis, 1997; Morrison, Ellis, & Quinlan, 1992). The effect of AoA has been shown to hold when objective measures of the age at which different words are learned are used in place of subjective (rated) measures (Ellis & Morrison, 1998; Morrison et al., 1997). In all of these studies AoA showed the highest correlation with naming latency of any variable investigated. Crucially, there is a high negative correlation between imageability and AoA ratings. For example, Gilhooly and Logie (1980) found a correlation of  $-.713$  between AoA ratings and imageability of the 1944 words in their norms. Ellis and Morrison (1998) also found a correlation of  $-.578$  in a set of 220 line drawings of objects.

Nonetheless, the AoA and the imageability effects might have independent sources. In relation to the imageability effect in deep dyslexia, Saffran, Schwartz, and Marin (1976) and Jones (1985) proposed that concrete words have well-defined and context-independent semantic representations, whereas the representations of abstract words are more vague, as they often depend upon the surrounding context. Similarly, Plaut and Shallice (1993) simulated the imageability effect in their connectionist model of deep dyslexia on the basis of the assumption that concrete words have a richer semantic representation than abstract words.

On the other hand, theoretical proposals about the possible source of the AoA effect have focused on quite different mechanisms. In particular, it has been argued that AoA affects the development of phonological representations. Brown and Watson (1987) proposed that the phonological representations of early acquired words may be stored in unitary form, whereas the phonological representations of later acquired words may be more fragmentary in nature. Morrison and Ellis (1995) and Ellis and Morrison (1998) discussed the acquisition of phonological representations with reference to self-organizing neural networks (e.g., Kohonen, 1984). They suggested that patterns introduced at later points of training will develop less effective representations because early patterns are allocated most of the network's resources (i.e., nodes).

A different explanation, more compatible with the PDP framework, has been recently proposed by Ellis and Lambon Ralph (2000). Using a standard backpropagation network, they

showed that when items introduced early into training are then joined by later sets of items that are trained alongside them in a cumulative and interleaved manner, performance of the network continues to favour the early set. They also demonstrated that the advantage for early-acquired patterns cannot be explained simply in terms of differences between early and late sets in cumulative frequency of training. In all these proposals, however, the explanation of the AoA effect is clearly different from the explanation of the imageability effect. If the imageability effect truly reflects semantic processing, it should contribute to word naming independently of AoA. Therefore, we obtained AoA ratings for the experimental stimuli from a separate group of subjects and used the value as a covariate in the item analyses.

The other main aim of our research was to investigate whether semantic processes contributed in the same way to a quite different language and orthography, Japanese Kanji. The PDP model has been claimed to be a general architecture of language processing that can be applied to any language, including Kanji (see Fushimi, Ijuin, Patterson, & Tatsumi, 1999). Thus, the prediction of imageability effects in reading English should generalize to Kanji reading. Moreover, the logographic nature of Kanji might entail a more prominent role of semantics in reading than would alphabetic scripts such as English. Kanji and English differ in terms of their "orthographic depth", which is roughly the degree to which the pronunciation of a word can be derived from the pronunciation of its parts. Thus, Serbo-Croatian and Italian are shallow, whereas English, which is more irregular, is deeper (Frost, Katz, & Bentin, 1987; Lukatela, Popadic, Ognjenovic, & Turvey, 1980). Zorzi et al. (1998b) predicted that readers differentially weight assembled phonology (i.e., the pronunciation derived from a direct spelling-sound mapping) and retrieved phonology (i.e., the pronunciation derived from lexical access) in a way that reflects the relative transparency of the mapping between spelling and sound in a specific language. As Kanji is a deeper orthography than English, reading aloud will rely more on lexical access, and hence the effects of imageability should be greater.

## EXPERIMENT 1

### Imageability effects in English

In the first experiment we investigated whether imageability effects would appear mainly for low-frequency exception words, as previously found by Strain et al. (1995). Frequency, regularity, and imageability were orthogonally manipulated in the experiment. In their Experiment 1 Strain et al. found a main effect of imageability and a (nonsignificant) trend towards the critical three-way interaction only in the subjects analysis. Given this somewhat weak result, it was important to replicate the experiment and to clarify the issue of whether the imageability effect truly reflects semantic processing, by using AoA ratings for the experimental stimuli as a covariate in the item analyses.

### Method

#### *Participants*

A total of 20 native English speakers (1 male and 19 females) aged between 19 and 48 years took part in this experiment. All had normal or corrected-to-normal vision. Each participant was paid a small fee for participating.

### *Apparatus*

A Macintosh IIsx computer with the software Psychlab (Gum & Bub, 1988) was used to present word stimuli in the centre of a computer screen. Reaction times were measured by the computer via a microphone connected to a voice key. Vocal responses were tape recorded for later checking.

### *Stimuli*

All experimental stimuli were taken from Strain et al.'s (1995) Experiment 1 (see their Appendix A). There were two sets of 48 words for this experiment, one of regular words and the other of exception words. Each set contained an equal number of high-frequency and high-imageability words (Hf-Hi), high-frequency and low-imageability words (Hf-Li), low-frequency and high-imageability words (Lf-Hi), and low-frequency and low-imageability words (Lf-Li).

Strain et al. (1995) classified a word as regular on the basis of two criteria: (1) pronunciation was regular in terms of grapheme to phoneme correspondences, and (2) it belonged to a consistent orthographic body neighbourhood. Words were classified as exception if their pronunciation conflicted with that obtained through grapheme-phoneme correspondence rules (Venezky, 1970). The set of exception words excluded (1) any word in which the body contains an irregular grapheme-phoneme correspondence but is pronounced like the majority of the words in the body neighbourhood (e.g., *bold*, *gold*) and (2) orthographically strange words (e.g., *yacht*). Word stimuli were classified as high or low frequency in accordance with the Kucera and Francis (1967) norms. High-frequency words had frequency values greater than 70 per million, whereas low-frequency words had values lower than 30 per million. Imageability ratings, placed on a 7-point scale, ranged from 1 (low imageability) to 7 (high imageability). Words with ratings between 1 and 4.3 were classified as low imageability, whereas words with ratings between 4.9 and 7 were classified as high imageability. Within each condition, words were closely matched between high- and low-imageability groups for initial phoneme, number of letters, and log frequency. Where it was not possible to match for initial phoneme, words were matched according to the sound similarity of the initial phoneme.

All experimental words were also rated for AoA by a group of 15 independent participants. The procedure used to collect the AoA ratings closely followed that employed by Morrison and colleagues (Morrison & Ellis, 1995; Morrison et al., 1992). The mean AoA values were used as a covariate in the item analysis of reaction times (see Appendix 1).

### *Procedure*

Participants were tested one at a time in a quiet room. Each trial began with the presentation of a fixation dot in the centre of the screen for 500 ms. The fixation dot was then replaced by a target word, which remained visible until a response was made. The inter-trial interval for all blocks was 1 s. Participants were instructed to read the target words aloud as quickly and accurately as possible.

The complete set of 96 words was presented in a series of four blocks, with the order of blocks counterbalanced across participants. Each experimental block began with three starter items, resulting in 27 items in each block. The order of presentation of test items was randomized by the computer program for each participant. The experimental blocks were preceded by a practice block of 20 words to allow participants to familiarize themselves with task requirements. The four experimental blocks were run with a short break between each block. Upon completion of the four experimental blocks, participants were asked to rate all experimental word stimuli for imageability on a 7-point scale. These ratings closely matched those collected by Strain et al. (1995) in terms of classification of items to the high vs. low imageability groups for each condition.

## Results

The three starter items in each block were removed from the data analyses. One test item and its composite eight-way matched set were also excluded from the analysis due to misprint of the test item. Analyses of variance were carried out on median naming latency and arcsine transformed error data by both subjects ( $F_1$ ) and items ( $F_2$ ). Voice key errors were excluded from the analyses.

*Response times.* Table 1 shows correct mean response times (RTs) (in milliseconds) with the effect of imageability on naming latencies for each condition. The RT data were analysed with a three-way analysis of variance (ANOVA, Word Type  $\times$  Frequency  $\times$  Imageability). The main effect of frequency was significant,  $F_1(1, 19) = 29.02$ ,  $MSE = 4388.69$ ,  $p < .0001$ ;  $F_2(1, 80) = 31.18$ ,  $MSE = 1539.18$ ,  $p < .0001$ , showing shorter naming latencies for high-frequency words (552 ms) than for low-frequency words (606 ms). The main effect of imageability was also significant,  $F_1(1, 19) = 25.59$ ,  $MSE = 640.96$ ,  $p < .0001$ ;  $F_2(1, 80) = 5.16$ ,  $MSE = 1539.61$ ,  $p < .05$ , indicating shorter naming latencies for high-imageability words (570 ms) than for low-imageability words (589 ms). The main effect of word type, although significant only by items,  $F_1(1, 19) = 2.69$ ,  $MSE = 4329.00$ ,  $p < .12$ ;  $F_2(0, 80) = 4.40$ ,  $MSE = 1539.61$ ,  $p < .05$ , suggests that regular words (570 ms) yielded shorter naming latencies than exception words (589 ms). There were no significant two-way or three-way interactions.

The item analysis of the RTs was also carried out using AoA as a covariate. The main effect of imageability disappeared,  $F_2 < 1$ , with no significant interactions with this factor. There was a significant effect of AoA,  $t(79) = 2.96$ ,  $p < .005$ . Moreover, rated AoA was significantly correlated with rated imageability,  $r = -.44$ .

*Errors.* Due to the error rate of 1.8% (32 out of 1760 responses) and a lack of errors occurring outside low-frequency exception words, an ANOVA of error data was not conducted (see Table 1).

Naming errors were categorized as one of four types. Visual-phonological word errors occurred when the participant responded to the target word presented with either a visually or

TABLE 1  
Imageability effects in mean reaction times<sup>a</sup> and error rates<sup>b</sup> for each condition in Experiment 1

Condition		High frequency			Low frequency		
		High imageability	Low imageability	Effect	High imageability	Low imageability	Effect
Regular	RT	547	553	6	583	599	16
	SD	36.3	32.8		56.6	39.8	
	Error	0	0	0	0	0.05	0.05
Exception	RT	541	569	28	611	634	23
	SD	36.3	36.4		35.6	57.4	
	Error	0	0.05	0.05	0.45	1.25	0.80

<sup>a</sup>In ms.

<sup>b</sup>In percentages.

TABLE 2  
Error rates<sup>a</sup> by frequency and imageability for exception words and comparative error rates from Strain et al.<sup>b</sup>

Experiment		High frequency		Low frequency	
		High imageability	Low imageability	High imageability	Low imageability
1	Regularization	0	0	1.25	0.42
	Other	0	0.05	0.30	0.95
Strain et al. <sup>b</sup>	Regularization	1.25	0.40	1.25	14.60
	Other	0.80	2.10	0.80	1.25

<sup>a</sup>In percentages.

<sup>b</sup>Strain et al. (1995, Experiment 1).

phonologically similar word. Visual–phonological nonword errors were those in which the participant produced a nonword that was visually or phonologically similar to the target word. Regularization errors occurred when the participant named an exception word as if it were a regular word. The only other errors were voice-key errors and were classified as such. A breakdown of error classes showed that 55% of the total errors, (39 words) were due to voice-key errors. For the remaining 32 errors, 30 occurred when naming low-frequency exception words (9 word errors, 13 nonword errors, and 8 regularization errors), and one nonword error occurred with both low-frequency regular and high-frequency exception words. The distribution of errors for exception words is shown in Table 2, where the data are shown in comparison to the original Strain et al.'s (1995) data.

## Discussion

The results of Experiment 1 failed to show the crucial three-way interaction between frequency, regularity, and imageability. It must be noted, however, that the three-way interaction was not significant in the original study either: Strain et al. (1995) could only find a trend in the subjects analysis. However, similar to Strain et al., we found a rather large main effect of imageability. This main effect was not discussed by Strain and colleagues, but the imageability effect was 19 ms for high-frequency regular words and 14 ms for low-frequency regular words in their Experiment 1, a result that was certainly not predicted by the PDP model, because in this model, the direct orthography-to-phonology computation is very accurate and efficient for high-frequency regular words, and therefore the reading of such words does not depend on the semantic contribution. In the present study, however, the main effect of imageability disappeared when AoA was used as a covariate in the item analysis, suggesting that the imageability effect was the result of a confound with AoA. The finding that naming times are faster for words learned early than for words learned late in life is in agreement with a number of previous studies that investigated the effects of AoA (e.g., Gerhand & Barry, 1998; Morrison & Ellis, 1995). High-imageability words are likely to be acquired earlier in life than low-imageability words. This was confirmed by the highly significant negative correlation

between imageability and AoA ratings,  $r = -.44$ , that we found in the set of experimental stimuli.

However, the disappearance of the main effect of imageability when AoA is accounted for does not rule out the possibility of a residual effect of imageability on low-frequency exception words. The stimuli in Experiment 1 did not produce the critical three-way interaction, leaving space for the possibility that it might appear with a different set of stimuli. To this end, we replicated in the following experiment Strain et al.'s (1995) Experiment 2, which produced a robust interaction between regularity and imageability in a set of low-frequency words.

## EXPERIMENT 2

### Imageability effects for low-frequency English words

The main concern in Experiment 2 was to enhance the sensitivity of the design by enlarging the set of stimuli and the number of participants and by focusing just on low-frequency words. This was done by using the set of stimuli used by Strain et al. (1995) in their Experiment 2. In their study, this set produced a robust interaction between imageability and regularity in both latencies and error data.

#### Method

##### *Participants*

A total of 42 native English speakers (11 males and 31 females) aged between 20 and 70 years took part in this experiment. All had normal or corrected-to-normal vision. Each participant was paid a small fee for participating.

##### *Stimuli*

The low-frequency words of Strain et al.'s (1995) Experiment 2 (see their Appendix B) were used in this experiment. There were two sets of 32 low-frequency words, one for regular words and the other for exception words. Each set contained 16 high-imageability and 16 low-imageability words. Low-imageability words were matched with high-imageability words within each condition. A total of 40 items were monosyllabic, and the remaining 24 were disyllabic.

Each of two experimental blocks consisted of three starter items and 32 experimental items. This was preceded by a practice block of 20 words to allow participants to familiarize themselves with the task. A brief rest period was allowed between each block.

Similar to Experiment 1, all experimental words were rated for AoA by a group of 15 independent participants following the procedure of Morrison and colleagues (e.g., Morrison & Ellis, 1995). The mean AoA values were used as a covariate in the item analysis of reaction times (see Appendix 1).

##### *Procedure*

The procedure including the apparatus was identical to that in Experiment 1. After the computerized experiment, participants were asked to rate all experimental word stimuli for imageability on a 7-point scale. These ratings closely matched those of Strain et al. (1995) in terms of items assignment to the high- vs. low-imageability groups for each condition.



TABLE 3  
Imageability effects in mean reaction times<sup>a</sup> and error rates<sup>b</sup> for each condition in  
Experiment 2

Condition	High word			Low word		
	High imageability	Low imageability	Effect	High imageability	Low imageability	Effect
RT	605	593	-12	612	675	63*
SD	66.4	69.3		64.9	84.4	
Error	0.5	0.1	-0.4	0.2	19.2	19.0*

<sup>a</sup>In ms.

<sup>b</sup>In percentages.

## Results

*Response times.* Table 3 shows correct mean RTs (in milliseconds) and error rates for each condition. A two-way ANOVA (Word Type  $\times$  Imageability) performed on the RT data showed that participants were significantly faster in naming regular words (599 ms) than exception words (643 ms),  $F_1(1, 41) = 43.93$ ,  $MSE = 1871.5$ ,  $p < .0001$ ;  $F_2(1, 60) = 12.34$ ,  $MSE = 2854.6$ ,  $p < .001$ . Participants also named high-imageability words (608 ms) significantly faster than low-imageability words (634 ms),  $F_1(1, 41) = 27.00$ ,  $MSE = 1035.0$ ,  $p < .0001$ ,  $F_2(1, 60) = 7.08$ ,  $MSE = 2854.6$ ,  $p < .05$ . The interaction between word type and imageability reached significance,  $F_1(1, 41) = 48.40$ ,  $MSE = 1206.1$ ,  $p < .0001$ ;  $F_1(1, 60) = 9.78$ ,  $MSE = 2854.6$ ,  $p < .005$ . Analyses of the simple main effect of this interaction showed that this interaction was due to a significant difference between high- and low-imageability exception words,  $F_1(1, 41) = 71.29$ ,  $MSE = 1171.81$ ,  $p < .0001$ ;  $F_2(1, 30) = 13.13$ ,  $MSE = 3640.8$ ,  $p < .005$ , but not for regular words,  $F_1(1, 41) = 2.59$ ,  $MSE = 1069.2$ ,  $p = .12$ ;  $F_2 < 1$ , indicating that the imageability effect appears only for low-frequency exception words.

Finally, the item analysis was performed on the RT data using AoA as a covariate. Although the main effect of imageability disappeared,  $F_2 < 1$ , the regularity by imageability interaction was still significant,  $F_1(1, 59) = 14.97$ ,  $MSE = 2179.1$ ,  $p < .001$ . There was a significant effect of AoA,  $t(59) = 4.43$ ,  $p < .001$ . AoA ratings also had a high negative correlation with the imageability ratings,  $r = -.76$ .

*Errors.* Voice-key errors were removed from the analysis. Naming errors were categorized as one of three types: (1) Visual-phonological word errors; (2) visual-phonological nonword errors, and (3) regularization errors. Regularizations occurred most frequently (90% of the total), as compared with visual-phonological word errors (5%) and visual-phonological nonword errors (5% see Table 4). Out of 127 regularization errors, 58 were caused by the words *cache* (produced by 34 participants out of 42) and *wrath* (produced by 24 participants out of 42).

A two-way ANOVA on error scores (arcsine transformed after a constant of 0.01 was added to each error score) showed significant main effects of word type,  $F_1(1, 41) = 137.74$ ,  $MSE = 0.0033$ ,  $p < .0001$ ;  $F_2(1, 60) = 9.26$ ,  $MSE = 0.0209$ ,  $p < .005$ , and imageability,  $F_1(1, 41) = 169.17$ ,  $MSE = 0.0018$ ,  $p < .0001$ ;  $F_2(1, 60) = 6.43$ ,  $MSE = 0.0209$ ,  $p < .05$ . The former effect

TABLE 4  
Error rates<sup>a</sup> in response to the various word types in Experiment 2  
and Strain et al.<sup>b</sup>

Experiment		High frequency		Low frequency	
		High imageability	Low imageability	High imageability	Low imageability
2	Regularization	0	0	1.68	17.82
	Other	0.47	0.15	0.31	1.23
Strain et al. <sup>b</sup>	Regularization	0	0	2.00	14.10
	Other	1.25	1.10	1.60	4.50

<sup>a</sup>In percentages.

<sup>b</sup>Strain et al. (1995, Experiment 1).

shows that error rates were higher for exception words (10.6%) than for regular words (0.3%). The latter effect indicates that naming of low-imageability words (9.7%) produced more errors than that of high-imageability words (1.2%).

There was a significant interaction of word type and imageability,  $F_1(1, 41) = 185.90$ ,  $MSE = 0.0018$ ,  $p < .0001$ ,  $F_2(1, 60) = 6.96$ ,  $MSE = 0.0209$ ,  $p < .05$ . Further analyses showed that this interaction was due to a significant difference between high- and low-imageability exception words,  $F_1(1, 41) = 188.18$ ,  $MSE = 0.0034$ ,  $p < .0001$ ,  $F_1(1, 30) = 6.73$ ,  $MSE = 0.0415$ ,  $p < .05$ , which indicates that imageability effects are produced only for low-frequency exception words.

## Discussion

The results of both latency and error analysis successfully replicated the findings of Strain et al. (1995). An important finding was that the imageability by regularity interaction survived the analysis of covariance. The main effect of imageability disappeared, but the crucial regularity by imageability interaction was still significant when AoA was partialled out as a covariate. The AoA effect shows that words are named faster when they are learned early in life, than when they are learned late. However, even if the imageability effect is largely driven by the high negative correlation between AoA and imageability ( $r = -.76$  in the item set of Experiment 2), there is a significant residual effect that is unaccounted for. This can be taken as strong evidence that the effect of imageability is truly limited to low-frequency exception words. Considering that words learned later in life tend to be lower in frequency, it is suggested that the imageability effect may be more pronounced for late acquired words.

## EXPERIMENT 3

### Imageability effects in naming two-character Kanji words

Kanji orthography is logographic in nature, and each Kanji character has no separate components that correspond to individual phonemes. Therefore, a Kanji character cannot be decomposed phonetically in the same way as the alphabetic scripts. Moreover, most Kanji characters

have more than one pronunciation: a KUN-reading of Japanese origin, and one or more ON-readings of Chinese origin. For KUN reading, a single character can be used as a word (with a concrete meaning in most cases) and it can also be combined with other characters to make multi-character words; whereas for ON reading there are hardly any single-character words. For most Kanji characters with more than one reading, the appropriate pronunciation is determined by the intra-word context—that is, the other character(s) with which the particular character combines to constitute the word in question (see Wydell, Butterworth, & Patterson, 1995).

Originally, the meaning of Kanji characters was thought to be accessed directly from its orthographic representation without need to access phonological information (Feldman & Turvey, 1980; Goryo, 1987; Kimura, 1984; Saito, 1981). That is, Kanji characters were thought to be read primarily, or even exclusively via a lexical-semantic route. However, more recent studies have provided evidence that the computation of phonology is automatic (Mizuno, 1997) and that semantics can be accessed in parallel from both orthography and phonology (Wydell, Patterson, & Humphreys, 1993).

In general, ON reading is used to pronounce most of the two-character Kanji words. The number of two-character words that are pronounced using the KUN reading is rather small. In terms of distributional statistics, ON reading can be therefore regarded as the *regular* (i.e., most frequent) pronunciation in the context of two-character Kanji words. However, the existence of KUN reading pronunciations for one or both characters of a two-character word introduces the issue of character-to-sound consistency (Wydell et al., 1995). Therefore, Kanji words can be consistent (each character has a single ON reading with no KUN reading) or inconsistent (at least one of the characters can be pronounced by both ON and KUN readings, depending on the word in which it occurs). A more crucial distinction, however, is whether the entire word is pronounced as ON or KUN. Given our characterization of ON reading as “regular” (i.e., statistically more frequent) pronunciation, we will consider KUN reading words as approximately analogous to English exception words. Recently, Wydell, Butterworth, Shibahara, and Zorzi (1997) have shown that two-character, ON-reading Kanji words are named faster and more accurately than KUN-reading Kanji words. However, Wydell et al. found no RT difference in Kanji naming between consistent ON reading words and inconsistent ON reading words. In other words, naming latencies were not affected by the existence of alternative readings of the characters per se, but they were affected by whether the pronunciation of the word was ON or KUN.

As a consequence of the different distribution of ON and KUN readings in two-character Kanji words, we can assume that the computation of phonology from print is less efficient for KUN reading words (also see Fushimi et al., 1999), in particular if they are of low frequency. Clearly, the activation of semantic representations would be most beneficial for these words (note that they also yield the slowest naming latencies; Wydell et al., 1997). We therefore predicted that, if the same basic reading processes are applicable to both alphabetic English and logographic Japanese Kanji, imageability effects should appear for low-frequency KUN words.

## Method

### *Participants*

A total of 16 native Japanese speakers (7 males and 9 females) aged between 19 and 30 years took part in this experiment. They were brought up in Japan at least to the age of 18 and had lived in the United

Kingdom for less than 18 months. All had normal or corrected-to-normal vision. Each participant was paid a small fee for participating.

### *Apparatus*

All the experimental equipment was identical to that used in Experiment 1 except that Japanese Kanji stimulus words were displayed on the computer screen by the Macintosh software SweetJAM 4.5 (1990) installed on the English operating system.

### *Stimuli*

The stimuli consisted of three sets of 32 two-character Kanji words with Consistent-ON (Con-ON), Inconsistent-ON (Inc-ON), and Inconsistent-KUN (Inc-KUN) readings chosen from a corpus of 2,357 Japanese nouns (Wydell, Quinlan, & Butterworth, in press).

- Con-ON: Each component character of a two-character Kanji word of this type has one possible ON pronunciation without alternative ON and KUN pronunciations.
- Inc-ON: At least one of the two characters has an alternative KUN reading. However, the correct pronunciation of this type of Kanji is the ON reading, although KUN readings can be used in other compound words.
- Inc-KUN: These are KUN reading compounds with each component character having an ON reading used in other words. The pronunciation of the whole word is KUN, although the typical pronunciation of each character across its word neighbourhood is an ON reading rather than the KUN reading.

Within each word type, stimuli were matched for frequency<sup>1</sup> and imageability ratings taken from Wydell et al. (in press). Each set contained equal numbers of high-frequency and high-imageability words (Hf-Hi), high-frequency and low-imageability words (Hf-Li), low-frequency and high-imageability words (Lf-Hi), and low-frequency and low-imageability words (Lf-Li). For each condition, half the stimuli were Kanji words with three morae, and the other half were words with four morae.

Kanji words were classified as high frequency when familiarity exceeded 4.0 (on a 7-point scale), otherwise they were categorized as low frequency. Words were assigned to the high-imageability group when the imageability value was equal to 5.5 or higher (on a 7-point scale), otherwise they were assigned to the low-imageability group. There were differences in the criteria for the assignment of Kanji words to high and low groups between the two dimensions. This is due to a trend toward higher scores on imageability ratings than on frequency ratings for Kanji words (e.g., the mean score of frequency was 4.09 and that of imageability was 5.41 in the present stimulus set). Table 5 shows the mean familiarity and imageability ratings for each type of word. Words in each set were matched by initial phoneme or sound similar to the initial phoneme where it was not possible to match for initial phoneme.

Similar to Experiments 1 and 2, 12 independent Japanese participants rated all experimental Kanji words for AoA. For each word, they were asked to provide separate estimates of AoA for the spoken form (age of spoken acquisition) and for the written form (age of written acquisition). Mean values of AoA (for both spoken and written AoA) were used as a covariate in the item analysis of reaction times (see Appendix 2).

<sup>1</sup>We used rated word familiarity (from Wydell et al., in press) to classify target words into high- and low-frequency groups because no objective frequencies were available for two-character Kanji word when we designed the current experiment (see Wydell et al., 1995). However, word frequency counts in Japanese Kanji are now available (Amano & Kondo, 2000).

TABLE 5  
Mean scores of familiarity, imageability, and AoA ratings for each condition in Experiment 3

Condition		High frequency		Low frequency	
		High imageability	Low imageability	High imageability	Low imageability
Con-ON	Familiarity	4.8	4.4	3.5	3.5
	Imageability	5.7	4.8	5.9	4.5
	Spoken AoA	3.7	5.1	4.2	5.5
	Written AoA	5.1	5.8	5.8	5.8
Inc-ON	Familiarity	5.2	5.0	3.4	3.2
	Imageability	6.1	4.8	6.1	4.8
	Spoken AoA	4.0	3.9	4.7	5.3
	Written AoA	4.9	4.8	5.2	5.2
Inc-KUN	Familiarity	5.2	4.6	3.5	2.9
	Imageability	6.4	4.7	6.4	4.7
	Spoken AoA	3.2	4.2	3.6	6.0
	Written AoA	5.0	4.9	5.4	6.2

### Procedure

The procedure was identical to that in Experiment 1 except that participants named the complete set of 96 Japanese Kanji words in a series of six blocks. Each experimental block began with two starter items, resulting in 18 items in each block. Before the experimental blocks participants were given a practice block of 18 items. After all experimental blocks, participants were asked to rate all experimental words for imageability on a 7-point scale. These ratings corresponded closely with those of Wydell et al. (in press).

### Results

Table 6 shows correct mean RTs (in milliseconds) and error rates, together with the imageability effects for each condition.

*Response times.* Correct RTs were submitted to a three-way ANOVA with word type (Con-ON vs. Inc-ON vs. Inc-KUN), frequency (high vs. low), and imageability (high vs. low) as factors.

The main effect of word type was significant,  $F_1(2, 30) = 17.18$ ,  $MSE = 12,832.9$ ,  $p < .0001$ ;  $F_2(2, 84) = 13.86$ ,  $MSE = 11,942.7$ ,  $p < .0001$ . This indicates that participants were slower in naming Inc-KUN words (838 ms) than Con-ON (743 ms) and Inc-ON (731 ms) words ( $p < .05$ , Scheffé test) with no difference between the Con-ON and Inc-ON condition. There were also significant main effects of frequency,  $F_1(1, 15) = 49.98$ ,  $MSE = 14,343.3$ ,  $p < .0001$ ,  $F_2(1, 84) = 41.99$ ,  $MSE = 11,942.7$ ,  $p < .0001$ , and imageability,  $F_1(1, 15) = 14.70$ ,  $MSE = 5954.9$ ,  $p < .005$ ;  $F_2(1, 84) = 5.10$ ,  $MSE = 11,942.7$ ,  $p < .05$ . These results show that high-frequency words (710 ms) produced shorter RTs than low-frequency words (832 ms), and that RTs for high-imageability words (750 ms) were faster than those for low-imageability words (792 ms).

TABLE 6  
Imageability effects in mean reaction times<sup>a</sup> and error rates<sup>b</sup> for each condition in Experiment 3

Condition		High frequency			Low frequency		
		High imageability	Low imageability	Effect	High imageability	Low imageability	Effect
Con-ON	RT	715	740	25	726	793	67
	SD	175.7	163.5		137.1	205.5	
	Error	0.8	1.6	0.8	0.8	10.2	9.4
Inc-ON	RT	661	664	3	846	753	-93
	SD	95.7	100.8		150.8	137.3	
	Error	1.6	0	-1.6	3.9	10.2	6.3
Inc-KUN	RT	744	736	-8	806	1068	262*
	SD	118.0	155.4		132.4	273.2	
	Error	0.8	2.3	1.5	6.3	24.2	17.9*

<sup>a</sup>In ms.

<sup>b</sup>In percentages.

\*= significant at  $p < .05$  by both subjects and items.

All two-way interactions were significant: word type by frequency,  $F_1(2, 30) = 17.10$ ,  $MSE = 6432.2$ ,  $p < .0001$ ,  $F_2(2, 84) = 4.21$ ,  $MSE = 11,942.7$ ,  $p < .05$ , word type by imageability,  $F_1(2, 30) = 19.54$ ,  $MSE = 6033.6$ ,  $p < .0001$ ;  $F_2(2, 84) = 3.11$ ,  $MSE = 11,942.7$ ,  $p < .05$ , and frequency by imageability,  $F_1(1, 15) = 10.05$ ,  $MSE = 6190.7$ ,  $p < .01$ ,  $F_2(1, 84) = 4.39$ ,  $MSE = 11,942.7$ ,  $p < .05$ . However, these were qualified by the critical three-way interaction between word type, frequency, and imageability,  $F_1(2, 30) = 30.54$ ,  $MSE = 4485.5$ ,  $p < .0001$ ;  $F_2(2, 84) = 6.45$ ,  $MSE = 11,942.7$ ,  $p < .005$ , showing that the imageability effect was reliable (both by subjects and by items) only for low-frequency Inc-KUN words,  $F_1(1, 15) = 33.54$ ,  $MSE = 16,354.9$ ,  $p < .0001$ ;  $F_2(1, 14) = 9.67$ ,  $MSE = 29,678.7$ ,  $p < .01$ .

Analysis of covariance of the item data was performed using AoA of spoken or written type as a covariate. The main effect of imageability disappeared for spoken AoA,  $F_1(1, 83) = 2.59$ ,  $MSE = 12,063.7$ ,  $p = .11$ , whereas there was still a marginal effect using written AoA,  $F_2(1, 83) = 3.96$ ,  $MSE = 11,887.5$ ,  $p < .05$ . The three-way interaction was still significant in both analyses with spoken AoA,  $F_2(2, 83) = 6.20$ ,  $MSE = 12,063.7$ ,  $p < .005$ , and written AoA,  $F_2(2, 83) = 5.72$ ,  $MSE = 11,887.5$ ,  $p < .01$ , indicating imageability effects for low-frequency Inc-KUN words. The effect of AoA was not significant for either spoken,  $t(83) = 0.40$ ,  $p = .69$ , or written,  $t(83) = 1.18$ ,  $p = .24$ , types. A significant correlation between imageability and AoA ratings was observed for ratings of spoken words,  $r = -.51$ , but not for ratings of written words,  $r = -.20$ .

**Errors.** Voice-key errors were removed from the analysis. As shown in Table 7, naming errors were categorized according to the following types: (1) Semantic errors (a different word semantically related to the target was produced); (2) visual errors (a different word visually (or graphically) similar to the target was produced); (3) reverse errors (the participant pronounced the second character correctly first, then named the first character); (4) start-stop errors (the participant pronounced the first character correctly, then stopped and named the

TABLE 7  
Error rates<sup>a</sup> of different types of errors for each condition in Experiment 3

Condition		High frequency		Low frequency	
		High imageability	Low imageability	High imageability	Low imageability
Con-ON	Semantic	0.78	0	0	0
	Visual	0	0.78	0	4.69
	Phonological	0	0.78	0.78	1.56
	Start-stop	0	0	0	3.91
	Reverse	0	0	0	0.78
	LARC	0	0	0	0
	Other	0	0	0	0.78
Inc-ON	Semantic	0	0	0	0
	Visual	0.78	0	2.34	0
	Phonological	0	0	0	0
	Start-stop	0.78	0	0	1.56
	Reverse	0	0	0	0.78
	LARC	0	0	0.78	5.47
	Other	0	0	1.56	0.78
Inc-KUN	Semantic	0	0	0	1.56
	Visual	0	0	0	2.34
	Phonological	0	0	0	3.13
	Start-stop	0.78	0.78	0.78	2.34
	Reverse	0	0	0.78	0
	LARC	0	0.78	4.69	3.13
	Other	0	0.78	0	10.16

<sup>a</sup>In percentages.

whole word correctly); (5) phonological errors (a different word phonologically similar to the target was produced); (6) LARC (Legitimate Alternative Reading of the Components; Fushimi et al., 1999; Patterson, Suzuki, Wydell, & Sasanuma, 1995; Wydell et al., 1995) errors (where the whole word pronunciation is wrong but the pronunciation of each character is a legitimate one in the sense that it is appropriate to other words containing these characters); and (7) other (errors that can not be categorized into the previous groups).

There were differences in the number of LARC errors between low- and high-imageability words for Inc-ON and Inc-KUN types with low frequency. (There were no LARC errors for Con-ON words because no alternative pronunciations exist for this type of word.) For Inc-ON, errors produced in reading low-imageability words were mainly LARC errors, as compared with a small number of LARC errors for high-imageability words. In contrast, for Inc-KUN, LARC errors were the main type of error produced for high-imageability words, but not for low-imageability words.

A three-way ANOVA on error scores (arcsine transformed after a constant of 0.01 was added to each error score) yielded significant main effects of frequency,  $F_1(1, 15) = 41.49$ ,  $MSE = 0.0079$ ,  $p < .0001$ ;  $F_2(1, 84) = 18.53$ ,  $MSE = 0.0083$ ,  $p < .0005$ , and imageability,  $F_1(1, 15) = 37.27$ ,  $MSE = 0.0045$ ,  $p < .0001$ ;  $F_2(1, 84) = 8.27$ ,  $MSE = 0.0083$ ,  $p < .01$ . These effects indicate that participants made more errors on low-frequency (9.2%) than on high-frequency

(1.2%) words, and that error rates for low-imageability words (8.1%) were higher than those for high-imageability words (2.3%). The effect of word type was significant only by subjects,  $F_1(2, 30) = 5.27$ ,  $MSE = .0101$ ,  $p < .05$ , although with a trend towards significance in the item analysis,  $F_1(2, 84) = 2.39$ ,  $MSE = 0.0083$ ,  $p = .10$ , showing that to some extent Inc-KUN words (8.4%) produced more errors than Con-ON (3.3%) and Inc-ON (3.9%) words.

There was a significant interaction between frequency and imageability,  $F_1(1, 15) = 37.06$ ,  $MSE = 0.0041$ ,  $p < .0001$ ;  $F_2(1, 84) = 8.26$ ,  $MSE = 0.0083$ ,  $p < .01$ , showing that the imageability effect was present only for low-frequency words (11.5% vs. 0.3%). The size of the imageability effect for low-frequency words was 9.4% for Con-ON, 6.3% for Inc-ON, and 17.9% for Inc-KUN. However, further analyses showed that the effect was reliable both by subjects and by items only for Inc-KUN words,  $F_1(1, 15) = 18.59$ ,  $MSE = 0.0151$ ,  $p < .001$ ;  $F_2(1, 14) = 5.42$ ,  $MSE = 0.0207$ ,  $p < .05$ .

## Discussion

The results show that imageability has a reliable effect on both naming latencies and naming accuracy only for low-frequency Inc-KUN words. The imageability effect was confirmed in the analysis of covariance with AoA as a covariate (naming RTs): the crucial three-way interaction between frequency, imageability and word type was still significant when the effect of either spoken or written AoA was partialled out in the analysis. Thus, these results parallel those of Experiment 2 (English). Imageability affects naming of low-frequency Inc-KUN words—that is, the type of words that we characterized as the (rough) equivalent to the English exception (or irregular) words.

Our finding of an imageability effect in naming two-character KUN words runs counter to the results of Yamazaki, Ellis, Morrison, and Lambon Ralph (1997), who showed no correlation between imageability and naming RTs of one-character KUN words. However, single characters that have KUN pronunciation are generally associated with concrete meanings, and their number is very small. It would be therefore harder to detect a sizeable imageability effect within the restricted sample of words (which are relatively high in concreteness) used by Yamazaki et al.

Moreover, when comparing Yamazaki et al.'s (1997) results and ours, there seem to be differences in the effects of AoA on naming RTs between single- and two-character Kanji words. That is, earlier acquired words are named faster than later acquired words in Yamazaki et al.'s set of single-character Kanji, whereas AoA has no significant effect on the naming speed of our set of two-character Kanji words. However, the mean spoken AoA is much lower for Yamazaki et al.'s single-character Kanji (2.68) compared to our set of two-character Kanji (4.60 for Con-ON, 4.45 for Inc-ON, and 4.26 for Inc-KUN)<sup>2</sup>. This suggests that AoA has little influence on the naming RTs of the two-character words because they are all learned relatively later in life. Finally, spoken AoA was correlated with imageability both in Yamazaki et al.'s study and in ours. This is because high-imageability words are likely to be learned earlier in life than low-

<sup>2</sup>Our ratings of age of written acquisition differ from those of Yamazaki et al. (1997) in that we asked participants to rate AoA on a 7-point scale, whereas Yamazaki et al. took them from the Gakunen-haitouhyou list published by the Japanese Ministry of Education.



imageability words when they are used in spoken form. In contrast the age at which the Kanji itself is learned as a written script is not necessarily correlated with the degree of imageability.

Of greater interest is the large number of LARC errors for low-frequency Inc-ON words when they were low in imageability. In general, characters of Inc-ON words are associated with concrete meanings when they are read with KUN pronunciations. The relatively high imageability of the character level, combined with the low imageability of the whole word, might therefore produce a bias towards KUN pronunciation, which would lead to the production of LARC errors.

Finally, there was no significant difference in naming latencies between Con-ON and Inc-ON words (i.e., no consistency effect). Although the component characters of Inc-ON words have alternative KUN pronunciations, ON readings are statistically typical for the Inc-ON words used in the present experiment. That is, when the first component character is combined with other characters, there are on average 43 combinations for ON reading and 3 combinations for KUN reading. In the case of the second component character, there are on average 32 combinations for ON reading and 2 combinations for KUN reading. Therefore, the lack of consistency effects may be due to the statistical typicality of the ON pronunciation for the component characters of Inc-ON words. Moreover, there was a non-significant trend towards Inc-ON words being named faster than Con-ON words. This may be due to the fact that the frequencies of the component characters in the Inc-ON word set (3.03 and 2.99 for each character) are significantly higher than those in the Con-ON word set (2.65 and 2.63 for each character),  $t(31) = 2.98, p < .005$  for the first character and  $t(31) = 2.25, p < .05$  for the second character, although there was no difference in the word frequencies between them,  $t(31) = 0.52, p = .61$ .

*Comparison between English and Japanese.* The imageability effect was reliable for low-frequency exception words in English (Experiment 2) and for low-frequency KUN words in Japanese Kanji (Experiment 3). However, a comparison of English and Kanji shows that the imageability effect is stronger in Kanji (see Figure 1). This was confirmed in a combined

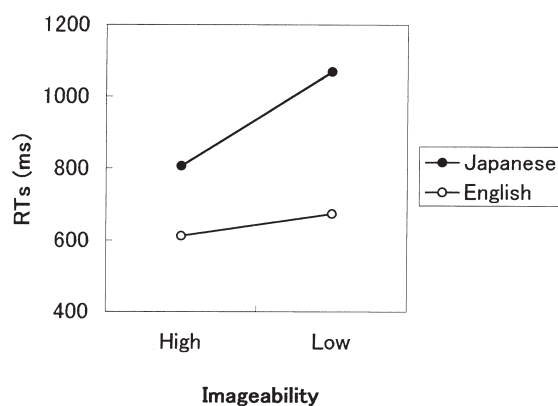


Figure 1. Mean naming latencies (ms) as a function of imageability for English low-frequency exception words (Experiment 2) and for Kanji low-frequency KUN words (Experiment 3).

analysis of Experiments 2 and 3 on the low-frequency exception/KUN words, with language (English vs. Japanese) as a between-subjects factor and imageability (high vs. low) as a within-subject factor. Both main effects were significant, indicating that English readers were faster than Japanese readers,  $F(1, 56) = 71.7, MSE = 27,756.57, p < .001$ , and that high-imageability words were named faster than low-imageability words,  $F(1, 56) = 76.97, MSE = 5238.7, p < .001$ . The claim that the imageability effect is larger in reading Kanji was substantiated by the highly significant interaction between language and imageability,  $F(1, 56) = 43.7, MSE = 5238.7, p < .001$ . It is important to note that the bigger size of the effect in Kanji is not the result of a larger difference in imageability ratings between low- and high-imageability words. The mean imageability scores for the Kanji low-frequency KUN words were 6.4 for high- and 4.7 for low-imageability words, whereas the mean scores for the English low-frequency exception words were 5.9 for high- and 3.2 for low-imageability words. Therefore, the difference in the mean scores would act in the opposite direction—that is, the imageability effect should be larger in English.

## GENERAL DISCUSSION

Participants are slower and more error prone in reading aloud low-frequency exception words, in particular when the words are low in imageability. The results partially confirmed the previous finding of Strain et al. (1995) that imageability affects naming of low-frequency exception words, and they extended the results to a very different orthographic system, Japanese Kanji. The imageability effect in Kanji was found for two-character KUN-reading words. Crucially, for both English and Kanji, we ruled out the hypothesis that the interaction between imageability and word type is produced by a confounding variable, AoA. We found that the overall imageability effect is driven by the high negative correlation with AoA; however, an imageability effect restricted to low-frequency exception words is still found when AoA is partialled out as a covariate.<sup>3</sup> This finding supports the hypothesis that the imageability effect and the AoA effect have independent sources. For instance, some authors have argued that the imageability effect would reflect the quality of semantic representations (e.g., Jones, 1985; Plaut & Shallice, 1993; Saffran et al., 1976), whereas the AoA effect would reflect the quality of phonological representations (e.g., Brown & Watson, 1987; Ellis & Morrison, 1998;

<sup>3</sup>Contrary to our results, Monaghan and Ellis (2002) recently showed that the interaction between regularity and imageability disappeared when AoA was entered as a covariate (this paper appeared after our paper was originally submitted). They further showed that when RT data taken from Strain et al. (1995) were reanalysed with AoA entered as a covariate, there was a significant regularity by imageability interaction although the main effect of imageability did not reach significance. Monaghan and Ellis suggest that such an interaction requires a significant main effect of imageability in order to claim that imageability affects the naming of low-frequency exception words. However, the PDP framework proposes that semantic representations will be automatically activated for all words. In this framework, the computation of orthography-to-phonology is too efficient and self-sufficient for regular words and high-frequency exception words. Therefore, naming these words does not depend on the semantic contribution. On the other hand, low-frequency exception words have the smallest impact on setting weights for orthography-to-phonology translation, and therefore the translation of orthography-to-phonology is somewhat inefficient, slow, or error prone. Thus, these low-frequency exception words benefit more from activation of semantic representations when they are high in imageability. In short, what the PDP model predicts is the regularity by imageability interaction, but not the main effect of imageability (see Strain et al., 1995; Strain, Patterson, & Seidenberg, 2002).

Morrison & Ellis, 1995) or a gradual loss of plasticity during the acquisition of new lexical items (Ellis & Lambon Ralph, 2000).

Imageability refers to the ease with which a word produces a mental image irrespective of whether it is concrete or not. In this respect, imageability and concreteness are dissociable although they are highly correlated with each other. There are two distinct accounts of how imageability is represented in the semantic system. One model argues that imageability is represented in the number of sensory features of a concept (Bird, Howard, & Franklin, 2000; Bird, Lambon Ralph, Patterson, & Hodges, 2000), whereas the other model claims that imageability reflects the total number of semantic features (semantic richness) available in a semantic representation (Plaut & Shallice, 1993). Although these models are distinguished by the manner in which imageability is represented in the semantic system, it is assumed by both that imageability effects are produced due to a differential weighting of the sensory features or the semantic features represented by words.

There are two possible explanations for why imageability effects are confined to low-frequency exception words. According to the PDP model of reading (Plaut et al., 1996; Seidenberg & McClelland, 1989), the computation of phonology from print results from the combination of two sources of activation, which are generated by the orthography-to-phonology (O-P) pathway and by the orthography-to-semantics-to-phonology (O-S-P) pathway. In principle, the O-P network could learn the correct pronunciation of both regularly and irregularly spelt words (Plaut et al., 1996); however, the computation of phonology is much more difficult for exception words, in particular when they are low in frequency, because their pronunciation is inconsistent with that of other words having similar spelling patterns. Therefore, low-frequency exception words greatly benefit from the contribution of semantic activation (i.e., the O-S-P pathway). When Plaut et al. trained the O-P network with an additional external input to the phonological units (representing the contribution of a putative semantic pathway), Plaut et al. observed a redistribution of labour between the (putative) semantic pathway and the O-P network. As the competence of the semantic pathway improved, they observed a shift in the reliance for correct reading of low-frequency exception words from the phonological to the semantic pathway. Thus, the effect of imageability in reading low-frequency exception words appears to fall out naturally from the PDP framework.

It must be noted, however, that the same results can be easily accommodated with two-process models of reading, such as the Dual Route Cascade (DRC) model (Coltheart et al., 1993; Coltheart & Rastle, 1994; Rastle & Coltheart, 1999) and the Connectionist Dual-Process model (Zorzi, 2000; Zorzi, Houghton, & Butterworth, 1998; Zorzi et al., 1998b). For instance, in a classic dual-route model such as DRC, semantic effects would only be observed for the slowest stimuli, which are just the low-frequency exception words. Exception words are read primarily via the lexical route, but in the case of low-frequency words processing would be sufficiently slow to allow semantic effects to emerge from processing in the lexical-semantic route.

One possible way to disentangle this issue might be to investigate whether the imageability effect is modulated by reading skill. In a recent version of the PDP model developed by Harm and Seidenberg (1999), both semantic and phonological pathways have been fully implemented, and the model learns the mappings from orthography to phonology and from orthography to semantics at the same time. Harm and Seidenberg examined the division of labour between the two pathways over time and found that the reliance on the semantic pathway

increases with skill level. Crucially, they also measured the size of the imageability effect for early and late points in training and found an increase of the effect over time. Notably, Strain and Herdman (1999) found that poor readers show a stronger imageability effect than good readers. This, however, would reflect an increased reliance on semantic processing as a consequence of the poor phonological processing skills (i.e., an “abnormal” redistribution of labour) and not the normal learning process that was tracked over time in the PDP model by Harm and Seidenberg (1999).<sup>4</sup>

*Cross-linguistic considerations.* The finding that semantic effects are similar in English and Kanji speaks to the issue of language-specific versus universal principles underlying the reading system. Although the converging results for the two languages suggest similar architectures and computational properties, the effect of imageability was more pronounced in reading Kanji than in reading English. One possible interpretation is that lexical-semantic processing is more weighted by the Japanese readers, as suggested by the orthographic depth hypothesis (e.g., Frost et al., 1987). A differential weighting of lexical phonology in reading was also predicted for different languages in the computational study of Zorzi et al. (1998b): “If direct (assembly) pathway and mediated (lexical) pathway were the basic starting architecture of all phonological mechanisms, the role of the mediated pathway would depend on the orthographic transparency of the language that the model is trained on. Hence, an opaque orthography such as Japanese Kanji . . . would mostly rely on the mediated pathway. . . . In other cases of shallow orthographies, the mediated pathway would be necessary at least for resolving minor inconsistencies, such as the stress assignment in Italian.” (pp. 1156–1157). Notably, recent neuroimaging evidence suggests exactly this kind of differential weighting in the comparison between English and Italian (Paulesu et al., 2000).

There is another explanation for the large imageability effects in Kanji reading.<sup>5</sup> According to the naming RT data, it took longer to articulate Kanji words than English words. This may happen due to differences in word length between English (1 or 2 syllables) and Kanji (3 or 4 morae) words as well as the relative transparency of spelling-sound mappings between them. If the effect of semantic information on reading takes some time to develop, a semantic variable such as imageability will have a stronger impact on slower responses (i.e., Kanji word naming) than on fast responses (i.e., English word naming). Therefore, the imageability effects on naming low-frequency words will be more pronounced in Kanji than in English. Further research should better clarify the issue of cross-linguistic differences in the cognitive architecture of reading.

## REFERENCES

- Amae, S., & Kondo, T. (2000). *Word frequency database. NTT database series: Lexical properties of Japanese* (Vol. 7). Tokyo: Sanseido.

<sup>4</sup>Note that the attempt to derive predictions from computational models without actually running the simulations can be very misleading: For instance, Zorzi (2000) demonstrated that a putative serial effect—that is, the position of irregularity effect (Rastle & Coltheart, 1999)—can be produced (contrary to predictions) by the parallel model of Zorzi et al. (1999b).

<sup>5</sup>We are grateful to Karalyn Patterson for suggesting this alternative explanation.

- Barry, C., Morrison, C.M., & Ellis, A.W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology*, 50A, 560–585.
- Bird, H., Howard, D., & Franklin, S. (2000). Why is a verb like an inanimate object? Grammatical category and semantic category deficits. *Brain and Language*, 72, 246–309.
- Bird, H., Lambon Ralph, M.A., Patterson, K., & Hodges, J. (2000). The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language*, 73, 17–49.
- Brown, G.D.A., & Watson, F.L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, 15, 208–216.
- Carello, C., Turvey, M.T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning: Advances in psychology* (Vol. 94, pp. 211–226). Amsterdam: North-Holland.
- Cipolotti, L., & Warrington, E.K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, 1, 104–110.
- Coltheart, M., Curtis, B., Atkins, R., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608.
- Coltheart, M., Patterson, K.E., & Marshall, J.C. (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1197–1211.
- 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, 204–256.
- Denes, E., Cipolotti, L., & Zorzi, M. (1999). Acquired dyslexias and dysgraphias. In G. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental neuropsychology* (pp. 289–317). Hove, UK: Psychology Press.
- Ellis, A.W., & Lambon Ralph, M.A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1103–1123.
- Ellis, A.W., & Morrison, C.M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 515–523.
- Feldman, L.B., & Turvey, M.T. (1980). Words written in Kana are named faster than the same words written in Kanji. *Language and Speech*, 23, 141–147.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123, 71–99.
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 104–115.
- Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology*, 49A, 417–446.
- Fushimi, T., Ijuin, M., Patterson, K., & Tatsumi, I. (1999). Consistency, frequency, and lexicality effects in naming Japanese Kanji. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 382–407.
- Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 267–283.
- Gerhand, S., & Barry, C. (1999). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, 27, 592–602.
- Gilhooly, K.J., & Logie, R.H. (1980). Age-of-acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behavior Research Methods and Instrumentation*, 12, 395–427.
- Goryo, K. (1987). *On reading* [in Japanese]. Tokyo: Tokyo Daigaku Shuppankai.
- Gum, T., & Bub, D. (1988). *PsychLab software*. Montreal, Canada: Montreal Neurological Institute.
- Harm, M.W., & Seidenberg, M.S. (1999). *Division of labor in the triangle model of visual word recognition*. Paper presented at the 1999 Meeting of the Psychonomics Society.
- Jacobs, A.M., Rey, A., Ziegler, J.C., & Grainger, J. (1998). MROM-P: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Jones, G.V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, 24, 1–19.
- Kimura, Y. (1984). Concurrent vocal interference: Its effects on Kana and Kanji. *Quarterly Journal of Experimental Psychology*, 36A, 117–131.
- Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer-Verlag.

- Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lambon Ralph, M., Ellis, A.W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase*, *1*, 363–369.
- Lukatela, G., Popadic, R., Ognjenovic, R., & Turvey, M.T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, *8*, 124–132.
- McCarthy, R.A., & Warrington, E.K. (1990). *Cognitive neuropsychology: A clinical introduction*. San Diego, CA: Academic Press.
- Mizurio, R. (1997). A test of a hypothesis of automatic phonological processing of Kanji words. *The Japanese Journal of Psychology*, *68*, 1, 1–8.
- Monaghan, J., & Ellis, A.W. (2002). What exactly interacts with spelling–sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 183–206.
- Morrison, C.M., Chappell, T.D., & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559.
- Morrison, C.M., & Ellis, A.W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 116–133.
- Morrison, C.M., Ellis, A.W., & Quinlan, P.T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*, 705–714.
- Patterson, K.E., & Hodges, J.R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *12*, 1025–1040.
- Patterson, K.E., Graham, H., & Hodges, J.R. (1994). Reading in Alzheimer's type dementia: A preserved ability? *Neuropsychologia*, *8*, 395–412.
- Patterson, K.E., Suzuki, T., Wydell, T., & Sasanuma, S. (1995). Progressive aphasia and surface alexia in Japanese. *Neurocase*, *1*, 155–165.
- Paulesu, E., McCrory, E., Fazio, F., Mononcello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorossu, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C.D., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*(3–5), 91–96.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K.E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domain. *Psychological Review*, *103*, 56–115.
- Plaut, D.C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377–500.
- Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 482–503.
- Saffran, E.M., Schwartz, M.F., & Marin, O.S.M. (1976). Semantic mechanisms in paralexia. *Brain and Language*, *3*, 255–265.
- Saito, H. (1981). Use of graphemic and phonemic encoding in reading Kanji and Kana. *The Japanese Journal of Psychology*, *52*, 266–273.
- Schwartz, M.F., Saffran, E.M., & Marin, O.S.M. (1980). Fractionating the reading process in dementia: Evidence for word-specific print-to-sound associations. In M. Coltheart, K. Patterson, & J.C. Marshall (Eds.), *Deep dyslexia* (pp. 259–269). London: Routledge & Kegan Paul.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.
- Strain, E., & Herdman, C.M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*, 347–359.
- Strain, E., Patterson, K., & Seidenberg, M. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1140–1154.
- Strain, E., Patterson, K., & Seidenberg, M.S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 207–214.
- SweetJAM 4.5 [Computer software]. (1990). Tokyo, Japan: A & A Company, Ltd.
- Van Orden, G.C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*, 181–198.
- Van Orden, G.C., Johnston, J.C., & Hale, B.L. (1988). Word identification proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 371–386.
- Van Orden, G.C., Pennington, B.F., & Stone, G.O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, *97*, 488–522.
- Venezky, R.L. (1970). *The structure of English orthography*. The Hague, The Netherlands: Mouton.

- Wydell, T., Butterworth, B., & Patterson, K. (1995). The inconsistency of consistency effects in reading: The case of Japanese Kanji. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1155–1168.
- Wydell, T., Butterworth, B., Shibahara, N., & Zorzi, M. (1997). The irregularity of regularity effects in reading: The case of Japanese Kanji. In *Proceedings of the 1997 Meeting of the British Experimental Psychology Society, Cardiff, UK*, (p. 60). Glasgow: Bell and Bain Ltd.
- Wydell, T., Patterson, K., & Humphreys, G. (1993). Phonologically mediated access to meaning for Kanji: Is a ROWS still a ROSE in Japanese Kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 491–514.
- Wydell, T., Quinlan, P., & Butterworth, B. (in press). *Japanese lexical database: 2,357 Japanese nouns rated on frequency, familiarity, and imageability, and other language statistics on Kanji characters/words*. Hove, UK: Psychology Press.
- Yamazaki, M., Ellis, A.W., Morrison, C.M., & Lambon Ralph, M.A. (1997). Two age of acquisition effects in the reading of Japanese Kanji. *British Journal of Psychology*, *88*, 407–421.
- Zorzi, M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 847–856.
- Zorzi, M. (in press). Computational models of reading. In G. Houghton (Ed.), *Connectionist models in psychology*. Hove, UK: Psychology Press.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998a). The development of spelling–sound relationships in a model of phonological reading. *Language and Cognitive Processes*, *13*, 337–371.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998b). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

Original manuscript received 18 December 2001

Accepted revision received 21 May 2002

## APPENDIX 1

The mean ratings of spoken AoA and imageability for each condition in Experiments 1 and 2

Experiment	Condition		High frequency		Low frequency	
			High imageability	Low imageability	High imageability	Low imageability
1	Regular	Spoken AoA	2.22	3.14	3.15	4.67
		Imageability	5.45	2.60	5.88	2.51
	Exception	Spoken AoA	2.21	2.68	3.28	4.54
		Imageability	6.19	2.42	5.99	2.60
2	Regular	Spoken AoA	–	–	3.34	5.45
		Imageability	–	–	6.01	2.93
	Exception	Spoken AoA	–	–	3.11	4.95
		Imageability	–	–	5.89	3.24

## APPENDIX 2

The mean ratings of spoken and written AoAs and imageability for each condition in Experiment 3

Condition		High frequency		Low frequency	
		High imageability	Low imageability	High imageability	Low imageability
Regular	Spoken AoA	3.68	5.04	4.20	5.45
	Written AoA	5.08	5.78	5.81	5.81
	Imageability	5.74	4.79	5.94	4.51
Inc-ON	Spoken AoA	3.98	3.89	4.66	5.28
	Written AoA	4.90	4.80	5.19	5.19
	Imageability	6.11	4.80	6.05	4.80
Inc-KUN	Spoken AoA	3.18	4.28	3.61	5.96
	Written AoA	4.97	4.94	5.40	6.18
	Imageability	6.39	4.73	6.36	4.68