Reading in a Foreign Language April 2016, Volume 28, No. 1
ISSN 1539-0578 pp. 43–62

Automatization and orthographic development in second language visual word recognition

Shusaku Kida
Hiroshima University
Japan

Abstract

The present study investigated second language (L2) learners' acquisition of automatic word recognition and the development of L2 orthographic representation in the mental lexicon. Participants in the study were Japanese university students enrolled in a compulsory course involving a weekly 30-minute sustained silent reading (SSR) activity with graded readers for 12 weeks. They completed the masked form-priming lexical decision task (LDT) before and after the in-class SSR activity. Results showed that participants exhibited signs of increasing automaticity of L2 word recognition (analyzed with the coefficient of variation), but could not develop their L2 orthographic representation (analyzed with the pattern of priming effects in the masked form-priming LDT). These findings suggest that automatization does not necessarily entail the development of orthographic representation, that is, the acquisition of automatic word recognition and the development of orthographic representation do not occur simultaneously. Instead, their development is asymmetrical.

Keywords: second language visual word recognition, automatization, orthographic representation, coefficient of variation, masked form-priming, sustained silent reading

Successful second language (L2) reading should require effective visual word recognition. The system of visual word recognition develops from the so-called alphabetic stage where words are recognized through letter-sound correspondence with unstable and less efficient processing and then through sight, where the processing is more rapid and flexible (Ehri, 1992, 1995, 2005). It is this later stage, often called the *orthographic stage*, which is typically regarded as the advanced level of visual word recognition (Castles & Nation, 2006; Perfetti, 1992; Share, 1995). Two characteristics often regarded as skilled orthographic processing are (a) automatic processing of word recognition (Ehri, 2005) and (b) fully developed orthographic representation (Perfetti, 1992).

Given that processing of a word is based on a representation of the word and that repeated processing of the same word results in the development of the representation, it is fair to assume that processing and representation have a highly interrelated relationship. However, to date, research on the two perspectives has been undertaken separately. Consequently, the relationship between the acquisition of automatic processing and the development of orthographic representation is not fully understood especially within the context of overall development of visual word recognition. The lack of research in this area has left a number of important questions unanswered. For example, how does the acquisition of automatic processing of a word relate to the development of its representation? Is the development of orthographic representation a prerequisite for the acquisition of automatic processing? Does the development of the two (automatic processing and representation) occur simultaneously or separately?

The present study aims to bridge the gap between the two perspectives of L2 word recognition development. Specifically, this study examines Japanese learners of English as a foreign language (EFL), their acquisition of automatic processing, and their development of orthographic representation in L2 through a required one-semester university-level reading course.

Automatic Processing and the Coefficient of Variation

One widely acknowledged phenomenon of word recognition is that experience and practice lead to automatic processing of words. Several characteristics have been proposed to describe the nature of automatic processing, which exemplify that it is fast, effortless, stable, or unintentional (DeKeyser, 2001). Although researchers have discussed these characteristics, they have faced difficulties in identifying precisely when learners achieve automatic processing (Segalowitz & Segalowitz, 1993). Nonetheless, there has been a general trend to make use of learners' latency data as a means to evaluate automatic processing. Latency data generally have some basic tendencies (Wagenmakers & Brown, 2007). First, distributions are not normal and are usually skewed to the right. Second, the skew increases with task difficulty. Third, the relationship between mean reaction time (RT) and spread of the distribution is linear; that is, "the spread of the distribution increases with the mean" (Wagenmakers & Brown, 2007, p. 830).

Utilizing the nature of the linearity of latency data and its distribution, Segalowitz and Segalowitz (1993) distinguished automatic processing from speed-up. According to Segalowitz and Segalowitz, practice can lead to "performance gains through qualitative changes in the functioning of the underlying processes through a restructuring effect" (p. 373) and they proposed using coefficient of variation (CV) as an index to examine operationally the automatization of information processing. CV is calculated as the standard deviation (*SD*) divided by mean RT and, thus, is expressed as CV_{RT}. Akamatsu (2009) showed how learners' word recognition developed with respect to the CV approach (Figure 1).

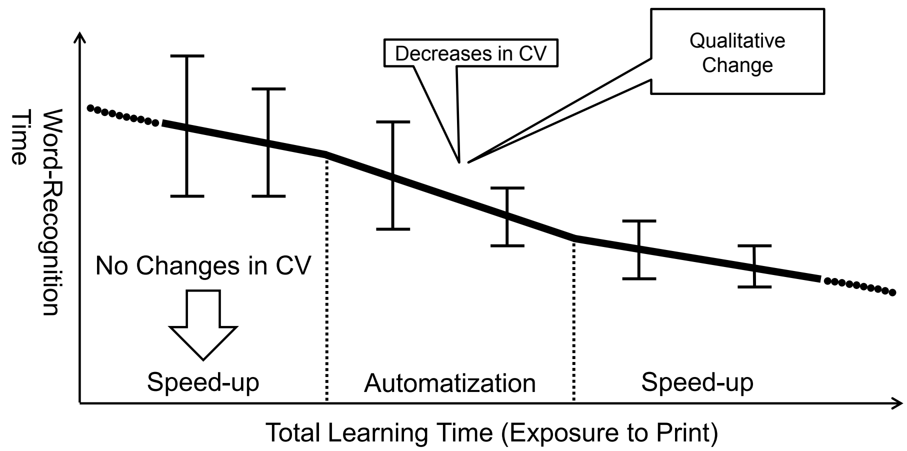

Figure 1. Time course of word recognition development from simple speed-up to automatization to the final speed-up phase (Akamatsu, 2009).

Suppose, for example, when the mean RT of L2 learners in a lexical decision task (LDT) is 1,000 ms and their *SD* is 100, the CV_{RT} is 0.10. Then, suppose that after they receive some training about word recognition, their mean RT is reduced from 1,000 to 800 ms and their *SD* is also reduced from 100 to 80. In this situation, the RT is drastically reduced and, therefore, some may argue that the learners have developed the automatic processing of the target words. However, from the perspective of CV_{RT}, this is not the case. This is because the reduction of the RT and the *SD* are proportionate and, therefore, CV_{RT} scores (0.10) remain the same from the pretest phase to the posttest phase (the left speed-up phase in Figure 1). Later, with more reading experience, learners show a disproportionate reduction in *SD*, in addition to a reduction in mean RT. At this stage, CV_{RT} values decrease resulting in automatization (the automatization phase in Figure 1). Finally, after the automatization period, another simple speed-up phase emerges. Therefore, from the CV perspective, in order to achieve automatization, learners must disproportionally reduce *SD*, in addition to the reduction of mean RT, resulting in a positive correlation between mean RT and CV_{RT}. Further, this implies that if some sort of training results in the development of automatic processing, the CV_{RT}–RT correlation should increase from before training to after training. Therefore, "[t]he crucial test for whether there is a difference between speedup and automatization, as suggested by Segalowitz, is whether, longitudinally, a decrease in mean RT produces a significant decrease in CV with an accompanying increase in CV–RT correlation" (Hulstijn, van Gelderen, & Schoonen, 2009, p. 563). In research that adopts the within-subjects design, such as pre and posttest pedagogical intervention studies, these three criteria (decrease in RT, decrease in CV_{RT}, and increase in CV_{RT}–RT correlations between pre and posttest phases) should be met for true automatization.

Here we review empirical studies that used CV_{RT} in a visual LDT and (at least) partly adopted a within-subjects design, focusing on the three criteria. The first study that used CV_{RT} as an index of automatic processing was conducted by Segalowitz and Segalowitz (1993). In their second experiment, the participants performed an LDT with 284 English words and nonwords. The stimuli were 35 baseline words; 15 words repeated six times each (i.e., 90 items in total), which served as repetition items; 35 homophone words; and 124 nonwords. The results showed that CV significantly correlated with RT for the baseline words. For the repetition words, they reported a significant decrease of mean RT and increase of CV–RT correlation. Further, for the initially fast processing students, CV significantly correlated with RT both at the first and at the last presentation. In contrast, for the initially slow processing students, CV did not significantly correlate with RT at the first presentation but the correlation was significant at the last presentation, possibly due to the practice effect. These results showed some empirical evidence for automatization. As Hulstijn et al. (2009) pointed out, however, "the authors do not report whether the CV in the repetition data decreased from the first to the sixth response, nor whether the decrease, if obtained, was significant" (Hulstijn, et al., 2009, p. 559). Therefore, it was unclear if their research fully satisfied the three criteria described above.

Segalowitz, Watson, and Segalowitz (1995) demonstrated a single participant's variability of RT data in an LDT. The materials were 120 base words for which the objective frequency was different (four frequency "bands" and 30 words for each) and 120 pseudo words. Further, the authors selected 30 additional words, 15 of which were used in the textbook in the course that the participant was taking and 15 that were not, and 30 corresponding pseudo words. The participant performed an LDT four times over a period of three weeks. Only Band 1 and words that were in the textbook had a tendency of reduced CV scores over time. The separate analyses of data with 10 words with reading experience and 12 control words showed that the change of CV score was significant for words with reading experience, but not significant for control words. These results meant that at least one of three criteria for true automatization was met (decrease of CV_{RT}), but as was the case with Segalowitz and Segalowitz (1993), it was not clear if all of the three criteria were met.

Next, Segalowitz, Segalowitz, and Wood (1998) studied 105 Canadian students learning French. They performed an LDT in six sessions over the period of one academic year. The materials were 300 words (consisting of 210 baseline words assumed to be known by the participants and 90 lesson words taken from class materials; thus, 35 baseline words and 15 lesson words for each session) and 300 pseudo words. The RT for the initially fast processing group significantly correlated with CV_{RT} throughout the research period. RT for the initially slow processing group did not significantly correlate with CV_{RT} at the initial test, but did at the other two tests. Further, the speed gain score and automaticity gain score for each participant were measured by subtracting initial RT from final RT and initial CV_{RT} from final CV_{RT}, respectively. There were significantly positive correlations between RT gain scores and CV_{RT} gain scores for both the initially fast and initially slow groups. Note, however, that this kind of gain data is misleading because "it could stem from the separation of the initial and final scores, as intended, or it could be primarily a function of either the first score or the second score" (Segalowitz, et al., 1998, p. 61). The authors therefore partialled out each participant's initial score from his or her final score and the residuals were used for the analyses; results showed that both groups increased automatic processing of visual word recognition. As Hulstijn et al. (2009) pointed out, however, the authors did not report whether the decrease of CVs of the two groups were significant; therefore, again, we do not know if the three criteria were fully satisfied.

Akamatsu (2008) asked 49 Japanese learners of English to draw lines to separate words in a string of letters (e.g., sunbendgivebearpen) over seven weeks (one session per week). The words were 150 monosyllabic English words of which 50 were target words in the LDT. In the LDT, 25 high frequency and 25 low frequency words and 50 pseudo words were used. The CV_{RT} score of low frequency words dropped significantly as a result of the training, while the score of high frequency words did not. Further, correlational analyses showed that RT and CV_{RT} significantly correlated with each other for low frequency words both in the pre and posttests, but not for high frequency words, either in the pre or posttests. In addition, the correlation between RT gain scores and CV_{RT} gain scores was significant for low frequency words but not for high frequency words. The same results were obtained for the residualized RT and CV_{RT} scores. Akamatsu's experiment met two criteria of true automatization (decease of RT and decrease of CV), but the last criterion (the correlation between CV–RT) decreased instead of increased (Hulstijn et al., 2009).

Finally, Hulstijn et al. (2009) reported CV data of previously published studies that had not originally shown CV analyses. First, van Gelderen, Schoone, Stoel, de Glopper, and Hulstijn (2007) tracked changes in reading comprehension of first language (L1) Dutch and L2 English by 389 learners using an LDT and sentence verification task. Data were collected when the participants were in Grade 8, 9, and 10. The results showed that, as opposed to Segalowitz' view of automatization, CV for LDT did not change significantly. Further, the CV–RT correlation remained relatively low, although significant *p* values were sometimes obtained, possibly because of the large number of participants. Second, Fukkink, Hulstijn, and Simis (2005) conducted an experimental training study with Grade 8 students with L1 Dutch and L2 English. The target words were 100 frequent words and 90 pseudo words. Word targets consisted of 40 trained words, 40 control words (appearing only in the pre and post-LDT), and 20 "context words" that appeared in the exercise. A series of analyses reported in Hulstijn et al. revealed that CV_{RT} did not change significantly from pre to posttest in most cases. The two studies reported in Hulstijn et al., therefore, failed to show support of true automatization.

In sum, results of previous studies are not straightforward. Some reported positive results for automatization while others did not. Further, some studies did not fully report the three criteria for true automatization; therefore, to date, the nature of automatization analyzed by the CV perspective is not clear. The other limitation of previous studies is that the nature of automatization was not clarified. According to Segalowitz's view, automatization is accompanied by qualitative changes or restructuring of the underlying cognitive system. From this statement, however, it is not clear *what* is qualitatively changed. Because CV is a tool for rejecting the simple-speedup null hypothesis (see Segalowitz, 2010 for details), even though CV values meet the three criteria for true automatization, the index does not explain the nature of automaticity per se. When considering this issue, therefore, we need to turn to other aspects of skilled visual word recognition. Given that processing of a word is carried out based on representation, orthographic representation may reflect the underlying qualitative change that takes place during the acquisition of automatic processing.

The Development of Orthographic Representation

According to Perfetti (1992), the development of representation can be understood as a process of increasing the precision of orthographic representation. Preciseness of orthographic representation is important because "[t]he advantage of a fully specified representation is that it is *determinant* with respect to the input features that will trigger it" (Perfetti, 1992, p. 157). This means that only the given word can activate its representation, rather than other, similar-looking words. Thus, if L2 learners do not have a precise orthographic representation, they may confuse the word with other similar-looking words (Ehri, 1995). This phenomenon has been widely reported in L2 reading research (Bensoussan & Laufer, 1984; Laufer, 1988). For example, Laufer (1988) reported that L2 learners have trouble differentiating words that are similar in form (e.g., *comprehensive* and *comprehensible*). Another related phenomenon is that L2 learners often perceive *unknown* words as similar-looking *known* words when reading L2 texts (Frantzen, 2003; Huckin & Bloch, 1993; Koda, 1997; Laufer, 1997). These errors can be attributed to impreciseness of the orthographic representation, since, as Perfetti suggests, if the learner has precise orthographic representation, similar-looking words do not activate the representation of the given word.

The assumption of the development from partial to precise orthographic representation can be empirically tested using the masked form-priming technique. The typical procedure is to present a row of hash marks (#####) followed by the prime (shown in lower case for around 50 ms) and then the target (shown in upper case). Participants are asked to perform some task (e.g., naming, lexical decision) on the target. The reason for presenting the prime and the target in different cases is "to ensure that the two stimuli are physically distinct" (Forster, Mohan, & Hector, 2003, p. 5).

Recently, this technique has been applied in developmental research. This can be achieved by taking into account neighborhood (*N*) metrics of the target stimulus. *N* is typically operationalized as the number of words that can be created from a particular word when one letter is changed (e.g., sale, male, safe, etc.). *N* is "a broad metric of the similarity of a word to other words" (Castles, Davis, Cavalot, & Forster, 2007, p. 167), implying that there are many similar words for high-*N* words. Therefore, it is assumed that these words require the development of orthographic representation; that is, if the representation of these words is not precise, the person frequently makes errors in recognizing the words. On the other hand, for low-*N* words, the development of the representation may not be as important, since only a few similar-looking words exist; hence, errors of word recognition do not occur frequently.

This is evident in L1 adult word recognition research, in which adults usually show facilitative priming effects when the target is a low-*N* word, while such effects are not observed when the target is a high-*N* word, when they perform a masked form-priming task (Forster, Davis, Schoknecht, & Carter, 1987). Thus, *N* metrics are useful in developmental research, because *N* values of the same word are consistent for adults, but change from low to high gradually over time for children (because *N* values are dependent upon vocabulary size). When children's written vocabulary is small, a high-*N* word is actually a low-*N* word in their mental lexicon, but later, as their vocabulary grows, the same word becomes a high-*N* word, resulting in the development of orthographic representation of the word. Therefore, it can be assumed that the priming effect should be observed when participants' orthographic representation is not precise, while the effect on the same word should be reduced when their orthographic representation becomes precise, in the case of high-*N* words. For this reason, the use of high-*N* words as experimental stimuli is theoretically important.

Castles et al. (2007) investigated English-speaking children's development of a word recognition system by a masked form-priming LDT. The participants in their experiment were 23 Grade 3 children and 24 adults. The Grade 3 children were re-tested two years later when they were in Grade 5 (n = 18). The researchers used 27 high frequency and high-N words. The primes used were substitution neighbors (SN), transposition neighbors (TN), and controls. SN primes were created by changing one letter from the target (e.g., *rlay* for *PLAY*) while TN primes were created by changing the position of two adjacent letters within the target (*lpay* for *PLAY*). Control primes were letter strings that did not share any letter in any position with the target (*meit* for *PLAY*). The position of substitution and transposition were varied almost equally across the primes (i.e., at the beginning, middle, and the end of the letter string). The rationale for using two different priming forms was that both types of primes were similar in form to the target, but the degree of similarity was different, that is, TN pairs are more similar to each other than SN pairs (Davis, 2006).

The results showed that adults' processing of the targets was not influenced either by the SN or TN, suggesting that their orthographic representation of these high-N words was precise and word recognition system was finely tuned (i.e., only the input stimulus that perfectly matched the internal orthographic representation could activate it). On the other hand, word recognition of Grade 3 children was roughly tuned and the orthographic representation was not precise, so that two types of primes could activate the target's representation, producing priming effects in the two prime conditions. Two years later in Grade 5, those children's word recognition system had developed because the vocabulary size became larger. Therefore, at this stage, SN primes no longer had the power to activate the target. However, since the word recognition system in Grade 5 was still developing and the orthographic representation was not completely precise yet, TN primes could activate the target, yielding the significant priming effect only in the TN condition.

The results of Castles et al. (2007) suggest that the word recognition system and orthographic representation develop, and that the masked form-priming procedure with SN and TN primes can be used to reflect the nature of orthographic representation. That is, by looking at the change of priming pattern, the development of orthographic representation can be investigated. However, only one study has investigated this in the L2 environment. Kida and Morita (2014) was the first study to investigate L2 learners' orthographic representation. The participants in their experiment were adult Japanese EFL learners. The experimental stimuli and procedure used in their experiment were almost the same as the original experiment by Castles et al. Results were that the SN and TN conditions showed similar facilitative priming effects. This finding was consistent with that of Castles et al.'s experiment with Grade 3 students, suggesting that, even in adults, the word recognition system and orthographic representation in L2 were at a relatively early stage of development.

Kida and Morita (2014) demonstrated that the masked form-priming LDT with SN and TN primes is applicable to adult L2 experiment; however, it is not clear if L2 orthographic representation of EFL learners can be changed over time in situations where EFL learners are exposed to large amounts of written L2 input. As Perfetti (2007) pointed out, it is assumed that the preciseness of orthographic representation depends on experience with words. Therefore, if learners had the opportunity for intensive exposure to L2 input, we could observe priming effects that were different from those obtained in Kida and Morita.

Research Questions

In sum, a mature word recognition system should contain automatic processing of words and precise orthographic representation. To date, it is unclear how automatic processing of words and orthographic representations relate to each other and how the two are acquired over time along with overall word recognition development. Therefore, the present study examines the acquisition of automatic word recognition and the development of orthographic representation using a pre-post within-subjects design. Two research questions were addressed as follows:

1. Is automatic word recognition acquired over time by adult EFL learners?
2. Is the development of orthographic representation achieved over time by adult EFL learners?

The hypothesis is that, if participants acquire automatic word recognition, we would observe a reduction of mean RT and CV_{RT} as well as an increase of the RT–CV_{RT} correlation from the pretest to the posttest. Further, if they achieve the development of L2 orthographic representation, we would observe a change of priming patterns in a masked form-priming LDT.

Method

Participants

Participants were 41 Japanese EFL students who enrolled in a compulsory English reading course at a national university in Japan. Students in this course were recruited for the present study because the course introduced the in-class sustained silent reading (SSR) activity with graded readers (Penguin Readers) in addition to standard text-based reading activities and tasks. Because large amounts of written exposure to L2 English is necessary for both automatic word recognition and L2 orthographic development, students enrolled in this course were deemed appropriate for the present study.

All participants provided informed consent before treatment and participation was voluntary. In order to analyze the data of participants who were most exposed to written L2 input, data from 21 students (12 males and 9 females) who attended all of the 12-week SSR activity were used for the analyses. Most of the students began learning English in junior high school at age 12 and had at least six years of formal English instruction. According to scores on the Test of English for International Communication (TOEIC)—a standardized English proficiency test developed by the Educational Testing Service—the students' English proficiency was intermediate. Participants' background information and their achievements in self-paced in-class SSR with graded readers are shown in Table 1 and Table 2, respectively.

Table 1. *English learning experience of the participants*

		M	*SD*	Minimum	Maximum
Age		18.67	0.58	18	20
Beginning age of English learning		12.38	1.16	9	13
Years of formal instruction		6.62	1.07	6	10
TOEIC score		565.48	98.89	330	735
Self-assessed rating:	Speaking	4.00	1.10	2	6
	Listening	4.24	1.34	2	7
	Reading	5.33	1.43	3	8
	Writing	4.67	1.20	3	7

Note: Self-assessed ratings indicate how proficient participants are in each skill from 1 (*minimum proficiency*) to 10 (*near-native proficiency*).

Table 2. *Number of books read by participants during the course*

	TOEIC Score	Easy Starts	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Tot.
Participant 1	640	1	2	2	1	0	0	0	6
Participant 2	650	5	1	1	0	0	0	0	7
Participant 3	580	0	6	0	0	0	0	0	6
Participant 4	515	0	0	0	3	0	0	0	3
Participant 5	500	0	0	8	2	1	0	0	11
Participant 6	640	0	0	3	0	0	0	0	3
Participant 7	655	0	0	2	2	0	0	0	4
Participant 8	460	3	1	1	1	0	0	0	6
Participant 9	605	0	0	4	0	0	0	0	4
Participant 10	535	6	3	3	0	0	0	0	12
Participant 11	595	0	0	3	0	0	0	0	3
Participant 12	415	0	0	0	3	0	0	0	3
Participant 13	730	0	0	1	0	1	0	0	2
Participant 14	520	0	0	2	0	0	0	0	2
Participant 15	330	7	0	0	0	0	0	0	7
Participant 16	540	3	1	2	0	0	0	0	6
Participant 17	520	0	3	2	1	0	0	0	6
Participant 18	735	0	0	5	0	0	0	0	5
Participant 19	495	0	0	0	0	1	0	0	1
Participant 20	590	0	2	2	2	0	0	0	6
Participant 21	625	0	0	3	3	0	0	0	6

Course Description

The in-class SSR activity was administered over 12 weeks. Students attended one class every week and each class was 90 minutes long. The teacher gave students various reading and vocabulary tasks together with activities based on passages from the textbook. After these text-based activities, the SSR activity was administered in the last 30 minutes of class. The Penguin Readers from *Easystarts* to *Level 6* were used in the course. All published books in the series were introduced in the course. Based on the general instructions for pleasure reading (e.g., Day & Bamford, 1998), participants were encouraged to (a) read books as much as possible, (b) choose books that meet their interests, (c) read books with pleasure, (d) change books any time they wanted if the book was not interesting or if it was too easy or difficult for them, and (e) avoid using the dictionary frequently.

Experimental Materials

The present study tried to select high-frequency basic words used in previous studies, which appeared in graded readers of any themes as well as any levels. Therefore, the main stimuli used in the present experiment were borrowed from Kida and Morita (2014), which adapted the stimuli used in Castles et al. (2007) to suit Japanese learners of English. These words were four or five letter high frequency English words. Kida and Morita (2014) replaced eight words from the original study by Castles et al., since no information about familiarity of Japanese EFL learners with these words was available. In order to choose replacement words, the following steps were taken (Kida & Morita, 2014).

(1) CELEX and Kucera-Francis frequency, and *N*-size of the original eight words were checked using the *N*-watch software (Davis, 2005).
(2) Based on the standard English word familiarity rating list for Japanese EFL learners developed by Yokokawa (2006), words with familiarity ratings above 4.0 (based on a 7-point scale) were selected and eight words were subsequently chosen with similar Kucera-Francis frequency, number of letters, and *N*-size to the original eight words.
(3) SN, TN, and control primes for these eight words were created in the same way as Castles et al. (2007).
(4) Pseudo words for the *no* response in the LDT were created based on the ARC nonword database (Rastle, Harrington, & Coltheart, 2002), and the number of letters and *N*-size of each pseudo word was matched to words for the *yes* response in the LDT.
(5) SN, TN, and control primes for these pseudo words were created in the same way as the yes response words.

Using the described procedure, 27 word targets and corresponding nonword primes, and 27 pseudo word targets and their nonword primes, were selected. Three counterbalanced lists were created from these 54 stimuli using a Latin square design. In this design, each participant was randomly assigned to one of the three lists, so that each target word was shown only once to one participant, however, across lists, all target words were shown in the three experimental conditions (SN, TN, and control conditions). Thus, participants would not encounter the same word more than once in an experiment.[1]

Apparatus and Procedure

Epson ST12E with Windows 7 Professional computers (32-bit, Core 2 Duo CPU, 2.00 GB RAM) was used in the experiment. The DMDX program (Forster & Forster, 2003) was used for the presentation of items and measurement of RTs and error rates of the LDT. The screen refresh rate was 16.67 ms.

LDT data were collected before and after the 12-week SSR. Participants were assigned to the same counterbalanced list condition of the LDT at pre and posttests. Procedures for the two phases were the same. First, participants read the instructions for the LDT. They were asked to judge as quickly and accurately as possible whether the presented letter strings were an English word. Then, the practice session began with items that were not used in the main session. In the practice session, five hash marks (#####) appeared on the computer screen for 800 ms. Immediately after, a prime was presented for 50 ms in lower case. Then, the target appeared in upper case until the participant's judgment was made. After the judgment, feedback was given to participants in terms of the accuracy of the judgment and the RT. Following the feedback, five hash marks for the next stimulus appeared, and so forth. Five items were used for the yes response and the other five for the no response. After the end of the practice session, the main session began. The procedure of the main session was the same as that of the practice session, except that participants did not receive any feedback. The presentation order of the experimental stimuli was pseudo-randomized for each participant.

The data trimming procedure adopted in the present experiment was the same as that used by Castles et al. (2007) and Kida and Morita (2014). Only correct responses were analyzed. Responses faster than 150 ms were treated as errors. Through this procedure, no data were treated as outliers.

In the following sections, statistical analyses are reported for the acquisition of automatization and the development of orthographic representation, separately. The acquisition of automatization was analyzed using RT data from the control condition for pre and posttests. Theoretically, the presentation of the control primes do not affect the processing of the target and, therefore, the results of the control condition can be treated as those of the normal LDT. Factors in the analyses were the version of the experiment (list 1, list 2, list 3) and time (pre and post). The variable, version of experiment, is not the theoretical focus of the present study. Further, following methods used in previous studies, correlational analyses were conducted. Next, the development of orthographic representation was analyzed by the RT and error data, separately. Factors were the version of the experiment (list 1, list 2, list 3), time (pre, post), and prime (SN, TN, Control). Again, the version of experiment was not the theoretical focus. In addition, separate analyses were conducted for participant analysis (F_1) and item analysis (F_2).

Results

The overall results are shown in Table 3. The results show that (a) both SN and TN prime conditions yielded similar amounts of priming effects in the pretest and posttest, and (b) the overall RTs in each condition became faster from the pretest to the posttest.

Table 3. *Mean and standard deviation of reaction time (RT) and error rates of each experimental condition at pre and posttest phases*

Prime type	Example	RT (ms)	Error rates (%)	Priming (ms)
Pretest				
Substitution	rlay/PLAY	1028 (258)	4.23 (5.53)	70
Transposition	rpay/PLAY	1011 (236)	6.88 (7.43)	87
Control	meit/PLAY	1098 (276)	10.05 (12.12)	
Posttest				
Substitution	rlay/PLAY	887 (174)	5.82 (5.69)	75
Transposition	rpay/PLAY	888 (156)	8.47 (13.57)	74
Control	meit/PLAY	962 (184)	6.35 (8.29)	

The Acquisition of Automatization: CV_{RT} Analyses

Descriptive statistics are shown in Table 4. There was no significant interaction of the two variables in the 3 (list) × 2 (time) analysis of variance (ANOVA) for RT in the control conditions (in the pre and posttest phases), $F\ (2, 18) = 0.85$, $p = .44$, partial $\eta^2 = .09$. This suggests that the main effect of time was not modulated by the list × time interaction. However, the main effect of time was significant, $F\ (1, 18) = 6.32$, $p = .02$, partial $\eta^2 = .26$. A separate ANOVA, with CV_{RT} as the dependent variable, also showed no significant interaction, $F\ (2, 18) = 0.70$, $p = .51$, partial $\eta^2 = .07$, again suggesting that the main effect of time was not modulated by the list × time interaction. However, in this case, no significant main effect of time was observed, $F\ (1, 18) = 2.62$, $p = .12$, partial $\eta^2 = .13$. The correlations between CV_{RT} and RT at the pre and posttests were significant, $r = .69$, $p < .01$ and $r = .75$, $p < .01$, respectively. These results suggest that two of the three criteria for true automatization were satisfied (decrease in RT and increase in CV_{RT}–RT correlations between pre and posttest phases).

Table 4. *Descriptive statistics of mean reaction time (RT) and mean cv_{rt} in the pre and posttest*

	Mean RT	*SD* RT	Mean CV_{RT}	*SD* CV_{RT}
Pretest	1098	276	0.28	0.14
Posttest	962	184	0.22	0.10

Note. CV_{RT} = ratio of the *SD* of RT to the mean RT.

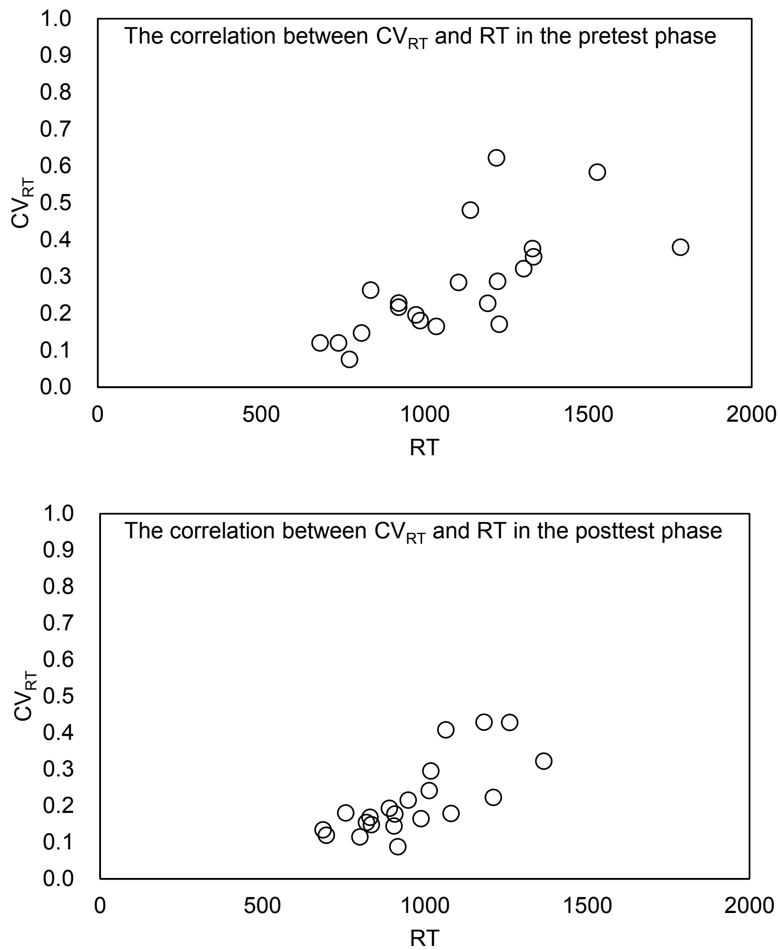

Figure 2. Correlations between coefficient of variation (CV_{RT}) and reaction time (RT) in pre and posttest phases.

In addition to the three criteria, the correlational analyses between RT gain scores and CV_{RT} gain, and between the residualized RT and CV_{RT}, were conducted in the same manner as in Segalowitz et al. (1998) and Akamatsu (2008). The correlations of gain scores and residualized scores were significant, $r = .52$, $p = .02$ and $r = .57$, $p = .01$, respectively. A series of these analyses showed that most criteria for automatization were satisfied and, therefore, participants increased automatic processing of visual word recognition by the end of the 12-week SSR activity.

The Development of Orthographic Representation: Priming Effects Analyses

Reaction time analyses. Results for the 3 (list) × 2 (time) × 3 (prime) ANOVA showed that there was no significant three-way interaction, F_1 (4, 36) = 1.66, $p = .18$, partial $\eta^2 = .16$, and F_2 (4, 32) = 0.98, $p = .43$, partial $\eta^2 = .11$. This suggests that the two-way (time × prime) interaction, the main focus of the current analysis, was not moderated by the three-way interaction. The interaction between time and prime was not significant for either participant analysis, F_1 (2, 36) = 0.25, $p = .78$, partial $\eta^2 = .01$ or item analysis, F_2 (2, 16) = 0.73, $p = .50$, partial $\eta^2 = .08$. The main effect of time was significant for both participant analysis, F_1 (1, 18) = 11.06, $p < .01$, partial $\eta^2 = .38$, and item analysis, F_2 (1, 8) = 29.57, $p < .01$, partial $\eta^2 = .79$. Both analyses showed relatively large effect sizes. The main effect of prime was also significant for the participant analysis, F_1 (2, 36) = 9.35, p < .01, partial η^2 = .32, and approached significance for the item analysis, F_2 (2, 16) = 3.36, p = .06, partial η^2 = .30.[2] The item analysis did not show a significant effect, but this might have been because of the small number of items. This is reflected in the moderate effect size. A pairwise comparison between conditions with the modified sequentially rejective Bonferroni procedure (Shaffer, 1986) showed that the difference between the SN and the control condition was significant, t (18) = 3.05, p < .01. The difference between the TN condition and the control condition was also significant, t (18) = 3.99, p < .01. No significant difference was found between the SN and the TN conditions, t (18) = 1.11, p = .28.

Error rate analyses. The same 3 × 2 × 3 ANOVAs were conducted for the error rates. The results showed that there was no significant interaction between the three variables, both for F_1 (4, 36) = 0.26, p = .90, partial η^2 = .03, and F_2 (4, 36) = 0.55, p = .70, partial η^2 = .06. As with the RT analyses, the non-significance and small effect sizes suggest that the interaction between time and prime was not modulated by the three-way interaction. The two-way interaction between time and prime was not significant either in the participant analysis, F_1 (2, 36) = 1.10, p = .34, partial η^2 = .06, or the item analysis, F_2 (2, 18) = 1.52, p = .25, partial η^2 = .14. This suggests that the patterns of error rates across the two data collection points were similar, but this may be possibly because of the floor effect. The main effect of time was not significant both for F_1 (1, 18) < 0.01, p = .96, partial η^2 < .01, and F_2 (1, 9) < 0.01, p = .95, partial η^2 < .01. These results were due to the floor effect. The main effect of prime was not significant for F_1 (2, 36) = 0.94, p = .40, partial η^2 = .05, and F_2 (2, 18) = 0.95, p = .41, partial η^2 = .10, also due to the floor effect.

Discussion

The present experiment examined the following two research questions: (1) Is automatic word recognition acquired over time by adult EFL learners? (2) Is the development of orthographic representation achieved over time by adult EFL learners? The results clearly showed a positive response to the first question but not to the second. In terms of the acquisition of automatization in L2 word recognition, the participants in the present experiment exhibited signs of increasing automaticity of word recognition. As discussed in the literature review, previous research investigating automatization through the CV perspective has been limited in terms of the data they reported (i.e., the three criteria for true automatization discussed in the literature review section). Therefore, it was unclear whether it is possible to achieve true automatization in L2 acquisition or whether using the CV approach is appropriate to investigate the automatization process in L2 reading acquisition. The results of the present study, however, showed that the CV approach is a useful tool to observe L2 learners' change in processing of visual word recognition, and that most criteria for true automatization could be satisfied over time. This suggests that the acquisition of automatic processing of visual word recognition in L2 is an achievable goal of EFL education.

As for the development of L2 orthographic representation, however, the present study did not show any evidence. The interaction between time and prime was not significant for either the participant analysis or item analysis. The non-significance and small effect sizes (partial η^2 = .01

for the participant analysis and partial $\eta^2 = .08$ for the item analysis) suggest that the patterns of priming across the two data collection points were similar, and, therefore, that participants' L2 orthographic representation did not develop. Compared to the developmental shift shown in Castles et al. (2007), the present study showed that participants' orthographic representation did not change from the pretest to posttest phase. Further, the pattern of priming effects was such that both SN and TN primes showed similar and significant priming effects. This result is consistent with the finding for Grade 3 native English-speaking children shown in Castles et al., which suggests that the orthographic representation of the present study participants remained at a relatively earlier stage of development. This is also consistent with the results of Kida and Morita (2014) in which L2 orthographic representation of Japanese EFL learners stayed at a relatively earlier stage of development. Combined with previous studies, therefore, the results of the present experiment suggest that the development of L2 orthographic representation is not easy to achieve in an EFL context.

The present results suggest asymmetrical development of automatic word recognition and orthographic representation in L2. That is, the acquisition of automatic processing precedes the development of orthographic representation in overall L2 visual word recognition development. The present study showed that the acquisition of automatic processing can be achieved without fully specified (precise) orthographic representation. Within the framework of recognition development shown in Figure 1, this means that the automatization stage (the middle stage in Figure 1) does not require the development of orthographic representation. In other words, the development of orthographic representation is not a prerequisite for automatization. Therefore, orthographic representation is not the cognitive system that experience qualitative change or restructuring as was proposed by Segalowtiz. Future research is expected to further explore how the acquisition of automatic processing and the development of orthographic representation relate each other and what cognitive mechanisms contribute to automatization.

There were several limitations to the present study. First, there is a possibility that asymmetrical development of automatic processing and orthographic representation, the main finding of the present study, was merely a methodological artifact. The two aspects (automatic and orthographic) of visual word recognition were analyzed using two different paradigms (i.e., CV values and the pattern of priming effects). Future research is necessary to overcome this limitation. Second, the present study did not reveal potential effects of participants' L1 orthographic system. Some previous studies (e.g., Akamatsu, 2003; Koda, 1990) showed that non-alphabetic L1 learners such as Japanese EFL learners are less efficient in processing alphabetic language such as English when compared to alphabetic L1 learners. Because there is a possibility that L2 learners with alphabetic and non-alphabetic L1 backgrounds process LDT in a different way, we can hardly make a claim about English learners from other L1 backgrounds. Third, the statistical analyses were limited in several ways. Statistical power may have been insufficient because data were analyzed for only 21 participants and 27 words. In addition, because only the control condition data were used in the CV analyses, the nonsignificant difference for CV_{RT} could be due to the small numbers of both participants and items. Further, because this study used separate statistical analyses for CV and priming, the issue of increased Type I error should be carefully considered when interpreting the results. Finally, the results obtained could not be attributed merely to the effect of the in-class SSR activity since it is difficult to exclude the possibility that learners were exposed to target words outside of the course. These limitations should be addressed in future research.

Conclusion

The present study demonstrated that Japanese EFL learners could acquire automatic processing of L2 visual word recognition over time. Two out of three criteria for true automatization were satisfied (decrease in RT and increase of CV_{RT}–RT correlation). Furthermore, the experiment showed correlations between RT and CV_{RT} gain scores and residualized RT and CV_{RT} scores. Although L2 learners acquired automatic processing, they did not develop L2 orthographic representations. The pattern of priming effects in the SN and TN conditions did not change over time, suggesting that orthographic representation level remained at a relatively earlier stage of development. Taken together, the acquisition of automatic processing and development of orthographic representation do not occur simultaneously and their development is asymmetrical.

Acknowledgments

I would like to thank the reviewers for their insightful comments and suggestions for revisions to earlier versions of the manuscript. This work was supported by the Japan Society for the Promotion of Science (JSPS) Grant Number 24720258, Grant-in-Aid for Young Scientists (B).

Notes

1. Because we analyzed data for 21 participants who attended the entire 12-week SSR activity, the experimental design resulted in an unbalanced number of participants in each list condition (eight participants in each of lists 1 and 2 and five in list 3). However, no significant pre-existing difference was observed between the three list groups in terms of their general English proficiency based on their TOEIC scores (list 1, M = 544. 38, SD = 104.83; list 2, M = 552.00, SD = 90.32; list 3, M = 595.00, SD = 103.23), F (2, 18) = 0.56, p = .58, partial η^2 = .06.

2. The main effect of prime was modulated by the significant list × prime interaction in the participant analysis, F_1 (4, 36) = 10.21, $p < .01$, partial η^2 = .54, but not in the item analysis, F_2 (2, 16) = 1.45, p = .26, partial η^2 = .15. It was the only case where a significant list effect was found. However, this was not considered to be a problem because the main focus of the priming analysis is the two-way interaction between time and prime, and this interaction was not modulated by the list factor. Therefore, the significant list × prime interaction in the participant analysis does not affect the main conclusion of the present study.

References

Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning*, *53*, 207–231. doi:10.1111/1467-9922.00216

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, *29*, 175–193. doi:10.1017/S0142716408080089

Akamatsu, N. (2009, September). Bairingaru mentaru rekishikon no tankyu: L2 tango ninshiki kenkyu [Exploring the bilingual mental lexicon: Research on L2 word recognition]. In S. Kadota (Chair), *Motto hannoujikan deta no katsuyou wo!!: Kenkyu oyobi kyoshitsu de ikani tsukauka* [*Let's use reaction time data more: How we could use it in research and classrooms*]. Symposium conducted at the meeting of JACET 48th Annual Convention Program, Sapporo.

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, *7*, 15–32. doi:10.1111/j.1467-9817.1984.tb00252.x

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*, 165–182. doi:10.1016/j.jecp.2007.01.006

Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 151–179). Psychology Press.

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*, 65–70. doi:10.3758/BF03206399

Davis, C. J. (2006). Orthographic input coding: A review of behavioural data and current models. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 180–206). Psychology Press.

Day, R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge University Press.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge University Press.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recording. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*, 116–125. doi:10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C. (2005). Learning to read words: Theory, findings and issues. *Scientific Studies of Reading*, *9*, 167–188. doi:10.1207/s1532799xssr0902_4

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *Quarterly Journal of Experimental Psychology*, *39A*, 211–251. doi:10.1080/14640748708401785

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. doi:10.3758/BF03195503

Forster, K. I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 3–37). Philadelphia, PA: Psychology Press.

Frantzen, D. (2003). Factors affecting how second language Spanish students derive meaning from context. *The Modern Language Journal*, *87*, 168–199. doi:10.1111/1540-4781.00185

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *The Modern Language Journal*, *89*, 54–75. doi:10.1111/j.0026-7902.2005.00265.x
Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153–176). Norwood, NJ: Ablex.
Hulstijn, J. H., van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*, 555–582. doi:10.1017/S0142716409990014
Kida, S., & Morita, M. (2014). *Development of L2 orthographic representation in late Japanese-English bilinguals*. Manuscript in preparation.
Koda, K. (1990). The use of L1 reading strategies in L2 reading: Effects of L1 orthographic structures on L2 phonological recoding strategies. *Studies in Second Language Acquisition*, *12*, 393–410. doi:10.1017/S0272263100009499
Koda, K. (1997). Orthographic knowledge in L2 lexical processing: A cross-linguistic perspective. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 35–52). Cambridge University Press.
Laufer, B. (1988). The concept of 'synforms' (similar lexical forms) in vocabulary acquisition. *Language and Education*, *2*, 113–132. doi:10.1080/09500788809541228
Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge University Press.
Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.
Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383. doi:10.1080/10888430701530730
Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362. doi:10.1080/02724980244000099
Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*, 369–385. doi:10.1017/S0142716400010845
Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*, 53–67. doi:10.1017/S0142716400010572
Segalowitz, N., Watson, V., & Segalowitz, S. (1995). Vocabulary skill: Single-case assessment of automaticity of word recognition in a timed lexical decision task. *Second Language Research*, *11*, 121–136. doi:10.1177/026765839501100204
Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of American Statistical Association*, *81*, 826–831. doi:10.2307/2289016
Share, D. L. (1995). Phonological recoding and self-teaching: *sine qua non* of reading acquisition. *Cognition*, *55*, 151–218. doi:10.1016/0010-0277(94)00645-2
Van Gelderen, A., Schoonen, R., Stoel, R. D., de Glopper, K., & Hulstijn, J. H. (2007).

Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, *99*, 477–491. doi:10.1037/0022-0663.99.3.477

Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830–841. doi:10.1037/0033-295X.114.3.830

Yokokawa, H. (Ed.). (2006). *Nihonjin eigo gakusyusya no eitango shinmitsudo: Mojihen* [*Familiarity ratings for visual English words by Japanese English learners*]. Tokyo: Kuroshio.

Appendix

The Experimental Stimuli of Kida and Morita (2014) Employed in the Present Study.

Word Target	SN prime	TN prime	Control prime	Non-Word Target	SN prime	TN prime	Control prime
ANGRY	angyr	angrf	twiek	COWCE	cowec	cowcp	sturc
<u>CRIME</u>	rcime	cgime	wholk	PLAIL	lpail	pqail	jerge
BAND	badn	banc	lese	KILE	kiel	kilg	froy
<u>LOCK</u>	olck	kock	smat	TOVE	otve	gove	waim
DARK	drak	derk	bleg	FUSK	fsuk	fzsk	palt
EACH	eahc	eafh	ibnd	ZOLC	zocl	zoxc	dabe
FAST	afst	gast	eben	DITE	idte	xite	lurg
GLAD	gald	ghad	porf	NURF	nruf	nqrf	cyce
HATE	haet	hati	obok	ZEAR	zera	zeai	bolf
HEART	herat	heaet	spliz	TREBE	trbee	treme	wharn
HORSE	ohrse	gorse	dwaul	SHARN	hsarn	kharn	chibe
<u>BIKE</u>	ibke	bdke	toal	DIRP	idrp	durp	yash
<u>COIN</u>	cion	coyn	desh	DERN	dren	dewn	nolm
KICK	kikc	kicm	glon	HAMP	hapm	hamx	grig
<u>TWIN</u>	wtin	pwin	yoap	FILP	iflp	iilp	nean
<u>RELAX</u>	rleax	rekax	smick	SLONT	solnt	slsnt	zeafe
NIGHT	ngiht	nilht	blaes	BRATE	barte	brlte	skuch
NORTH	nroth	nosth	bleck	DEACE	daece	delce	boint
NOSE	noes	nosp	beda	MALK	makl	malh	torp
<u>FRUIT</u>	rfuit	fwuit	bleve	GEAFE	egafe	grafe	splip
PLAY	lpay	rlay	meit	LOSP	olsp	rosp	jabe
SALE	slae	sase	obth	DEAT	daet	dejt	zick
SHAPE	shaep	shaie	diert	FLANE	flaen	flaxe	shrou
SLIDE	sldie	slire	crong	DRINE	drnie	drife	karsh
<u>MOUSE</u>	muose	mduse	trant	SPINK	sipnk	smink	yorce
THING	thnig	thiog	slark	SWICE	swcie	swige	drurt
WHITE	whiet	whitn	flarm	BREAP	brepa	breah	skaul

Note: The eight underlined words were replaced with words from the original study by Castles et al. (2007).

About the Author

Shusaku Kida is an associate professor at the Institute for Foreign Language Research and Education at Hiroshima University where he teaches English mainly to undergraduate students. He is interested in second language vocabulary acquisition and visual word recognition. E-mail: skida@hiroshima-u.ac.jp