Japanese Speakers’ Confusion of Phonemically Contrasting English Words: A Link Between Phoneme Perception and Lexical Processing

Ryo Komaki and Reiko Akahane-Yamada

ATR Human Information Science Laboratories, Kyoto, Japan
komaki@atr.jp, yamada@atr.jp

Abstract
This study examined the link between phoneme perception and lexical processing in a second language (L2) through four experiments on testing and training of word translation. Results showed that translation performance generally differed as a function of perceptual difficulty of phonemes in words. In contrast, perceptual difficulty did not always influence the translation performance during extensive laboratory training. These results suggest that the link between phoneme perception and lexical processing in L2 is stronger in longitudinal vocabulary acquisition than in the learning by short-term extensive training.

1. Introduction
The mental lexicon is said to include representations at many levels, including phonology, orthography, syntax, and semantics. In L2 vocabulary acquisition, learners face confusions at many levels.

With regard to the problem in phonological representation, L2 learners have difficulty distinguishing L2 phonemic contrasts that their native language (L1) does not have. It has been demonstrated that native speakers of Japanese have difficulty identifying /r/-/l/ contrasts in American English (AE) [1], and that the extent of this difficulty varies depending on the syllabic position where the contrast occurs in a word [2]. This “position effect” has been observed in several studies in /r/-/l/ perception [3, 4] and production [5], where Japanese listeners and non-Japanese listeners were compared.

Another serious problem for L2 learners concerns the failure in accessing appropriate semantic representations because of an instability in various information in their L2 mental lexicon. Several studies have found that phonological information notably played an important role in L2 mental lexicon. It was demonstrated that words in an L2 mental lexicon were strongly connected to each other with phonological links [6], and that L2 learners at a low-proficiency level relied on sound similarities between L2 words rather than associated meanings in L2 vocabulary learning [7]. These findings imply that learners of L2, especially at low-proficiency levels, tend to confuse the meaning of L2 words that are phonologically close to each other.

The present study focused on the relationship between L2 auditory perception and L2 lexical access. We hypothesize that one of the reasons for lexical confusion might be connected with perception of phonemes. In Experiment 1, we examined the possibility that lexical confusion in English words contrasting in /r/ and /l/ depends on the difficulty of phoneme perception manipulated by the position of the contrast in the word. Then, in Experiment 2, the relationship between lexical confusion and phoneme perception was further investigated through an extensive training in L1-to-L2 correspondence of these words. In Experiments 3 and 4, other phonemic contrasts were tested to vary the difficulty of phoneme perception.

2. General method
2.1. Participants
All participants were native speakers of Japanese, most of whom were undergraduates. None of them reported any history of speech or hearing impairment or have lived abroad for more than three months. All of them participated in only one of the four experiments.

2.2. Stimuli
Pairs of English words minimally contrasting in consonants such as /r/ and /l/ (“right”-“light”) were used. The stimulus words were presented auditorily and/or visually. For auditory presentation, they were pronounced by native speakers of AE and were saved into separate sound files at 16-bit resolution with a sampling frequency of 22,050 Hz. For visual presentation, Japanese words were also prepared, each of which was a representative translation for each English word.

2.3. Task
The following three tasks were used in combination with one of two modes (test mode/training mode). Participants were to carry out all the tasks in a two-alternative forced-choice (2AFC) paradigm. During the training mode, they received feedback (chime or buzzer) for each response, while they did not during the test mode. When they gave the wrong response in a trial within the training mode, the trial was repeated again as a “correction trial”.

Perception: A stimulus word was auditorily presented over headphones. Two alternatives were visually presented on the computer screen; one was the word audi-
torily presented (target), and the other was a counterpart of the minimal pair including the target (distractor). The task was to identify the auditory stimulus.

**Japanese-to-English (JE) translation:** A stimulus word in Japanese was visually presented on the screen, beneath which two alternatives in English were simultaneously presented; one was an English word corresponding to the presented Japanese word (target), and the other was an English counterpart to the target (distractor). The task was to choose an appropriate translation for the stimulus.

**English-to-Japanese (EJ) translation:** The task was the same as that in JE translation, but the direction of translation was reversed.

3. L2 Lexical confusion: /r/-/l/ contrast

3.1. Experiment 1: translation

3.1.1. Method

Stimuli consisted of 205 pairs of English words (410 words) contrasting in /r/ and /l/ in one of seven positions: initial singleton (IS), initial consonant cluster (ICC), initial triple consonant cluster (ITC), intervocalic (INT), medial consonant cluster (MCC), final consonant cluster (FCC), and final singleton (FS). Seven participants performed the JE/EJ translation task in the test mode separately into different blocks. Within each block, each stimulus was presented once in a random order. The order of the two blocks was counterbalanced across participants.

3.1.2. Results and discussion

An averaged translation accuracy was calculated for each participant, each translation direction, and each position except ITC and MCC, for the sake of comparison with the accuracy of phoneme identification [4]. A repeated-measures analysis of variance (ANOVA) was applied to arcsine-transformed accuracy scores, with translation direction and position as independent variables. The results indicated a significant main effect of position \(F(4, 24) = 39.13, p < 0.01\) (Figure 1). Then, Ryan’s method was applied for multiple comparison, which indicated that the accuracy significantly varied according to position as follows: ICC<IS<INT<FCC<FS \([MSe = 0.01, p < 0.05]\). In comparison with Komaki & Choi [4], the difficulty in perception and the difficulty in translation showed a similar tendency as a function of position.

Because there seemed to be a significant imbalance in familiarity and/or frequency of words by positions, the data was reanalyzed for a limited set of pairs. The pairs were selected so that they included at least one English word which Japanese undergraduate students would have typically learned. Results showed that the tendency as a function of position observed in the analysis of all pairs was still witnessed, implying a link between phoneme perception and lexical processing.

3.2. Experiment 2: Effect of training

3.2.1. Method

Stimuli consisted of 169 pairs of English words (338 words) contrasting in /r/ and /l/. The experimental session consisted of a pretest, repeated training blocks, and a post-test across six days. An incidental memory retention test was also conducted about five months after the post-test. Twelve participants performed the JE/EJ translation task in the test mode at each test. Balancing the pretest accuracy, they were divided into two groups. During the training, one group was trained using JE translation (N=6), and another was trained using EJ translation (N=6), in the training mode. All of them repeated a block of 338 trials for 18 times.

3.2.2. Results and discussion

An averaged translation accuracy was calculated for each participant, each test phase, each position, each translation direction of the test, and each training condition, and then normalized with an arcsine transformation method, to be submitted to a mixed model ANOVA. The results of the ANOVA, where test phase, position, and test direction were treated as within-subjects factors and training condition as between-subjects factor, indicated a significant main effect of test phase \(F(2, 20) = 223.81, p < 0.001\) and position \(F(4, 40) = 55.20, p < 0.001\), and a significant interaction between the two factors \(F(8, 80) = 3.39, p < 0.005\). This interaction is shown in Figure 2. The interaction among test phase, po-

![Figure 1: Translation accuracy as a function of position: Identification accuracy in Komaki & Choi [4] was shown in line for comparison.](image1)

![Figure 2: Translation accuracy as a function of position at three test phases.](image2)
position, and training condition was marginally significant \( F(8, 80) = 2.15, p < 0.05 \). Results of Ryan’s test for subsequent multiple comparisons are summarized as follows: (1) Regardless of training condition and position, translation accuracy significantly increased in the post-test but decreased in the retention test. (2) The performance in the retention test was significantly better than that in the pretest for INT, FCC, and FS, whereas it decreased to the level of the pretest for IS and ICC. (3) The position effect observed in the pretest disappeared (in JE training condition) or slightly remained (in EJ condition) in the post-test, but appeared again in the retention test. These results suggest that a sufficient amount of training using minimal pairs can improve L2 learners’ lexical confusion. Moreover, a robust position effect replicated 5-month after the training implies the presence of a link between phoneme perception and lexical processing.

4. L2 Lexical confusion: Various contrasts

4.1. Experiment 3: translation

4.1.1. Method
In order to vary the degree of difficulty of phoneme perception, we adopted four phonemic contrasts in addition to /r/-/l/, based on confusion matrices of AE consonants by native speakers of Japanese [8]. Stimuli consisted of 459 pairs of English words (918 words) minimally contrasting in one of five phoneme pairs: /r/-/l/ (RL: difficult), /b/-/v/ (BV: moderately difficult), /s/-/z/ (STH: moderately difficult), /b/-/g/ (BG: easy), and /p/-/k/ (PK: easy). Fifteen participants performed the JE translation task in the test mode. The stimuli were divided into five blocks by phonemic contrast and were presented once in a random order within each block. The order of the blocks was counterbalanced over participants.

4.1.2. Results and discussion
The arcsine-transformed accuracy averaged for each participant and each contrast was submitted to a one-way ANOVA. The results showed a significant main effect of contrast \( F(4, 56) = 49.62, p < 0.001 \) (Figure 3). Results of a subsequent test using Ryan’s method showed that the accuracy significantly differed according to phonemic contrast as follows: RL < BV < BG < PK < STH. Furthermore, the same re-analysis of the data as done in Experiment 1 still showed a significant main effect.

In comparison with the results of Yamada et al. [8], the difficulty in phoneme perception and in translation showed a similar tendency except for STH. It appears that the STH pairs were easier to translate than those of the other four contrasts due to visual dis-similarity of the orthography between members of each pair. There existed many pairs where each word had the same number of letters and differed by only one letter in RL, BV, BG, and PK, whereas there was no such pair in STH. This might have invoked better performance in STH. In summary, the link between phoneme perception and lexical processing is still implied by the results of the four contrasts except STH, though orthographical factors as well as phonological factors might influence lexical processing.

4.2. Experiment 4: Effect of training

4.2.1. Method
The overall experimental design was almost the same as that adopted in Experiment 2, except that (1) the BV, STH, BG, and PK contrasts in addition to RL were used, (2) the difficulty of the translation was controlled to be the same level among RL, BV, and BG, (3) multiple retention tests with shorter intervals were arranged, and (4) the Perceptual task in the test mode for perceptually difficult contrasts (RL, BV, and STH) was also conducted in the pretest and the post-test.

The RL, BV, STH, and BG pairs (40 pairs each) were assigned as experimental stimuli and the PK pairs (160 pairs) were assigned as filler stimuli and control stimuli. Seventeen participants took part in the experiment composed of two stages; the training stage consisted of a pretest, training blocks, and a post-test over seven days, and the retention stage consisted of the alternation of pseudo-training blocks and retention tests (3-days/1-week/2-weeks/3-weeks/1-month after the post-test) over 13 days. All the participants had to participate in the training stage but were able to finish the retention stage at any retention test.

4.2.2. Results and discussion
An averaged translation accuracy and identification accuracy were calculated for each participant, each test phase, each phonemic contrast, and each training condition, and then normalized with an arcsine transformation method to be submitted to a mixed model ANOVA. Analyses were conducted for two groups: (a) those who completed the training stage (all participants: N=17) and (b) those who completed the whole retention stage (N=8).

As for the translation performance, the ANOVA for (b) was run, where test phase, phonemic contrast
(RL/BV/STH/BG: PK was excluded because it appeared only in control stimuli), and test direction were treated as within-subjects factors and training condition was treated as a between-subjects factor. The result showed a significant main effect of test phase \( F(6, 36) = 48.86, p < 0.001 \) and phonemic contrast \( F(3, 18) = 12.26, p < 0.001 \), and a significant interaction between the two factors \( F(18, 108) = 2.21, p < 0.01 \). Results of subsequent analyses of simple effects and analyses based on Ryan’s method are summarized as follows: (1) A training effect was found in all contrasts, though performance in the post-test significantly differed in the following manner: RL>BV<STH, (2) After a month, translation accuracy of RL and BV significantly decreased, whereas that of STH and BG remained at the level of the post-test (Figure 4). The ANOVA for (a) generally indicated results that are consistent with those for (b).

Figure 4: Translation accuracy as a function of phonemic contrast at seven test phases for the (b) group: N=8.

As for the perceptual identification, the results of the same mixed model ANOVA indicated a significant main effect of test phase \( F(1, 15) = 11.67, p < 0.01 \), though participants underwent no training on perception. Contrary to our assumption, participants did not show a statistical difference among RL, BV, and STH in the Perception task in the pretest. Therefore, we cannot discuss the relationship between perception and lexical processing by comparing these phonemes.

By comparing the result in the Translation task with that in the Perception task, it was revealed that the difficulty of perception across phonemic contrasts affected the efficiency of retention in some regards (RL and BV vs. BG), but did not in another regards (RL and BV vs. STH). Thus, the results in Experiment 4 still implies a link between phoneme perception and lexical processing, but not strongly enough to generalize the robust link found in the Experiment 2.

5. General discussion: A link between perception and lexical access in L2

From the results of the four experiments, we can conclude that the link between phoneme perception and lexical processing does exist but does not always appear robustly. As a whole, importance of accurate phoneme perception in long-term vocabulary learning was demonstrated (Experiment 1 and 3), though it was not true for the learning induced by extensive translation training (Experiment 2 and 4). This is probably because learners rely heavily on phonological information when retrieving L2 words from long-term memory. In terms of pedagogy, the robust replication of the position effect in /r/-/l/ contrast suggests that L2 learners of English might not be able to overcome lexical confusion unless they minimize phonological confusion at an early stage of learning. However, the influence of non-phonological factors on lexical processing is also important, as the influence of orthographic factors was implied in Experiment 3. Therefore, systematic investigation of the influence of various factors such as phonological, orthographic, and semantic factors, are necessary to clarify confusions in L2 mental lexicon.

6. Acknowledgments

This research was supported in part by the Telecommunications Advancement Organization of Japan.

7. References