Article

Psychological Reports
2017, Vol. 120(5) 785–804

DOI: 10.1177/0033294117707946
journals.sagepub.com/home/prx
SAGE

Examining the Temporal Development of Phonetic and Lexical Learning in Second Language

Jeong-Im Han, Moongee Jeon, and Sujin Oh
Department of English, Konkuk University, Korea

Abstract

The purpose of this study was to investigate how second-language (L2) learners lexically encode confusable phonemes. Given the inconsistency of previous studies on whether and if so how learners can establish separate lexical representations of confusable categories, we examined (1) how phonetic categorization and lexical encoding abilities were developed at the early stage of learning and (2) whether there are any differences in those abilities between the words with a sound pair from a corresponding native language (L1)-dominant category and those lacking such category. Native speakers of Korean learned Arabic words with these two types of sound pairs for four days and then their phonetic categorization and lexical processing abilities were evaluated in AXB discrimination and lexical decision tasks, respectively. The results showed that phonetic categorization of the words with a sound pair from an L1-dominant category developed very early. With success in their discrimination abilities, L2 learners began to overcome lexical competition from the words with such a sound pair. By contrast, learners showed poor sound discrimination and lexical encoding skills for words with a sound pair lacking an L1-dominant category. This suggests that (1) L2 learners' accurate phonetic categorization abilities are prerequisite to success in L2 lexical encoding and (2) lexical representations of the L2 words with confusable phonemes depend on the distinct types of sound category matchup between L1 and L2.

Keywords

Phonetic and lexical learning, Korean learners of Modern Standard Arabic, AXB discrimination task, lexical decision task

Corresponding Author:
Jeong-Im Han, Department of English, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 143-701, Korea.
Email: jhan@konkuk.ac.kr

Introduction

Recognizing words spoken in a second language (L2) can be difficult because of the activation of spurious competitor words from the native language (L1) as well as the L2. It has been shown in numerous studies that L2 learners experience parallel competition from both languages when they process spoken words in the lexicon of the target language (Broersma, 2012; Darcy, Daidone, & Kojima, 2013; Marian & Spivey, 2003; Nakai, Lindsay, & Ota, 2015; Ota, Hartsuiker, & Haywood, 2009; Pallier, Colomé, & Sebastián-Gallés, 2001; Spivey & Marian, 1999; Weber & Cutler, 2004 among others). For example, when L2 learners recognize the minimal-pair L2 words differing by an L2 contrast, as in *kettle* and *cattle*, native speakers of Dutch appear to treat them as if they were homophones. Many researchers hold the view that this lack of lexical distinction in L2 arises from L2 learners' inabilities to distinguish the relevant L2 contrasts (Broersma, 2012; Nakai et al., 2015; Ota et al., 2009; Pallier, Bosch, & Sebastián-Gallés, 1997; Pallier et al., 2001). Pallier et al. (1997), for instance, examined the perception of Catalan vowels /e/ and /ɛ/ by Spanish-dominant and Catalan-dominant bilinguals, given that these languages, as Romance languages, have very similar vowels, but Spanish lacks the /ɛ/ phoneme. When presented with a series of vowels in a vowel continuum from /e/ to /ɛ/, Spanish-dominant bilinguals have difficulties in perceiving this contrast despite being extremely fluent in Catalan. Pallier et al. (2001) further investigated the influence of these perceptual difficulties on their lexical processing. When minimal-pair words differing by these two difficult vowels (e.g., /net/ "granddaughter" vs. /nɛt/ "clean") were presented to Spanish-dominant bilinguals, they appeared to treat the minimal-pair words as if they were homophonous in spoken-word recognition. Similarly, Broersma (2002) showed that when English nonwords, which were generated by substituting a corresponding L1 vowel (e.g., /ɛ/) for a difficult vowel (e.g., /æ/) from real English words (e.g., *geng* /gɛŋ/ from *gang* /gæŋ/), were presented to listeners who were native speakers of Dutch, they appeared to be falsely identified as real English words.

Weber and Cutler (2004), however, proposed that L2 learners can have separate lexical representations for minimal pairs even when their perception of the relevant L2 phonemic contrasts is inaccurate. Based on the fact that English /ɛ/ and /æ/ are mapped onto a single Dutch category, which is phonetically closer to the English /ɛ/, they asked native English listeners and Dutch–English bilinguals to click on the picture that corresponded to what they heard from among four pictures on the computer screen which included the target (e.g., *panda* with /æ/ in the first syllable), the competitor (e.g., *pencil* with /ɛ/ in the first syllable), and two fillers (e.g., *beetle*, *bottle*). Although both the target and the competitor have confusable vowels such as /æ/ in *panda* and /ɛ/ in *pencil*, the confusability of the vowel in spoken-word recognition is asymmetric: *pan* - activated *pencil*, but *pen*- did not activate *panda*. This asymmetry was not observed in native English listeners. Weber and Cutler (2004) explained this asymmetric pattern of word activation by arguing that L2 learners have indeed encoded the /ɛ-æ/ contrast lexically, even though they have difficulties perceiving the contrast in online auditory word-identification tasks. As to the reason for such lexical asymmetry for the category /ɛ/, the researchers argued that activation of the targets containing a more similar L2 category (*pencil*) is more selective in word recognition than that of the targets with a new category (*panda*) because Dutch contains the category /ɛ/ but not /æ/. Similar findings have been shown later in Cutler, Weber, and Otake (2006) and Escudero, Hayes-Harb, and Mitterer (2008). All these studies suggest that the two categories causing difficulties are not fully merged in lexical representations and that the distinction is preserved.

Given the inconsistency of previous studies, Nakai et al. (2015) argued that L2 learners' inabilities to discriminate confusable L2 contrasts are a necessary, but not sufficient, condition for spurious activation of L2 words including such confusable phonemes. To test this hypothesis, the authors conducted an auditory word-learning experiment over three consecutive days using Greek- and Japanese-speaking learners of English whose English proficiency was native like. Given that English /s/ and /ʃ/ are matched to a single Greek phoneme /s/, as are English /b/ and /v/ to a single Japanese phoneme /b/, novel words were created by changing the onset or offset of the real English words which began with the confusable phonemes (e.g., *shentimump* from *sentiment*; *venefop* from *benefit*). When auditory lexical decision and phoneme-monitoring tasks were given to the Greek-speaking and Japanese-speaking learners of English, as well as the native English speakers, it was shown that (1) the two groups of L2 learners did not differ from the native English speakers in their ability to identify the difficult phonemes in L2 and (2) more importantly, the two L2 groups did not differ significantly from each other in the results of lexical decisions. These results suggest an absence of spurious activation of L2 words containing such minimal-pair phoneme sequences. A close inspection of the results from individual subjects revealed that there was a negative relationship between the subjects' ability to discriminate the difficult L2 contrast and the strength of spurious activation of L2 words with such contrast. Nakai et al. (2015) interpreted these results to suggest that L2 learners who can discriminate relevant difficult L2 contrasts with native-like accuracy do not show spurious activation of such contrasts during L2 lexical access.

It is, however, not clear in previous studies that argue for separate lexical representations for confusable sound categories whether L2 learners have established target-like phonological encoding in the lexical representation for each, or whether their phonological forms are lexically separated but not target-like (Darcy et al., 2012). Even in Nakai et al. (2015), it was shown that although the results of the onset-phoneme monitoring task did not significantly differ between the L1 and L2 listeners, those of L2 listeners in the lexical decision task were still significantly different from those of the native English listeners, suggesting the difficulty of achieving a target-like lexical representation.

Darcy et al. (2013) further explored this specific issue and examined the degree to which a new contrast was target-like in L2 lexical representation, based on the lexical decision and phonetic categorization patterns of American-English learners of Japanese and German. The target contrasts were long and short consonants in Japanese and two front rounded vowels in German, with neither contrast occurring in English. The results showed that learners can discriminate the confusable contrasts with high accuracy, resembling native speakers' performance, and these contrasts were encoded phonologically at the lexical level but not in a target-like manner. This is because the encoding of confusable contrasts is dependent upon the native category. For example, the American-English speakers have no geminate consonants, which led them to encode a geminate category in Japanese as a poor match to their native singleton category. One interesting result in this study was that advanced learners' lexical encoding was less dependent on their native category, though they also showed an asymmetric lexical access pattern. This suggests that L2 learners gradually encode the phonological contrasts in L2 as their general proficiency increases.

The purpose of the present study is to expand on these previous findings of asymmetries in phonetic and lexical access. Given the inconsistent results from previous studies, this study explicitly traces the path of phonetic and lexical learning in L2 learners, with a special focus on the following:

1. How are phonetic and lexical contrasts in L2 learned during the earliest stage of learning?
2. Is there any difference in the temporal development of lexical and phonetic learning according to the sound category matchup between L1 and L2?

Specifically, the first goal of the present study is to evaluate whether less efficient word processing by learners is due to inadequate perception of the target sounds in L2. Results from the previous studies are not consistent in that in some previous studies (e.g., Pallier et al., 2001), unreliable discrimination of the L2 sounds was shown to lead to complete merging of lexical entries with those sounds, whereas other studies provided evidence for L2 learners' establishment of separate lexical representations for confusable L2 sounds (e.g., Cutler et al., 2006). In the present study, participants are asked to learn aspects of a language which they have never been exposed to and then their phonetic as well as lexical learning patterns are examined. This will provide a chance to show how L2 learners establish the lexical representations of confusable sound categories in L2 from the beginning stage of the learning process. To our knowledge, no work published thus far has examined the phonetic and lexical learning during the earliest stage of learning.

The second purpose of the study is to examine whether an asymmetric pattern of lexical access is a result of lexical dominance of one member of the L2 contrast over the other. Weber and Cutler (2004) and Cutler et al. (2006) argued that the phoneme which is closer to the native category emerges as the dominant category and that the dominant category is more active during lexical access. Similarly, Darcy et al. (2013) suggested that it is more difficult to establish separate lexical representations for certain sound contrasts than for others. If a contrast cannot be mapped as a dominant versus nondominant category in a given L1, it is possible that establishing a lexical contrast in any form will be more difficult. In order to test this hypothesis, we create two types of sound pairs, where one pair includes a dominant versus nondominant category and the other pair, categories which are equally nondominant. If these two types of sound pair show different lexical access patterns, this points to a need for a more detailed analysis of the lexical representations of L2 words which show so-called lexical homophony. Namely, the temporal development of lexical and phonetic learning should be examined considering the sound category matchup between L1 and L2.

To address these issues, we investigate the phonetic categorization and lexical decision patterns of the Arabic consonants by native Korean speakers using an adapted format of the word-learning paradigm (Nakai et al., 2015; Rastle, McCormick, Bayliss, & Davis, 2011). Korean speakers learn a set of Arabic nouns in four consecutive daily sessions and then their accuracy in phonetic categorization is tested in an AXB discrimination task and their lexical processing abilities in a lexical decision task. The reason for the choice of Arabic is that (1) it is typologically distant from the Korean language and (2) Korean speakers are indeed rarely exposed to the Arabic language. Hence, this study can provide a chance to directly trace the very early stages of learning.

The target categories in Arabic are two types of consonantal phoneme pairs in Modern Standard Arabic (MSA): One phoneme pair consists of the voiceless velar and uvular stops, /k/ and /q/, with the other pair being the voiceless uvular and pharyngeal fricatives, /χ/ and /ħ/ (Mahmoud, 2013; Showalter, 2012; Watson, 2002). These two pairs of consonants can be distinctly matched up against the available Korean categories (Lee & Jongman, 2012; Shin, Kiaer, & Cha, 2013).

Given that there is a lack of acoustic and aerodynamic studies on MSA, and as far as we know, there has been no published work examining Korean learners' assimilation pattern for MSA stops, it is difficult to provide a clear picture on the matching relationship between the target stops in MSA and Korean. However, based on the results of a few previous acoustic studies of MSA and Korean, it is assumed that the velar and uvular stops in MSA, /k/ and /q/, might be perceptually mapped to a single consonant in Korean, which is a voiceless velar stop, /k/. The velar and uvular stops in MSA, /k/ and /q/, might be perceptually mapped to the velar stop, /k/, because Korean has only one stop category for velar/uvular/pharyngeal places of articulation.[1] By contrast, the two target fricatives in MSA, /χ/ and /ħ/, are new to the Korean speakers in that neither appears in Korean, and the category most similar to these fricatives might be a voiceless glottal fricative, /h/. Thus, both target sound pairs would show a two-to-one matching relationship between L2 and L1, which would lead the Korean participants to have difficulties in discriminating the two target sounds. It has been shown in many studies on L2 speech perception that it is difficult to learn to discriminate L2 phonemes when the contrasting phonemes are perceived by L2 listeners to be similar to a single phoneme in their L1 (e.g., Best, 1995; Best & Tyler, 2007). However, in the mapping of /k/ and /q/ in MSA to /k/ in Korean, MSA /k/ would be perceived as a closer match to the Korean category /k/, whereas MSA /χ/ and /ħ/ would not show any dominant category relationship with the relevant Korean category. Recall that activation of the targets containing the more dominant, similar L2 category is more selective in word recognition than that of the targets with the nondominant category, which can lead to distinct lexical representations for minimal pairs with confusable phoneme contrasts (Cutler et al., 2006; Weber & Cutler, 2004, cf. Escudero et al., 2008). We set out to test this hypothesis through the comparison of the lexical (and phonetic) learning paths for these two types of target sound pairs.

Our predictions are as follows: If the dominant-category-based argument is correct, learners would be slower over time in learning the lexical encoding of the MSA pair /χ/ and /ħ/ than the pair /k/ and /q/. When learners are exposed to the /k-q/ pair, the /k/ member, which is a closer match to the native phoneme, would emerge as the dominant category for lexical activation, and this lexical dominance of one member of the L2 contrast over the other would lead to contrastive lexical representations of the words with this contrast. On the other hand, when the MSA /χ-ħ/ pair is presented, such asymmetry in the lexical activation would not be observed because these two sounds are equally remote from the corresponding L1 phoneme. As a result, these two categories might be fully merged in the lexical representation, and learners would continue to encode words with this contrast poorly in the lexical representation.

Method

Participants

A total of 28 native speakers of Korean (15 females and 13 males) participated in this study. All participants were born, raised, and educated in Korea and had never lived outside Korea for longer than a year. None of the speakers had any familiarity with Arabic. The participants had no familiarity with the Arabic language, but rather with languages which have similar types of sounds to the target segments in MSA. All participants began to study English when they were in kindergarten or elementary schools, but mostly focused on reading. As stated in the introduction, they did not have experience living in English-speaking countries for more than a year. Based on these facts, they were assumed to be far from bilingual. In addition to English, 8 of 28 participants learned another foreign language such as Chinese ($n = 4$), Japanese ($n = 3$), and French ($n = 1$) when they were in high school as a regular subject. However, that was limited to 2 hours per week. Furthermore, Chinese and Japanese have no phonemes similar to uvular and pharyngeal fricatives in MSA, and French has a voiced uvular fricative, which is very different from the MSA target fricatives. Thus, these language experiences do not seem to influence the results of the present study. Their ages ranged from 19 to 28 (mean age of 23.1). All participants reported normal speech and hearing, normal vision or vision corrected to normal, and received a stipend for their participation in the experiment.

Stimuli

Thirty-six MSA nouns referring to objects, each of which was easily represented by a picture, were chosen as target words. All target words were monosyllabic or bisyllabic and began with target phonemes. Each set of six test words had one of the six test phonemes (/k/, /q/, /χ/, / ħ/, /m/, and /t/) as an onset, as shown in Table 1.

As described earlier, /k/ versus /q/ is a contrast with a dominant L1 category in that these two phonemes are matched to the Korean /k/, while /χ/ and / ħ/ are equally nondominant phonemes for the Korean speakers. In addition to these two types of target pairs, two additional phonemes such as /m/ and /t/ were included as controls, both of which appear in MSA and Korean as separate phonemes. All test words were associated with pictures which were black and white line drawings, most of which were taken from Snodgrass and Vanderwart (1980) and five of which were drawn on our own. The complete set of test words and their corresponding pictures are presented in Appendix 1. The materials for the AXB discrimination and lexical decision tasks were recorded by two native speakers of MSA, one female and one male. The female speaker was from Saudi Arabia (age 18), and the male speaker was from Yemen (age 23). Both speakers recorded the stimuli in a soundproof booth with a Shure KSM44 microphone and a Tascam (HD-P2) solid-state recorder. The recorded tokens were digitized in Praat at a sampling rate of 44.1 kHz and saved as 16-bit computer audio files to be used in the experiment.

Table 1. Example stimuli from Modern Standard Arabic.

Phoneme pair	Pronunciation	Meaning	Pronunciation	Meaning
/k/ vs. /q/	kursi	"Chair"	qalam	"Pen"
/χ/ vs. / ħ/	χenzer	"Pig"	ħareq	"Fire"
/m/ vs. /t/	maktab	"Desk"	telfaz	"TV"

Procedure

Each participant took part in a four-day long, consecutive experimental session, as schematized in Table 2. The participants were tested individually in a semi-soundproof booth which had inside a computer with E-prime installed.

The detailed procedures for each session were as follows: on Day 1, the participants were told that they would learn novel words from a language they had never been exposed to, which represented simple objects. They were informed that they would learn 36 words in total over the course of three sessions over three days and then be tested on their knowledge of the newly learned words on Day 2 through Day 4. First, each participant was presented two times with a randomized list of all 36 words with the corresponding pictures. When the spoken form of each word was played through headphones (Sennheiser HD 590), the corresponding picture appeared on the screen of the computer. The participants were asked to memorize the association of spoken words to pictures. The learning was self-paced in such a way that the participants proceeded from one picture to the next by clicking the space bar on a keyboard. The actual time each participant used to memorize all the new words with pictures varied, but did not exceed 10 minutes.

The first learning session (Learning 1) followed the presentation of the stimuli, which consisted of two blocks. In the first block, each participant heard the spoken form of a novel word via headphones. At the same time, two different pictures appeared at the center of the screen, one of which was associated with

Table 2. The design of the sessions over four days.

Day	Session	Task
1	1	Presentation of words with pictures
	2	Learning 1
	3	Test
2	1	Learning 2
	2	Test
	3	AXB discrimination task
	4	Lexical decision task
3	1	Learning 3
	2	Test
	3	AXB discrimination task
	4	Lexical decision task
4	1	AXB discrimination task
	2	Lexical decision task

the spoken word while the other was not. Each participant was asked to click on the target picture. If he or she thought that the picture on the left was associated with the spoken word, the "1" key on the keypad was to be pressed, but the "3" key was to be pressed if the picture on the right was associated with the spoken word. If the response was correct, the word "correct" (in English) appeared on the screen; however, if the response was incorrect, the word ""wrong" (in English) appeared on the screen. There was a 500-ms blank-screen interval between pictures and the Inter-stimulus Interval (ISI) was 3 seconds. Each target word was heard two times in a randomized order and the position of the pictures on the screen was randomized. In the second block, each participant heard the spoken form of a novel word via headphones and was asked to shadow the word. The ISI was again 3 seconds. The total number of tokens was 144 (36 words × 2 repetitions × 2 blocks) and the whole learning session of two blocks lasted approximately 15 minutes.

After training, participants took a test that consisted of 36 words to examine how well they memorized the words, and their scores were computed for accuracy of responses. Participants heard the spoken form of the test word via headphones. At the same time, four different pictures appeared at the center of the screen, one of which was associated with the spoken word while the others were not. Each participant was asked to click on the appropriate button for the target picture on the keypad such as "1", "2", "3", or "4". The position of the pictures on the screen was randomized, and the ISI was 5 seconds.

On Day 2, the participants took the second learning session (Learning 2), which started just as the first learning session except that four, not two, pictures appeared with numbers at the center of the screen. Thus, each participant was asked to click on the target picture by pressing the corresponding number on the keypad. The learning session was followed by the same memorization test with 36 words, again, as in the first learning session. After the test, participants took part in a categorical AXB discrimination task. The stimuli were presented in triads through headphones with a 500 ms ISI. Participants had 3000 ms to respond before the next trial was initiated. The number of correct responses was then calculated. In the AXB test, the first (A) and the third (B) tokens came from different phonemic categories and the listeners needed to decide whether the second token (X) belonged to the same category as "A" or "B". For a given test item (e.g., *kursi*), a nonword beginning with a confusable consonant (*qursi*) was invented and presented to test for any discrimination between the target phonemes, /k/ and /q/. Two counterbalanced orderings for the triplets were used, *kursi*-*kursi*-*qursi* (AXB) or *qursi*-*kursi*-*kursi* (BXA), which resulted in 72 test triplets (36 test words × 2 orderings). These 72 trials were presented in two randomized blocks separated by a short break. Within each trial, the first and the third tokens were spoken by the same female speaker, and the second token was spoken by a different male speaker. This was to ensure that listeners focused on relevant phonetic properties that link two tokens as members of the same category without being distracted by indexical, phonetically irrelevant variation.

After the discrimination task, a lexical decision task was given to the participants. The 36 test words were used as real-word tokens. Furthermore, 24 nonwords were created based on MSA phonotactic constraints, all of which were bisyllabic or monosyllabic words beginning with the sounds which were different from the target phonemes (e.g., *baraf*). In addition, 36 nonwords were created in terms of changing the initial consonant of the test words with a confusable sound (e.g., *kursi* "chair" $\rightarrow$ *qursi*). These words were expected to be considered real words in the early stage of learning, but nonwords when participants completely learned the target words (and/or phonemes). In total, there were 192 tokens (96 items $\times$ 2 repetitions), which were randomized. Stimuli were divided into 2 blocks, where the real word and the nonword created with the modification of the initial consonant (e.g., *kursi* and *qursi*) were not in the same block. After a short practice, participants were instructed to decide as quickly as possible whether or not each token they heard was a real word of the language they learned during the learning session by pressing the "1" key on the keypad for real words and the "3" key for nonwords. Participants had 3000 ms to respond before the next trial was initiated. Before they began the task, they were informed that the number of tokens for real words and nonwords could be different. The number of correct responses was then calculated.

The detailed procedure of Day 3 was essentially the same as that for Day 2. The only difference was the increased number of the pictures presented on the screen in the Day 3 learning session, which was eight, in order to make that test more challenging than the previous learning set. On Day 4, the participants started the session with the same AXB discrimination and lexical decision tasks as on Day 3. All procedures took place in Korea; and the directions for the detailed procedures were given in Korean.

Results

The main analysis of both the AXB discrimination and the lexical decision tasks had a two-way within-subjects analysis of variance (ANOVA) design. In this experimental design, the independent variables were Sound Pair Group (/k-q/, /χ-ħ/, and /m-t/) and Day of Learning (Day 2, Day 3, and Day 4) and the dependent variables were accuracy scores for the AXB discrimination and *d*-prime (d') values for the lexical decision. Specifically, a two-way repeated-measures ANOVA was performed on the accuracy scores and d' values as a function of Sound Pair Group and Day of Learning.

Before presenting the results of AXB discrimination and lexical decision tasks, it is important to determine whether participants learned the associations between the novel words and their corresponding pictures. Table 3 summarizes

Table 3. Mean (standard deviations) proportions of correct responses in the word-learning tests.

	Day 1	Day 2	Day 3
/k/ vs. /q/	.94 (.08)	.97 (.04)	.99 (.02)
/χ/ vs. / ħ/	.95 (.06)	.98 (.03)	.98 (.02)
/m/ vs. /t/	.97 (.04)	.99 (.02)	.99 (.02)

the mean proportions of correct responses from the tests conducted at the end of each learning session.

A two-way repeated-measures ANOVA was performed on the accuracy scores corresponding to participants' correct selection of the target pictures based on audible stimuli. There was a significant main effect of the Sound Pair Group (/k-q/, /χ- ħ/, and /m-t/) for the accuracy scores, $F(2,54) = 4.94$, $p < .05$. There was also a significant effect of Day of Learning (Day 2, Day 3, and Day 4) for the accuracy scores, $F(2,54) = 13.76$, $p < .01$. The interaction effect of the two independent variables (Sound Pair Group and Day of Learning), however, was not statistically significant, $F(4,108) = 1.04$, $p = .388$. The least significant difference (LSD) pairwise comparisons for Day of Learning indicated that the differences among the Days of Learning (i.e., Day 1 vs. Day 2, Day 1 vs. Day 3, and Day 2 vs. Day 3) were statistically significant across the three Sound Pair groups (all p's $< .01$). In other words, the scores increased from the first to the third day. The LSD pairwise comparisons for the Sound Pair Group showed that there was a significant difference between /k-q/ and /m-t/ pairs ($p < .01$) and between /χ- ħ/ and /m-t/ pairs ($p < .01$). The difference between /k-q/ and /χ- ħ/ pairs, however, was not significant ($p = .949$). The statistical results reveal that participants reached a ceiling by the end of the Day 3 learning session in terms of accuracy, regardless of the sound pairs. Based on these results, we assume that after three days of learning, the participants fully encoded the novel words, or associations of the spoken forms and the corresponding pictures. Here, we should note that this does not necessarily mean that the participants established native-like phonological representations for the test words.

Given the solid word learning of the participants, we are ready to examine their responses in the AXB discrimination and lexical decision tasks. First, Table 4 summarizes mean accuracy scores in the AXB discrimination for each sound pair across three days.

A two-way repeated-measures ANOVA was carried out on the mean accuracy scores, with Day of Learning (Day 2, Day 3, and Day 4) and Sound Pair Group (/k-q/, /χ- ħ/, and /m-t/) as within-participant factors. The main effect of Sound Pair Group, $F(2,54) = 573.74$, $p < .01$, was significant, but there was no significant effect of Day of Learning, $F(2,54) = 2.38$, $p = .102$, and no significant

Table 4. Mean (standard deviations) proportions of correct responses in the AXB discrimination tasks for each sound pair across three days.

	Day 2	Day 3	Day 4
/k/ vs. /q/	.94 (.05)	.96 (.03)	.95 (.03)
/χ/ vs. / ħ/	.64 (.09)	.64(.09)	.67 (.08)
/m/ vs. /t/	.98 (.04)	.99 (.01)	.99 (.03)

effect of interaction between Sound Pair Group and Day of Learning, $F(4,108) = 1.674$, $p = .161$. As expected, discrimination accuracy scores of the sound pair groups were significantly different from each other. The LSD pairwise comparisons showed that the proportions of discrimination accuracy scores of the three sound pairs (/k-q/, /χ- ħ/, and /m-t/) significantly differed from each other ($p < .01$) in all three sessions. On Day 2, pair /χ- ħ/ showed significantly lower accuracy scores than pair /k-q/, which in turn showed lower accuracy scores than the control pair (/m-t/). The discrimination accuracy scores of pair /k-q/ seemed to be rather close to those of the control pair, even though there was still significant difference between these two types of sound pairs. These differences among the three sound pair groups were maintained unto Day 4. The discrimination accuracy scores for pair /k-q/ were already high, up to .94 on the second day of learning, whereas pair /χ- ħ/ showed relatively low scores (.64) as compared to the other types of sound pairs on the second day of learning, and slightly increased on the fourth day (.67), which was, however, not a significant improvement. These results indicate that the sound pair whose one member is closer to the corresponding L1 category than the other is more easily discriminated than that without such dominant category.

In the lexical decision task, the accuracy scores was converted to d' values (Macmillan & Creelman, 2005). A two-way repeated-measures ANOVA was carried out on the mean d' values, with Day of Learning (Day 2, Day 3, and Day 4) and Sound Pair Group (/k-q/, /χ- ħ/, and /m-t/) as within-participant factors. The main effects of Day of Learning, $F(2,54) = 8.721$, $p < .05$, and Sound Pair Group, $F(2,54) = 71.916$, $p < .01$, were significant, but their interaction effect was not significant, $F(4,108) = 1.051$, $p = .385$. Results for the mean d' values in the lexical decision tasks conducted over three days for three sound pairs are shown in Table 5.

Table 5 shows that first, participants' d' values (from LSD pairwise comparisons) for the sound pair groups in the lexical decision tasks were significantly different from each other (/k-q/, /χ- ħ/, and /m-t/) across the three days ($p < .01$). Similar to the discrimination scores, mean d'' values for the lexical decision task for pair /χ- ħ/ showed significantly lower values (.46) than those for pair /k-q/ (1.32), which in turn showed lower values than those for the control pair (/m-t/) (3.83)

Table 5. Mean (standard deviations) d' values in the lexical decision tasks for each sound pair across three days.

	Day 2	Day 3	Day 4
/k/ vs. /q/	1.32 (.72)	2.06 (1.45)	1.92 (1.42)
/χ/ vs. / ħ/	.46 (.40)	.62 (.77)	1.25 (1.31)
/m/ vs. /t/	3.83 (2.14)	4.47 (2.27)	4.58 (2.57)

on the second day. The group difference was maintained on the fourth day (1.25 for /χ- ħ/, 1.92 for /k-q/, and 4.58 for the control pair). Particularly interesting were the values of pair /k-q/, which were close to the control pair (at a ceiling, approximately) in the discrimination task, whereas in the lexical decision task, their d' values were still different from those of the control pair on each day. As to the effect of Day of Learning, the LSD pairwise comparisons showed that d' values for the three sound pairs significantly increased from Day 2 to Day 3, but had no further increase from Day 3 to Day 4. Given the possibility that the small, but significant difference between the two target pairs (/k-q/ and /χ- ħ/) might be mitigated due to the inclusion of the control pair which was close to a ceiling, only the results of the target pairs were further analyzed statistically. The mean d' values were submitted to a two-way repeated-measures ANOVA, with Day of Learning (Day 2, Day 3, and Day 4) and Sound Pair Group (/k-q/ and /χ- ħ/) as within-participant factors. The main effects of Day of Learning, $F(2,54) = 6.640$, $p < .01$ and Sound Pair Group, $F(2,54) = 27.831$, $p < .01$, were significant, and their interaction effect was marginally significant, $F(2,54) = 2.304$, $p = .079$. The LSD pairwise comparisons showed that in pair /k-q/, d' values increased significantly from Day 2 to Day3, but no further significant increase toward Day 4 was observed. On the other hand, pair /χ- ħ/ showed no significant increase of d' values from Day 2 to Day 3, but a significant increase was observed from Day 3 to Day 4. These results indicate that participants seemed to show a difference only in one day in learning path of lexical access as well as their overall d' values for lexical access between the two target pairs, /k-q/ and /χ- ħ/. Namely, during the earliest stage of learning, participants began to show improvement of lexical processing for pair /k-q/ earlier than for pair /χ- ħ/.

As Tables 4 and 5 show, the temporal development of phonetic (AXB discrimination) and lexical (lexical decision) learning revealed a somewhat different pattern according to the specific sound pairs involved. First, regarding the control pair (/m-t/), participants did not seem to have difficulties in discriminating the two sounds that are clearly distinguished as separate phonemes in both L1 and L2 and more importantly, lexical competition from the novel words beginning with members of this pair did not seem to arise. The participants showed relatively high scores for both discrimination and lexical decision on the second day of learning and this pattern extended to the last day of learning (Day 4).

Second, the overall pattern of temporal development for pair /χ- ħ/ revealed that participants showed relatively low discrimination scores and *d′* values in the lexical decision task. The mean discrimination scores were approximately .64 on Day 2 and slightly increased during the interval between Day 2 and Day 4, suggesting that the participants had difficulties in discriminating these two sounds. At the same time, there was lexical competition among the novel words with these two sounds on Day 2. Without noticeable improvement in participants' discrimination abilities, their lexical processing began to be facilitated only slightly and their overall scores for lexical access remained relatively low even at the final stage of learning (Day 4). Finally, pair /k-q/ showed a somewhat complicated pattern: the participants learned to discriminate these two sounds from the second (or the first) day of learning and showed relatively high accuracy scores (close to a ceiling), whereas their lexical processing scores were still far lower than those of the control pair, even though they were much higher than those of pair /χ- ħ/.[2]

General discussion

The purpose of the present study was to investigate how L2 learners lexically encode confusable phonetic categories. Specifically, we investigated (1) how phonetic and lexical contrasts in L2 are learned during the earliest stage of learning and (2) whether there is any difference in the temporal development of lexical and phonetic learning between the sound pair with a corresponding L1-dominant category and one without such category. In order to address these issues, we employed a word-learning experiment where the native speakers of Korean learned MSA words beginning with two different types of sound pairs, namely, those with, versus those without, a corresponding L1-dominant category. Consequently, their phonetic categorization of the sound pairs and lexical decision patterns of the novel words with those sounds were examined.

Our results show that overall L2 learners' phonetic categorization abilities directly influence their lexical processing skills. However, and more importantly, not all sound pairs show a parallel pattern, and there is asymmetry between the sound pairs with a corresponding L1-dominant category and those without such category. The pair /k-q/ served as the case where /k/ in MSA is phonetically closer to the corresponding L1 (Korean) sound than /q/, whereas the pair /χ- ħ/ is equally remote from the corresponding L1 sound phoneme. Consequently, participants' discrimination of the pair /k-q/ was far better than that of the pair /χ- ħ/ even at the first stage of learning and this pattern was maintained during the whole learning process. With proven success in learners' discrimination abilities, they began to overcome lexical competition from the words with a member of pair /k-q/. In contrast, L2 learners' inabilities to discriminate pair /χ- ħ/ led to poor lexical processing of the novel words with members of this pair even in the final day of learning. This asymmetry in the results supports the hypothesis that if one member of the contrast in question is a dominant category, it produces a more selective lexical activation pattern than a nondominant category. This selective activation leads to the preservation of the contrast in question, not fully merging the two categories in the lexical representation. In contrast, if both members of the contrast in question are equally nondominant as in the pair /χ- ħ/, such selective lexical activation would not be established, which can hamper the lexical encoding of the words with this contrast.

Recall that Weber and Cutler (2004) found an asymmetric pattern in the recognition of novel words with confusable phonemes by Dutch-speaking learners of English such that words containing /æ/ triggered looks to pictures corresponding to words containing /æ/ and those containing /ε/, whereas words containing /ε/ only triggered looks to pictures corresponding to words containing /ε/. The researchers put forth a possible explanation for this finding such that /ε/ is phonetically closer to the L1 category, which emerges as the dominant category for auditory word recognition. At the lexical processing level, the dominant category is more accurately recognized than the nondominant category. This asymmetry leads to separate lexical representations for the novel words with the dominant and nondominant categories, even though the two sound categories are still confusable. Hence, Weber and Cutler's argument for this asymmetry is further supported by the present results with inclusion of the contrast pair which is nonexistent in L1 (i.e., pair /χ- ħ/).

The present results also have important implications for the exact path of phonetic and lexical learning in L2. It is likely that L2 learners' lexical decision scores for the novel words can be high only if they are good at discriminating the target sounds. If we look at the results for pair /k-q/, learners' discrimination accuracy scores were approximately .94 on the second day of learning, which increased slightly on the following day. Crucially, their d' values in lexical decisions increased significantly from Day 2 to Day 3, that is, 1.32 to 2.06, which decreased somewhat to 1.92 on Day 4, however. By contrast, the discrimination accuracy scores of pair /χ- ħ/ remained low, and began to show significant improvement in lexical decisions later in the learning process, between Day 3 and Day 4 (.62 to 1.25). The present results suggest that the development of phonemic categories is likely to precede accurate lexical processing. If a phonemic contrast is not accurately perceived, these difficulties will lead to inappropriate phonological representations at the lexical level. These results are in accordance with those of previous studies (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Nakai et al., 2015). Nakai et al. (2015) argued more explicitly that L2 learners' spurious activation of L2 words stems from their inabilities to discriminate L2 contrasts, and thus these discrimination inabilities are a necessary, but not sufficient condition for spurious activation of L2 words containing such phonemic contrasts. Darcy et al. (2013) also mentioned that even a high level of accuracy in phonetic categorization would not guarantee accurate lexical encoding of a difficult contrast.

There is a body of literature, however, arguing that L2 learners may establish lexical contrasts but still exhibit persistent perceptual difficulties in the discrimination of phonemic contrasts in categorization tasks (Cutler et al., 2006; Escudero et al., 2008; Weber & Cutler, 2004, among others). These studies showed that L2 learners may acquire knowledge of lexical contrasts despite persistent perceptual difficulties with enough input from the target language. This apparent discrepancy between the previous studies and the present study might be attributed to the tasks employed in the studies, namely, lexical decision versus eye-tracking. In the lexical decision task, as employed in the present study, participants needed to build up an adequate representation of the stimulus presented during the task and compare it with the information stored in their phonological lexicon, whereas in eye-tracking studies, as employed in previous studies such as those listed above, it was not clearly shown whether participants accurately encoded the target phonological contrasts. As mentioned in the introduction, it is possible that participants' phonological forms are lexically separated but not target like (Darcy et al., 2013). Namely, it was shown in eye-tracking studies by Weber and Cutler (2004) and Cutler et al. (2006) that L2 learners have established separate lexical representations for confusable sound categories, but this does not necessarily mean that their lexical encoding was as adequate as that of native speakers.

Critically, the present study showed only the earliest stage of learning for difficult sound contrasts of an unfamiliar language. With enough input from the target language, they can gradually encode the phonological contrasts in L2, with less dependence on their L1. The present results alone cannot clearly show how much input is required to show substantial improvement for the target pairs (in particular pair /χ- ħ/), and whether participants can reach a ceiling in the end. However, it has been suggested in a previous study (Darcy et al., 2013) that the lexical processing abilities for the pair without a corresponding L1-dominant category might be improved with more input. Darcy et al. (2013) found that lexical processing difficulties due to confusable sound categories can be solved with more experience in L2. In their study, the performance of intermediate- and advanced-level L2 learners was directly compared and significant differences were observed between these two groups. They interpreted these results to suggest that L2 learners were in the process of establishing a more native-like lexical access, and that their lexical representations were gradually encoding all phonological contrasts accurately. The L2 learners in the present study might also be in the process of learning to accurately encode those sound categories lexically.

In sum, examination of the phonetic and lexical learning path of the MSA words with pairs /k-q/ and /χ- ħ/ by Korean learners demonstrates that L2 learners' accurate phonetic categorization abilities are prerequisite to the success of L2 lexical encoding: L2 learners' lexical encoding begins to develop if they perform at a ceiling in the phonetic categorization. Comparison of the results between the pairs /k-q/ and /χ- ħ/ further shows that there is asymmetry in the lexical as well as phonetic learning between sound pairs with a corresponding

L1-dominant category and those without such category. While much work remains to be done to fully understand the interactions between the phonetic and lexical processing, this study has provided a clear picture for the temporal development of phonetic and lexical learning at the earliest stage.

Appendix 1

Table 6. Test words and their corresponding pictures.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A2A01011545). We thank anonymous reviewers and Cory Scherer for their helpful comments.

Notes

1. In addition, in terms of voice onset time (VOT), Modern Standard Arabic (MSA) /k/ and /q/ are both considered to be short-lag in that the VOT values of /k/ were reported to range from 20 to 40 ms in Khattab (2000) and 40 to 60 ms in Mitleb (2001), and AlShareef (2015) reported the mean VOT value of /k/ as 42.7 ms, while the mean VOT value of /q/ was reported to be approximately 25 ms (AlDahri, 2013). On the other hand, the mean VOT values of the lenis (e.g., /k/) stop were reported to be 63 ms (Kang & Guion, 2008), or 52 ms (Lee & Jongman, 2012), averaging three places of articulation (bilabial, alveolar, and velar). Thus, MSA /k/ and /q/ might be matched to the Korean /k/.
2. In order to gain more insight into the path of lexical as well as phonetic learning for each sound pair, Pearson correlation tests were further run on individual participants' discrimination scores in the AXB discrimination task and their correct responses in the lexical decision task for each sound pair. Overall, significant correlation was found between the participants' performance in discrimination and lexical decisions: $r(252) = .723$, $p < .05$. However, this result mainly derived from the results for pair /k-q/: $r(84) = .277$, $p < .05$, and the other two pairs, /χ- ħ/ and /m-t/, did not show significant correlations between these two scores, $r(84) = .103$, $p = .353$ for pair /χ- ħ/; $r(84) = .046$, $p = .678$ for pair /m-t/. These results suggest that the Korean participants who could discriminate the MSA /k-q/ contrast could reliably begin to overcome lexical competition from novel words starting with a member of this contrast. The insignificant correlation results for pair /m-t/ might have arisen from the fact that the participants already fully discriminated the target contrasts and also no lexical competition was shown in the lexical decision task. Similarly, the results of the Pearson correlation test for pair /χ- ħ/ might have to do with poor performance in both discrimination and lexical decision tasks. These results suggest that if a certain contrast is successfully discriminated, the lexical processing of the words with this contrast begins to improve in the early stage of learning, whereas if there is any contrast which is poorly discriminated, the lexical processing of the words with this contrast remains unsuccessful. However, the data included accuracy values of the two tasks that were not independent across the three days, which require better statistical techniques to put forth any strong argument. We would like to thank an anonymous reviewer for pointing out this fact.

References

AlDahri, S. S. (2013). A study for the effect of the emphaticness and language and dialect for voice onset time (VOT) in Modern Standard Arabic. arXiv preprint arXiv:1305. 2680.

AlShareef, S. (2015). *Cross-language analysis of stop closure duration and aspiration in English and Arabic* (PhD dissertation). California State University, Fresno.

Best, C. T. (1995). A direct realist view on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Philadelphia, PA: John Benjamins.

Broersma, M. (2002). Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors. In J. H. L. Hansen & B. L. Pellom (Eds.), *Seventh International Conference on Spoken Language Processing* (pp. 261–264). Denver, CO: Center for Spoken Language Research.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*, 1205–1224.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*, 269–284.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*, 372–420.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, K. C., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research*, *28*, 5–40.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*, 680–689.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*, 345–360.

Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, *124*(6), 3909–3917.

Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *Amsterdam Studies in the Theory and History of Linguistic Science Series*, *4*, 1–38.

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association*, *42*(2), 145–169.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Mahmoud, S. A. M. (2013). Discrimination of Arabic contrasts by American learners. *Studies in Second Language Learning and Teaching*, *2*, 261–291.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*, 97–115.

Mitleb, F. (2001). Voice onset time of Jordanian Arabic stops. *Journal of the Acoustical Society of America*, *109*(5), 133–135.

Nakai, S., Lindsay, S., & Ota, M. (2015). A prerequisite to L1 homophone effects in L2 spoken-word recognition. *Second Language Research*, *31*, 29–52.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, *111*, 263–269.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*, B9–B17.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native–language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, *12*, 445–449.

Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1588–1594.

Shin, J., Kiaer, J., & Cha, J. (2013). *The sounds of Korean*. UK: Cambridge University Press.

Showalter, C. E. (2012). *The influence of novel orthographic information on second language word learning: The case of native English speakers learning Arabic* (Unpublished master's thesis). The University of Utah, Provo, UT, USA.

Snodgrass, J. G., & Vanderwart, M. (1980). A Standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281–284.

Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.

Author Biographies

Jeong-Im Han is Professor in the Department of English at Konkuk University, Seoul, Korea. Her current research focuses on speech planning, spoken-word production, L2 sound acquisition, and phonetics-phonology interface.

Moongee Jeon is an associate professor in the Department of English at Konkuk University, Seoul, Korea. His main research interests include text processing, text cohesion, language psychology, experimental psychology, and cognitive science.

Sujin Oh is a PhD student in the Department of English at Konkuk University, Seoul, Korea. Her area of interest is L2 sound acquisition, psycholinguistics, and English and Korean phonetics.