Author copy

System 91 (2020) 102249

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Homophones facilitate lexical development in a second language

Jiang Liu [a,*], Seth Wiener [b]

[a] Department of Languages, Literatures and Cultures & Linguistics Program, University of South Carolina, 709 Welsh Humanities Building, 1620 College St., Columbia, SC, 29208, USA

[b] Department of Modern Languages, Carnegie Mellon University, 160 Baker Hall, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA

A R T I C L E I N F O

Article history:
Received 3 August 2019
Received in revised form 14 December 2019
Accepted 26 March 2020
Available online 30 March 2020

Keywords:
Homophones
Lexical development
Mandarin Chinese
Word learning

A B S T R A C T

This study examined how previously learned words affect the acquisition of new words in a second language (L2). Twenty L1 English adults in a second semester L2 Mandarin Chinese class performed a three-day word learning experiment involving 10 tonal monosyllabic minimal pairs (e.g., *ku4* 'warehouse' – *ku1* 'to cry'). Half of the words were homophonous with previously learned words, while the other half were not. Half of the words contained syllables with high token frequencies, while the other half contained syllables with low token frequencies. Daily training included self-paced sound-image associations presented via headphones and computer, followed by a 4-alternative-forced-choice word identification task. A naming task was included on Day 3. On Day 1, participants showed a facilitative homophone effect: new words that were homophonous with previously learned words were more accurately identified than those that were not homophonous. On Day 2 and 3, new words that were homophonous with previously learned words were recognized faster than those that were not homophonous. Syllable token frequency did not affect accuracy or response time; neither homophone status nor syllable token frequency affected naming accuracy. Thus, previously acquired phonological and lexical information can positively affect L2 lexical development.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

This study examines the role of homophones in adult second language (L2) lexical development. Homophones—words that are phonologically identical but differ in meaning, such as 'maid' and 'made'—are thought to exist in all spoken languages (Caramazza, Costa, Miozzo, & Bi, 2001) and therefore must theoretically play some role in L2 acquisition (Lagrou, Hartsuiker, & Duyck, 2011; Nakai, Lindsay, & Ota, 2015). Importantly, homophones allow for predictions regarding how familiar phonological information is mapped and accessed to different lexical and often different orthographic information. Previous homophone research has primarily focused on already acquired homophones in first language (L1) listeners (e.g., Ferrand & Grainger, 2003; Pexman & Lupker, 1999; Pexman, Lupker, & Jared, 2001) or interlingual homophones among proficient bilinguals (e.g., Lemhöfer & Dijkstra, 2004; Schulpen, Dijkstra, Schriefers, & Hasper, 2003). To the best of our knowledge, no study has systematically investigated how previously acquired words that are homophonous with new spoken words affects

* Corresponding author.
E-mail addresses: jiangliu@mailbox.sc.edu (J. Liu), sethw1@cmu.edu (S. Wiener).

https://doi.org/10.1016/j.system.2020.102249
0346-251X/© 2020 Elsevier Ltd. All rights reserved.

Author copy

beginner L2 lexical development. One reason for this gap in the literature is European languages such as English and Spanish have a relatively sparse homophone neighborhood in which there are typically only two to three homophone mates (e.g., 'none' and 'nun'; Duanmu, 2007, 2009); L2 learners of these languages do not typically encounter a large number of words that are homophonous with previously acquired words.

In contrast, Mandarin Chinese (hereafter 'Mandarin') has nearly three times the homophone density of English (Duanmu, 2007, 2009). This is partly due to the simple syllable structure in Mandarin: there are approximately 400 (C)V(C) syllable types (where C and V represent consonants and vowels) and just over 1300 unique syllable-tone[2] combinations in Mandarin (DeFrancis, 1986; Duanmu, 2007). As a result, the number of orthographically and semantically unique homophone mates for Mandarin syllable-tone words can range from 2 to 48 (Packard, 2000; Zhou & Marslen-Wilson, 1995). For example, there are six homophones for the spoken syllable-tone *dao3* /tao3/(导, 岛, 捣, 倒, 祷, 蹈). On average, each spoken syllable-tone monosyllable corresponds to about 11 homophonous Chinese characters (Tan & Perfetti, 1998). As a result of this relatively high degree of homophony, L2 learners of Mandarin are far more likely to encounter new words that are homophonous with previously learned words. In the present study, we explore how such previously learned phonological and lexical information affects the acquisition of new spoken words.

1.1. Literature review

Because orthography-to-sound mappings differ across alphabetic (e.g., English) and logographic (e.g., Mandarin) languages (Perfetti & Tan, 1998; Perfetti, Zhang, & Berent, 1992), the results from visual homophone studies are thought to be driven, in part, by the language tested (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; Ferrand & Grainger, 2003; Grainger & Ferrand, 1994; Jared, Levy, & Rayner, 1999; Pexman, Luper, & Jared, 2001; Rubenstein, Lewis, & Rubenstein, 1971; Rubenstein & Rubenstein, 1971; Sakuma, Sasanuma, Tatsumi, & Masaki, 1998). This body of research has consistently found a processing disadvantage for English homophones. For example, Rubenstein et al. (1971) used an English visual lexical decisions task to test words with and without homophones. The authors proposed that homophone activation occurs at the phonological level (e.g., /nʌn/) and all orthographic/lexical representations (e.g., 'none' and 'nun') are activated. This activates competition among possible words and delays lexical decision response times (RTs), even in biasing contexts (e.g., Chen & Boland, 2008; Onifer & Swinney, 1981; Pexman et al., 2001; Swinney, 1979).

In contrast, a processing advantage has commonly been reported for Mandarin homophones (e.g., Chen & Ding, 2005; Chen, Vaid, & Wu, 2009; Zhou & Shu, 2008; Ziegler, Tan, Perry & Monfant, 2000). Ziegler et al. found in both naming and lexical decision tasks RTs were faster for Chinese characters with homophones than for characters without homophones. The authors claimed that phonological familiarity, the increased number of homophone mates, and the lack of visual similarity between homophones (e.g., 岛 'island' and 倒 'to fall' are both pronounced as *dao3* but share no overlapping radicals[3]) contributed to the processing advantage. These findings, however, are far from conclusive as auditory-only lexical decision in Mandarin results in a processing disadvantage. For example, Wang, Li, Ning, and Zhang (2012) found that words with high homophone density led to slower RTs as a result of increased competition at the lexical level.

Thus, for English and other alphabetic languages (see Pexman et al., 2002), the homophone disadvantage is thought to stem from competition at the orthographic/lexical level in which there is much more orthographic overlap and similarity (e.g., 'made' and 'maid') but fewer homophone mates. In contrast, for Mandarin and other logographic languages (see Hino, Kusunose, Lupker, & Jared, 2013), the homophone advantage is thought to stem from high phonological familiarity, a lack of orthographic overlap and similarity, and a greater number of homophone mates. However, the particular task used to examine homophone processing (cf. Wang et al., 2012), a homophone's token frequency, the frequency of homophone mates, as well as phonology frequency and neighborhood density can all further modulate lexical access in L1 listeners (see Kerswell, Siakaluk, Pexman, Sears, & Owen, 2007; Lukatela & Turvey, 1994a, 1994b, 1999).

To what degree spoken homophones affect early Mandarin lexical development remains an open question. Limited evidence from L1 Mandarin children and L2 adult learners suggest that repeated exposure to homophonous words can affect visual word recognition and result in writing errors in which a semantically incorrect but phonologically homophonous character is used. For example, in a set of dictation studies (Yi, Shu, & Zhang, 2001; Zhang, Zhang, Zhou, & Shu, 1999), sixth grade Mandarin-speaking children were asked to write down the second morpheme of disyllabic spoken words, e.g., *kuai4* 快 'fast', for 飞快 *fei1kuai4* 'very fast.' Researchers found one third of the errors were homophone mate replacements, especially if the homophone mate had high word frequency. For example, writing *kuai4* 块 'money unit' for the second morpheme in the target disyllabic word *fei1kuai4* 飞快. Similar errors have been reported in corpus studies of L2 Mandarin learners' writings (Guo, 2011; Zhang, 2013), such as the replacement of *zuo4* 做 'to make' with *zuo4* 作 'to do' for the first morpheme in the target disyllabic word *zuo4shi4* 做事 'to work.' These results suggest that L1 Mandarin children and L2 adult learners are at least sensitive to the relatively dense homophone neighborhoods that occur in Mandarin and that this sensitivity may affect different lexical processes related to speech.

[2] The syllable is considered segmental information while the tone is considered suprasegmental information. Tone refers to the four pitch patterns that can co-occur with a syllable. In terms of word meaning, a syllable with a rising pitch can be lexically different from the same syllable with a falling pitch.

[3] Many Chinese characters are formed by combining different semantic and phonetic radicals. For example, the character 清 (pronounced as *qing1* and meaning 'clean'), is produced by combining a semantic radical 氵 meaning water and a phonetic radical 青 *qing1*.

Author copy

In the present study, we examine how new auditory words—without the interference of orthography—are acquired by adult L2 Mandarin learners. This allows us to examine whether previously learned words that match a new word's phonology (but differ in its meaning) facilitate L2 lexical development. We operationalize lexical development as learning to recognize the sound to meaning mapping (identification), accessing that sound-meaning mapping quickly (recognition response time), and producing the phonology of that mapping correctly (naming). Although previous studies have not explicitly tested L2 homophone acquisition, experimental findings suggest that prior phonological and lexical knowledge can positively affect adult L2 word learning. For example, Stamer and Vitevitch (2012) taught L1 English-L2 Spanish adults engaged in intermediate L2 classroom learning 16 novel Spanish words that varied in neighborhood density. The authors found that L2 Spanish learners better identified new Spanish words with dense neighborhoods than new Spanish words with sparse neighborhoods. The authors attributed this advantage to the partial phonological overlap that existed between known words and new representations (see also Bradlow & Pisoni, 1999; Storkel, 2001; Storkel, Armbrüster, & Hogan, 2006; Wilcox & Medina, 2013 Vitevitch, Storkel, Francisco, Evans & Goldstein, 2014). These findings suggest that a newly encountered L2 word may be learned more accurately if it is similar sounding to other previously learned words. A new L2 word that sounds nothing like previously learned words may be learned less accurately.

Novel to the present study, we test Stamer and Vitevitch's claim that partial phonological overlap between known words and novel words strengthens new representations by exploring whether this overlap is limited to segmental information only or also includes suprasegmental information in the form of lexical tone. Because beginner L2 learners have notoriously poor tone perception and production abilities (e.g., Hao, 2012, 2018; Pelzl, 2019; Wang, Spence, Jongman, & Sereno, 1999; Wiener, 2017; Wiener, Chan, & Ito, 2020), learners may be unable to differentiate crucial tone information required for minimal pair acquisition (e.g., Wong & Perrachinone, 2007). As an example, our learners are presumed to have already acquired the syllable-tone word *ku4*裤 'pants' but no *ku1* word. We therefore test how *ku4*库 'warehouse' and *ku1*哭'to cry,' (*ku1* and *ku4* only differ in tone) are affected by the presence (or absence) of previously learned homophones. To address this question, we train and test adult L1 English-L2 Mandarin learners engaged in structured elementary classroom learning. We use referent identification and picture-naming tasks involving natural Mandarin vocabulary to assess the learning outcome and lexical configuration (i.e., sound and meaning) of spoken words (e.g., Leach & Samuel, 2007; Stamer & Vitevitch, 2012; Storkel, 2001; Storkel et al., 2006; Storkel, 2004, 2009). We train and test participants for three consecutive days in order to ensure that the learned words become consolidated into each learner's lexicon and engage with known words (e.g., Dumay & Gareth Gaskell, 2007; Leach & Samuel, 2007; Qin & Zhang, 2019).

1.2. Research questions

1. Are new words that have homophone mates identified more accurately than new words without homophone mates? Does syllable token frequency interact with this learning?
2. Are new words that have homophone mates recognized faster than new words without homophone mates? Does syllable token frequency interact with this timing?
3. Are new words that have homophone mates named more accurately than new words without homophone mates? Does syllable token frequency interact with this naming?

If L2 Mandarin learners—like L2 Spanish learners (Storkel et al., 2006)—benefit from previously learned phonology, L2 Mandarin learners may show a learning effect in which new words that have homophone mates are identified more accurately than words without homophone mates (RQ1). Given the relatively limited phonology and morphology involved in Mandarin word learning, a main effect of token frequency may also occur in which words comprised of high token frequency syllables are identified more accurately than words comprised of low token frequency syllables (e.g., Wiener, Ito, & Speer, 2018; Wiener, Lee, & Tao, 2019; Wiener & Lee, 2020).

Because orthography is not involved in our experiment, we do not expect words with homophone mates to be recognized faster (RQ2) than those without homophone mates (e.g., Ziegler, Tan, Perry, & Montant, 2000). A main effect of syllable token frequency may occur such that high token frequency words are recognized faster than low token frequency words (e.g., Zhou & Marslen-Wilson, 1994, 1995).

If L2 learners perform like L1 speakers in terms of homophone naming (e.g., Ziegler et al., 2000), words with homophone mates may be named more accurately than those words without mates (RQ3). A main effect of syllable token frequency may occur (e.g., Chen et al., 2009; Chen, Chen, & Dell, 2002) such that only high token frequency homophones are named more accurately than low token frequency homophones.

2. Materials and methods

2.1. Participants

Twenty-two native English speakers (11 male; 11 female; mean age = 19; SD = 0.9; age range: 18–21) enrolled in a second semester Mandarin class at a public U.S. university took part in this study. All participants started to learn Mandarin at the university level and had less than one year of classroom learning. At the time of testing, participants had been studying for

Author copy

roughly four classroom hours a week for 19 weeks. No participant was a heritage speaker of Mandarin or any other tonal language, and no participant self-reported speaking an additional language. All participants had normal hearing and normal or corrected-to-normal vision. Participants' pure pitch perception was controlled using Tonometric (Mandell, 2018). All but two participants were able to reliably differentiate two pure tones at 15 Hz or lower (remaining mean = 7 Hz; SD = 6 Hz). Those two participants were removed from further data analysis, resulting in a total of twenty participants analyzed in the present study. All participants received class credit and a small payment for their participation. All aspects of the study were approved by the Institutional Review Board (IRB) of the first author's university. Before the experiment, all participants read the IRB letter. As all participants were enrolled in a Mandarin class at the time of the experiment, participants were explicitly informed that the decision to participate or withdraw from the experiment would not affect their grade.

2.2. Stimuli

We first built a learners' text corpus to identify possible syllable-tone homophones and non-homophones. We identified all the vocabulary introduced in the participants' textbook (*Integrated Chinese, level 1, part 1*; Liu & Yao, 2009) and found that L2 learners were exposed to 115 unique syllable types during the first semester of classroom vocabulary learning. For example, we identified that the participants had presumably learned the syllable-tone combination of *ku4* 裤 'pants' but not the syllable *ku* combined with the other three lexical tones. Thus, we identified the unlearned homophone *ku4* 库 'warehouse' as well as the unlearned non-homophone *ku1* 哭 'to cry' as target words. This design is similar to Ziegler et al. (2000) and resulted in two homophone status conditions: homophone and non-homophone. Because L2 learners had acquired a limited number of words at the time of testing, we were unable to include all tonal pairs as target homophone/non-homophone words, and thus were unable to include tone as an experimental factor. We will return to this issue in the limitations section.

Syllable token frequency was operationalized following Ziegler et al. (2000) and Wang et al. (2012). We identified all the previously learned words containing our target syllables. For example, the target homophone and non-homophone pair *ku4* and *ku1* shared the same *ku* syllable with the previously learned morpheme, *ku4* 裤 'pants.' Syllable token frequency was calculated as the cumulative frequency of all *ku4* morphemes' frequency that appear in the participants' text. Based on Ziegler et al. (2000) and Wang et al. (2012), we categorized our stimuli into high syllable token frequency if its token count was ≥8 and low syllable token frequency if its token count was <8. Thus, homophone status and syllable token frequency were crossed, resulting in 4 conditions with 5 items per condition for a total of 20 target monosyllabic words or 10 minimal pairs differing in tone. See Appendix for full stimuli.

Once the 20 target words were identified, we carried out a pretest on our participant population in order to establish that participants knew the 10 presumed homophone mates but did not know the 20 new words. Participants were shown written Chinese characters for these 30 monosyllabic words and asked to write the syllable-tone Pinyin romanization and meaning for each word. None of the 20 new words were correctly identified by the participants.[4] The 10 previously learned words' Pinyin and meaning were correctly identified by all participants. For some participants, some of the learned morphemes were not identified initially. But once provided with the words in which the morpheme appears in the text, the participants identified the morpheme.

The 20 target words were recorded in a sound-attenuated booth by a female native Mandarin speaker from Beijing. Recordings were made using Praat (Boersma & Weenink, 2018) and saved at 44 k Hz/16 bits. All stimuli were clearly enunciated so as to promote L2 speech learning (e.g., Escudero, Benders, & Wanrooij, 2011). Each of the 20 words was paired with a hand-drawn color image designed to establish the sound and meaning association (see Fig. 1; online for color version).

2.3. Procedure

Participants came to the lab for three consecutive days of training and testing. At the start of the first day of training, participants took part in the pretest vocabulary screening, a brief language background questionnaire, and the Tonometric pitch test (Mandell, 2018).

After completion of these tasks, participants began the training procedure, which followed the same order on each day. First, participants took part in a self-paced passive listening task. Participants were shown an image centered on a computer monitor while the syllable-tone audio label was simultaneously presented over headphones. Participants were instructed to memorize the image–audio pair and then mouse click on the image at their own pace to advance to the next item. Each of the 20 monosyllabic words was repeated four times in the learning session for a total of 80 pseudo-randomized trials with a 1-s intertrial interval.

After the passive listening task, participants underwent a 4-alternative-forced-choice (4AFC) word identification task. Four images were presented on the screen (Fig. 1) as participants simultaneously heard the target image's audio label over headphones. Participants were told to click on the image that matched the perceived audio as quickly and accurately as possible. The target (e.g., *ku4* 'warehouse') was presented along with a tonal competitor (e.g., *ku1* 'to cry') and two distractors, both of which differed in syllable and tone (e.g., *ta3* 'pagoda'; *chen2* 'to sink'). The positions of the target, competitors, and

[4] We acknowledge that participants may have independently learned the spoken target word outside of class but not the character. Given our participants' low proficiency and limited outside interaction with native Mandarin speakers, we consider this as a relatively minor potential confound.

Author copy

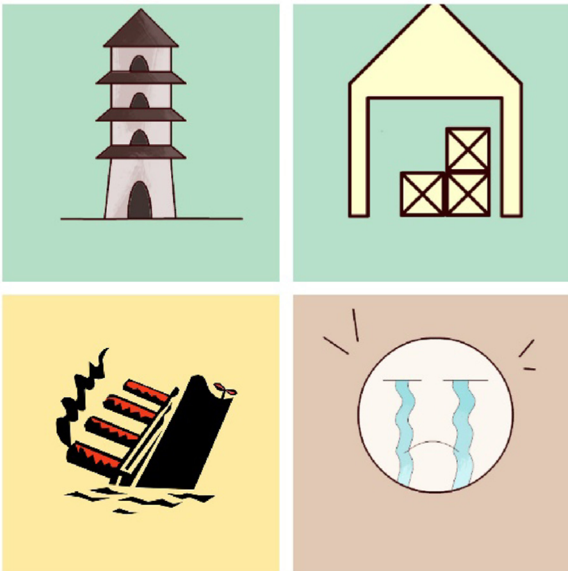

Fig. 1. Example slide from 4AFC task. Top right: target word *ku4* 'warehouse'; Bottom right: tonal competitor *ku1* 'to cry'; Bottom left and top left: distractors *chen2* 'to sink' and *ta3* 'pagoda'.

distractors were counterbalanced within each testing session and across all three days of testing. This design resulted in 60 unique slides across three days of testing (20 slides per day). On the first two days of testing, feedback was provided in the form of a red circle highlighting the target after the participant clicked. After the red circle appeared, participants were instructed to click again to proceed to the next trial with a 1-s intertrial interval. On the third day of testing, no feedback was provided. In total, the passive listening and 4AFC task took approximately 45 min. All tasks were presented on a PC using E-prime 2.0 (Psychological Software Tools Inc.).

On the third day after the 4AFC task was completed, participants took part in an additional naming task. Participants were shown a fixation for 1-s, followed by an image of one of the learned words. Participants were asked to loudly and clearly produce the image's syllable-tone into a microphone. After producing the word, participants clicked the mouse to proceed to the next trial. All 20 words were presented in a pseudo-randomized order such that the same syllable was not presented in consecutive trials. All speech productions were recorded using Praat (Boersma & Weenink, 2018) at 44 k Hz/16 bits. In total, the naming task lasted about 5 min. Participants' recordings were cleaned and trimmed and presented to 10 native Mandarin speakers from China who listened to the audio over headphones and reported the perceived syllable-tone in Pinyin.

3. Results

3.1. 4AFC mouse click accuracy

All 20 participants demonstrated well-above chance (25%) accuracy on Day 1, followed by consistent improvement across Day 2 and Day 3. Mean accuracy in terms of 95% confidence intervals were first calculated for each day: Day 1 [0.68, 0.72]; Day 2 [0.87, 0.91]; Day 3 [0.90, 0.93]. Errors were almost exclusively on the tonal competitor with distractor errors accounting for accuracy decreases of 0.03 (Day 1), 0.02 (Day 2), and less than 0.01 (Day 3).

To test whether 4AFC accuracy differed across Day and Conditions, a mixed effects logistic regression model was built with the *lme4* package in R (version 3.5.1; R Core Team, 2017) with homophone status and syllable token frequency as sum coded variables (−1, 1) and Day as a continuous variable. Model fit, including selection of predictors and their interactions, the random effect structure, and reported *p*-values were derived from the χ^2-test of the change in deviance between the models with and without the effect of interest. Table 1 reports the final model output and R code.

A main effect of Day was found indicating that accuracy improved as a function of daily training. A post-hoc comparison revealed that this effect was primarily due to the improvement between Day 1 to Day 2 ($\beta = 1.46$, $SE = 0.38$, $Z = 3.82$, $p < .001$). The improvement from Day 2 to Day 3 was much smaller and not statistically significant ($\beta = 0.70$, $SE = 0.48$, $Z = 1.51$, $p = .13$). A main effect of homophone status was also found, indicating new words with previously learned homophone mates were learned more accurately than those without homophone mates. Post-hoc comparisons revealed that this effect was primarily driven by the Day 1 results ($\beta = 0.43$, $SE = 0.19$, $Z = 2.20$, $p = .03$). On Day 2 and 3, no difference was observed between the homophone and non-homophone conditions ($ps > .05$). A null effect of syllable token frequency was found ($p = .19$). All two-way and three-way interactions were null ($ps > .05$).

Author copy

Table 1
Summary of fixed-effects coefficient estimates and significance values for outcomes on the 4AFC task. R code: glmer(Correct ~ Syllable + Homophone + Day+(1|Item)+(Day|Subject), family = "binomial").

	Estimate	Std. Error	z value	*p*
(Intercept)	−0.06	0.40	−0.15	0.87
Syllable	−0.16	0.12	−1.30	0.19
Homophone	0.27	0.12	2.22	0.03
Day	1.15	0.25	4.50	<.001

Fig. 2 (left) plots the mean 4AFC accuracy results across the three days of testing by homophone conditions. This figure summarizes the results of RQ1: participants' overall mean accuracy improved with each day of training; however, improvement primarily occurred between Day 1 and Day 2. On Day 1, participants identified new words that were homophonous with previously learned words more accurately than new words that were not homophonous with previously learned words.

3.2. Correct 4AFC mouse click response times

Incorrect mouse clicks and correct RTs faster than 500 ms or slower than 5000 ms (4% of all correct trials) were removed from further analyses and treated as outliers (see Whelan, 2008). The remaining correct mouse clicks were analyzed across each day of testing. Mean RT in terms of 95% confidence intervals were first calculated for each day (ms): Day 1 [2421, 2548]; Day 2 [2262, 2370]; Day 3 [2210, 2318]. These mean daily RTs revealed a trend in which participants correctly identified the target faster as a function of training.

To test whether RTs differed across Day and Conditions, a mixed effects linear regression model was built with homophone status and syllable frequency as sum coded variables and Day as a continuous variable. Model selection and effect structure followed the method previously outlined. Table 2 reports the model output and R code.

The RT model revealed a main effect of Condition indicating participants identified targets with homophone mates faster than targets without homophone mates. A marginal main effect of Day was found, indicating participants identified the target marginally faster with each day of training. A significant two-way interaction between syllable token frequency and homophone status was also found. To identify the locus of this two-way interaction, daily subset models were built. On Day 1, neither homophone status nor syllable token frequency, nor their interaction affected RTs ($ps > .05$). On Day 2, a main effect of homophone status was found ($\beta = -225.24$, $SE = 78.59$, $t = -2.87$, $p = .01$), indicating that new words with homophone mates were identified faster than new words without homophone mates. Neither syllable token frequency nor its interaction with homophone status affected RTs ($ps > .05$). On Day 3, again a facilitative homophone effect was found ($\beta = -204.35$, $SE = 75.92$, $t = -2.69$, $p = .02$) indicating faster RTs for new words with homophone mates. Neither syllable token frequency nor its interaction with homophone status affected RTs ($ps > .05$). Thus, the Day 2 and Day 3 main effect of homophone status contributed to the main model's overall interaction across all three days. No specific interaction was observed in the daily models.

Fig. 2 (right) summarizes the mean correct 4AFC RTs results across the three days of testing. This figure summarizes the results of RQ2: participants' mean RT became marginally faster with each day of testing. A homophone status effect emerged

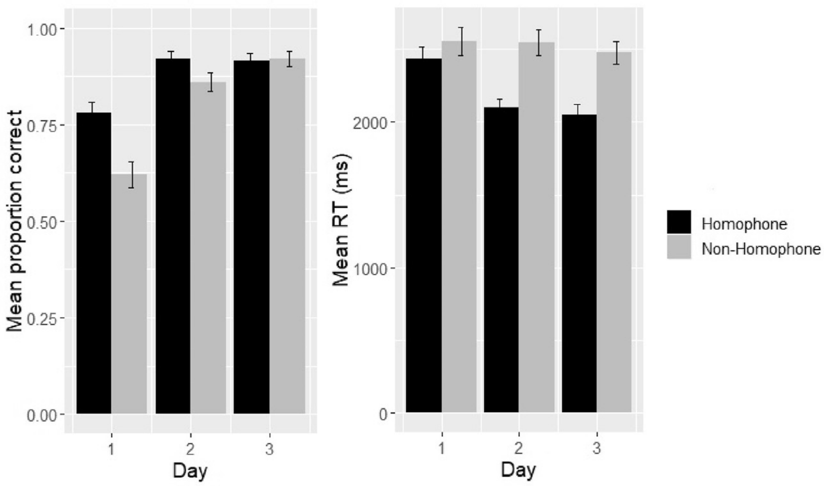

Fig. 2. Mean 4AFC accuracy (left) and mean correct response time (right) results by day and Homophone status. Error bars indicate standard error of the mean.

Author copy

Table 2
Summary of fixed-effects coefficient estimates and significance values for correct RT mouse clicks. R code: lmer(RT ~ Syllable + Homophone + Day + Syllable * Homophone + (1 | item) +(day | subject)).

	Estimate	Std.Error	t value	*p*
(Intercept)	2602.16	153.40	16.96	<.001
Syllable	29.06	44.12	0.659	0.51
Homophone	−167.30	44.15	−3.790	<.001
Day	−116.08	61.52	−1.88	0.06
Syllable:Homophone	103.59	44.14	2.34	0.02

on Day 2 and Day 3 as those new words with homophone mates were recognized faster than those without homophone mates.

3.3. Naming accuracy

A participant's production was considered correct if a native Mandarin listener's reported syllable-tone matched the target's intended syllable-tone. The ten native speakers demonstrated substantial inter-rater agreement ($\kappa = 0.63$; see Landis & Koch, 1977). Because learners' tonal productions were highly varied and not the direct focus of the present study, utterances in which only 1 or 2 of the raters disagreed with the other raters were used for analysis. That is, productions correctly or incorrectly identified by 8, 9, or 10 of the raters were analyzed. This resulted in 85% of the production data being analyzed with the remaining 15% removed due to rater disagreement among 3 or more of the native raters. On average, participants correctly named 92% of the analyzed trials; 95% confidence interval: [0.89, 0.94]. To test whether homophone status and syllable token frequency affected word naming accuracy, a mixed-effects logistic regression model was built following the previously outlined approach. R code: glmer(correct ~ Homophone * Syllable+(1|image)+(1|subject), family = "binomial"). A null effect of homophone status ($\beta = 0.03$, $SE = 0.21$, $Z = 0.17$, $p = .86$), syllable token frequency ($\beta = -0.04$, $SE = 0.21$, $Z = -0.20$, $p = .84$), and their interaction ($\beta = -0.11$, $SE = 0.21$, $Z = -0.52$, $p = .59$) were found. Fig. 3 plots the production accuracy results. This figure summarizes the results of RQ3: words with homophone mates were named as accurately as words without homophone mates.

4. Discussion

The complexity and vastness of the lexicon make systematically studying L2 lexical development prone to many confounds such as close coupling of phonological form and spelling in alphabetic languages like English. Mandarin, however, allows for the reduction of many of these confounds while at the same time providing a relatively straightforward manipulation of other phonological characteristics (see Myers, 2010, 2012 for discussions). Because Mandarin has a large number of homophones and the syllable can also represent a morpheme and word (Zhou & Marslen-Wilson, 1994, 1995), Mandarin serves as an ideal language to explore how phonological and lexical characteristics of known words affect different linguistic processes (e.g., word recognition, naming), particularly among beginner L2 learners.

In this study, we taught adult L2 Mandarin learners new spoken syllable-tone words paired with an image for three consecutive days. These words varied in their syllable token frequency and whether they were homophonous with previously acquired words. We found that overall participants learned new words with relative ease and became marginally faster at identifying the target with each consecutive day of testing. By the third day, much to our surprise, participants could recognize an average of over 80% of the items and named roughly 90% of the items correctly, (as judged by native Mandarin listeners). We present four findings from our study, which not only shed light on how homophones affect L2 lexical learning and access, but also carry pedagogical implications for L2 Mandarin word learning.

First, we found a learning advantage for new words that were homophonous with previously learned words. On the first day of testing, participants correctly identified an average of 78% of the homophone condition and only 62% of the non-homophone condition. These results suggest that prior representations were used to immediately bootstrap new word learning. For example, the target word *ku4* ('warehouse') was presumably acquired more accurately than *ku1* ('to cry') on Day 1, in part, because participants already had a *ku4* representation ('pants') in place at the time of learning.

Our interpretation of this facilitatory effect is in line with Stamer and Vitevitch's (2012) claim that if a new word has many previously learned words that sound similar to it through partial phonological overlap, the new word may be easier to learn relative to a new word with fewer neighbors and/or no phonological overlap (see also Bradlow & Pisoni, 1999; Storkel, 2001; Storkel et al., 2006; Wilcox & Medina, 2013). Novel to the present study, we extended Stamer and Vitevitch's claim by demonstrating that this overlap is not only limited to segmental information but can also include suprasegmental information in the form of lexical tone. For L2 learners of Mandarin, learning improves when new words share complete syllable-tone phonological overlap with previously learned words. This benefit, however, appears to be short lived. On Day 2 and 3, we found no accuracy difference between the homophone conditions, suggesting that once the new words were consolidated and integrated into the learners' lexicons, any phonological overlap benefits were weakened.

Author copy

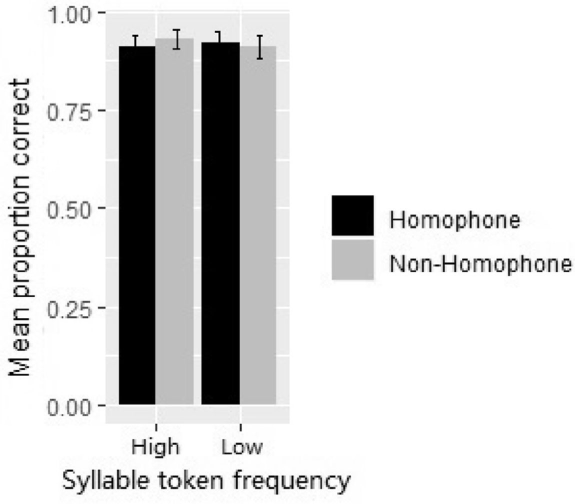

Fig. 3. Mean naming accuracy on Day 3 by Homophone status and Syllable frequency. Error bars indicate the standard error of the mean.

Second, we found a processing advantage for homophones on Day 2 and 3 in our analysis of correct response times (RTs). L2 learners identified targets that were homophonous with previously learned words faster than targets that were not homophonous with previously learned words. This finding is in line with Ziegler et al. (2000) and Chen et al. (2009) who found a facilitatory homophone effect using visual stimuli, but in conflict with Chen and Ding (2005) and Wang et al. (2012), both of whom showed that L1 Mandarin speakers exhibit an inhibitory homophone effect using auditory stimuli. There are at least two potential explanations for our pattern of results. One hypothesis is simply the L2 learners did not have enough homophones in their respective lexicons at the time of testing for an inhibitory effect or processing disadvantage to emerge. Whereas we are certain that all participants had at least one homophone for the target words in the homophone condition, it is unlikely that they had more than two or three for any given target. Thus, it may be the case that as Mandarin word learning continues and denser homophone neighborhoods emerge, L2 learners may experience a more L1-like inhibitory homophone effect during auditory processing. Future research on intermediate and advanced L2 learners is needed to evaluate this claim.

An alternative hypothesis for our RT difference is that participants were simply more certain in identifying the phonology of homophone targets than that of non-homophone targets. This is apparent from the Day 1 accuracy advantage. Moreover, because these forms were homophonous with previously learned syllable-tone combination, L2 learners had a potential bias towards certain tones when hearing a specific syllable. In other words, learners had accrued far more exemplars of particular syllable-tone combinations compared to other syllable-tone combinations. Learners were perceptually less certain about new non-homophone targets given that participants had rarely encountered words with those syllable-tone combinations. Under this usage-based account, the delay in response time for the non-homophone targets may also reflect the use of a knowledge-based tone processing mechanism (e.g., Wiener & Ito, 2016; Wiener & Turnbull, 2016) in which listeners evaluated the likelihood of the syllable-tone combination before arriving at a decision. In the current study, there were 197 errors in total among which 182 were tonal errors, namely, the participants chose the tonal competitor of the target. Post-hoc error analyses revealed that across all three days of testing, 56% of the tonal mistakes were due to participants clicking on the more probable tone for the given syllable, i.e., the word with a homophone mate rather than the word without a homophone mate. A two-proportion z-test revealed that these proportions were significantly different from one another $z = -2.31$, $p = .02$, $d = 0.12$. Future work can track eye movements to examine whether L2 learners—like L1 listeners (Wiener & Ito, 2015)—demonstrate similar probability-based predictions for targets in sparse tonal neighborhoods.

Third, we found no homophone effects in our naming data. Unlike L1 Mandarin speakers who showed an overall homophone advantage (Ziegler et al., 2000) or a limited homophone density effect in naming RTs (Chen et al., 2009), our L2 learners showed no sensitivity to either homophone status or syllable token frequency in terms of naming accuracy. Our null results are likely due to a combination of factors. First and foremost, we note that we did not examine RTs in naming. It may be that our participants showed a naming RT difference in line with Ziegler et al. (2000). Future work will need to examine RTs to evaluate this claim. Second, our participants reached an unexpected high level of accuracy. All participants correctly named over 90% target words (cf. less than 50% in Stamer & Vitevitch, 2012 by L2 Spanish learners). Given this relatively high accuracy and likely ceiling effect, we believe that on Day 3 participants had become too familiar with the 20 test items and any homophone effects had become weakened by that point in testing. Future work can explore whether a naming difference emerges on Day 1 or Day 2 before the representations become fully integrated and consolidated. We also speculate that the previously reported naming differences by L1 Mandarin speakers were likely modulated by orthographic and lexical frequency. In the current study we used sound-to-image stimuli all of which appeared with equal frequency to the L2 learners.

Author copy

Thus, neither visual frequency disparities nor orthographic information existed. Ziegler et al. (2000) found that native Mandarin speakers named morphemes without any homophones slower than morphemes with homophones. Yet, in Mandarin, morphemes without any homophones are somewhat rare and likely low in lexical frequency and high in visual complexity when compared to words with homophones (Liu et al., 1990). Thus, it may not be appropriate to use visual characters to examine word naming given the inherent frequency differences involved.

Finally, we found an overall null effect of syllable token frequency in all of our tasks. This was unexpected given the robust effect of syllable token frequency in L1 Mandarin speakers and intermediate/advanced L2 speakers (e.g., Chen et al., 2009; Chen et al., 2002; Wiener et al., 2018, 2019; Zhou & Marslen-Wilson, 1994, 1995). We speculate that this null finding may be a result of the estimated frequency counts using the learners' textbook, the relatively small stimuli set (20 items), and the unexpectedly high familiarity/accuracy participants achieved. Whereas L2 learners can show an immediate sensitivity to phonological and lexical distributional information during word recognition (e.g., Bradlow & Pisoni, 1999; Hayes-Harb, 2007; Imai, Walley, & Flege, 2005; Meador, Flege, & MacKay, 2000), these studies typically made use of a larger number of items and a wider range of distributions. We note that our syllable token calculations were far less precise than that used in previous L2 studies (e.g., Wiener et al., 2018). Our pattern of results does not necessarily rule out the possibility that our participants were sensitive to other distributional features such as lexical and orthographic frequency, or that our participants would demonstrate frequency effects when tested with other tasks that tap into the full lexicon (e.g., gating; Wiener et al., 2019).

Taken together, our results are in line with Jiang's (2000) theoretical proposal of L2 lexical development. Importantly, our finding that previous phonological overlap can help bootstrap new L2 word learning complements Jiang's argument that L2 lexical representations develop in stages with phonological (and orthographic) information acquired first prior to semantic and morphosyntactic information. We extend previous research by demonstrating that an already acquired lexeme can be used to help acquire a new lemma and thereby potentially aid or strengthen the formal stage of L2 lexical development in which initial links are relatively weak (see De Bot, Paribakht, & Wesche, 1997; Levelt, 1989). To what degree phonological overlap—or exact phonological match as in the present study—of a lexeme affects syntactic and morphological information, i.e., the mediation and integration stages, remains unclear. Future work involving measures of sentence processing is needed to tease apart how later stages of L2 lexical development occur with respect to homophones and non-homophones.

Our finding that homophones can play a positive role in L2 lexical development also has pedagogical implications. The current study did not include any orthographic training during word learning. Given the simple syllable structure in Mandarin and relatively reduced number of target words in our study, our beginner level learners acquired the new words' pronunciation and meaning quickly as shown by the high accuracy rate reached on Day 2. These results suggest that when the cognitive load associated with learning Chinese characters is removed from word learning, L2 learners may be able to form richer lexical representations. This proposal is in line with previous findings that showed orthographic information can interact with phonological representations of newly learned words and hinder acquisition (Hayes-Harb, Nicol, & Barker, 2010). Follow-up studies can replace the images with Chinese characters to test whether homophone status results in the same learning outcome when orthographic input is involved. Although the ultimate goal of L2 vocabulary learning is to establish pronunciation-orthography-meaning links, for learners acquiring an L2 that is drastically different from their L1 (e.g., L1 English-L2 Mandarin), it may be more beneficial to establish a partial linkage between two of the three levels at early stage of learning (see Barcroft, 2015). This link can gradually be extended to include Pinyin first followed by characters second. Previous research has shown a strong correlation between using Pinyin transcription and visual word recognition accuracy among L2 Mandarin learners who are native English speakers (Everson, 1998; Jiang, 2008). As a result, some L2 textbooks now prioritize the development of spoken Mandarin and use Pinyin primarily instead of Chinese characters (see Kubler, 2017; Zhao, 2011).

Finally, we note that the participants in the present study performed at a high level with respect to their accurate tone identifications and productions. This was unexpected given the literature on poor L2 tone perception and production (see Yang, 2015; Zhang, 2018), and L2 learners tendency to cite tone as a contributing factor as to why they stopped learning the language in a classroom setting (Cai, Chen, & Wang, 2010; Wu & Ortega, 2013). The present study was unable to fully investigate to what degree tonal homophones are accurately perceived and produced (e.g., Liu & Wiener, 2018). Because learners knew a small number of words, we did not compare all tone combinations as target-tonal competitors. We suspect that this limitation contributed in some way to the results. Future work will need to tease apart the role of homophony and tone more carefully. Additional limitations to our study include our relatively small stimuli set, reduced tonal competition between only two tones on screen (rather than three or all four), and the use of immediate feedback without a control condition (i.e., no feedback). To what degree these factors influenced our results remain unclear.

5. Conclusion

The present study served as an initial exploration into how spoken homophones without orthographic information affect adult L2 lexical development. We found evidence to support the claim that new words that were homophonous with previously learned words were acquired more accurately and accessed faster than new words that were not homophonous with previously learned words. Homophone status, however, did not appear to affect the naming of newly learned words. These findings suggest that L2 learners benefit from exposure to new words containing partial and full phonological overlap with previously learned words, including tonal overlap. Thus, previously learned phonological and lexical forms help bootstrap L2 lexical development.

Author copy

CRediT authorship contribution statement

Jiang Liu: Conceptualization, Methodology, Formal analysis, Visualization, Investigation, Writing - original draft. **Seth Wiener:** Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Investigation, Software, Writing - review & editing.

Appendix

Experiment stimuli

Homophone		Non-homophone	
High Syllable Token Frequency	Low Syllable Token Frequency	High Syllable Token Frequency	Low Syllable Token Frequency
ta1 塌	shu1 叔	ta3 塔	shu3 鼠
'to collapse'	'uncle'	'pagoda'	'mouse'
ren2 仁	chen4 称	ren3 忍	chen2 沉
'nuts'	'scale'	'to tolerate'	'to sink'
wan3 碗	jie3 解	wan2 丸	jie1 接
'bowl'	'to untie'	'(meat) ball'	'to pick up'
hai2 骸	ku4 库	hai3 海	ku1 哭
'skeleton'	'warehouse'	'ocean'	'to cry'
shi4 柿	hong2 虹	shi1 湿	hong1 烘
'persimmon'	'rainbow'	'wetness'	'to bake'

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2020.102249.

References

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins Publishing Company.
Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [computer program]* Version 6.0. 43. .
Bradlow, A., & Pisoni, D. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America, 106*(1), 2074–2085.
Cai, J., Chen, J., & Wang, C. (Eds.). (2010). *Teaching and learning Chinese: Issues and perspectives*. Charlotte, NC: Information Age Publishing, Inc.
Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1430.
Chen, L., & Boland, J. E. (2008). Dominance and context effects on activation of alternative homophone meanings. *Memory & Cognition, 36*(7), 1306–1323.
Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751–781.
Chen, B. G., & Ding, A. H. (2005). Homophone effects in the recognition of Chinese character: The evidence of phonology influencing the graphic processing of Chinese character. *Psychological Exploration, 25*(4), 35–39 (in Chinese).
Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language & Cognitive Processes, 24*(7–8), 967–982.
Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.
De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition*, 309–329.
DeFrancis, J. (1986). *The Chinese language: Fact and fantasy*. University of Hawaii Press.
Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). Oxford University Press.
Duanmu, S. (2009). *Syllable structure: The limits of variation*. Oxford University Press.
Dumay, N., & Gareth Gaskell, M. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35–39, 11(2), 149-175.
Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America, 130*(4), 206–212.
Everson, E. M. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing pinyin. *The Modern Language Journal, 82*(2), 194–204.
Ferrand, L., & Grainger, J. (2003). Homophone interference effects in visual word recognition. *Quarterly Journal of Experimental Psychology, 56A*, 403–419.
Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language, 33*(2), 218.
Guo, L. (2011). Analysis the learning status of homophone about foreign students. *Education and Literacy*, (2), 83–84 (in Chinese).
Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269–279.
Hao, Y. C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech, 61*, 135–152.
Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23*(1), 65–94.
Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech, 53*(3), 367–381.

Author copy

Hino, Y., Kusunose, Y., Lupker, S. J., & Jared, D. (2013). The processing advantage and disadvantage for homophones in lexical decision tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 529.
Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America, 117*(2), 896–907.
Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128*(3), 219.
Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47–77.
Jiang, X. (2008). *Research on teaching literacy and reading in CFL*. Peking University Press (in Chinese).
Kerswell, L., Siakaluk, P. D., Pexman, P. M., Sears, C. R., & Owen, W. J. (2007). Homophone effects in visual word recognition depend on homophone type and task demands. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 61*(4), 322.
Kubler, C. C. (2017). *Basic Mandarin Chinese: Speaking & listening*. Singapore: Tuttle Publishing.
Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 952.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
Leach, L., & Samuel, A. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353.
Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition, 32*(4), 533–550.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Bradford.
Liu, Y., Liang, N., Wang, D., Zhang, S., Yang, T., Jie, C., et al. (1990). *Modern Chinese word frequency dictionary for commonly used words*. Beijing: Astronautics Publishing House.
Liu, J., & Wiener, S. (2018). Perception and production of syllable-tone homophones by second language learners of Chinese. *Journal of the Acoustical Society of America, 144*(3), 1722-1722.
Liu, Y. H., & Yao, D. Z. (2009). *Integrated Chinese: Zhong wen ting Shuo Du Xie* (3rd ed.). Boston, MA: Cheng & Tsui.
Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123*(2), 107–128.
Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123*(4), 331–353.
Lukatela, G., & Turvey, M. T. (1999). Reading in two alphabets. *American Psychologist, 53*, 1057–1072.
Mandell, J. (2018). *Tonometric*. http://jakemandell.com/.
Meador, D., Flege, J. E., & MacKay, I. R. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3*(1), 55–67.
Myers, J. (2010). Chinese as a natural experiment. *The Mental Lexicon, 5*(3), 421–435.
Myers, J. (2012). Chinese as a natural experiment. In G. Libben, G. Jarema, & C. Westbury (Eds.), *Methodological and analytic frontiers in lexical research* (pp. 155–169). Amsterdam: John Benjamins.
Nakai, S., Lindsay, S., & Ota, M. (2015). A prerequisite to L1 homophone effects in L2 spoken-word recognition. *Second Language Research, 31*(1), 29–52.
Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9*(3), 225–236.
Packard, J. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.
Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a second language. The Journal of the Chinese Language Teachers Association, USA, 54*(1), 51–78.
Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in visual Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 101–118.
Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In R. Frost, & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227–248). Amsterdam: North-Holland.
Pexman, P. M., Luper, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 139–156.
Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 53*(4), 323.
Pexman, P. M., Lupker, S. J., & Reggin, L. D. (2002). Phonological effects in visual word recognition: Investigating the impact of feedback activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 572–584.
Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *Plos one, 14*(12). https://doi.org/10.1371/journal.pone.0221498.
R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10*, 645–657.
Sakuma, N., Sasanuma, S., Tatsumi, I. F., & Masaki, S. (1998). Orthography and phonology in reading Japanese kanji words: Evidence from the semantic decision task with homophones. *Memory & Cognition, 26*(1), 75–87.
Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1155.
Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism: Language and Cognition, 15*(3), 490–502.
Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337.
Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied PsychoLinguistics, 25*(2), 201–221.
Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language, 36*(2), 291–321.
Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*(6), 1175–1192.
Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 645–659.
Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing, 10*(3–5), 165–200.
Vitevitch, M. S., Storkel, H. L., Francisco, A. C., Evans, K. J., & Goldstein, R. (2014). The influence of known-word frequency on the acquisition of new neighbours in adults: Evidence for exemplar representations in word learning. *Language, Cognition and Neuroscience, 29*(10), 1311–1316.
Wang, W. N., Li, X. J., Ning, N., & Zhang, J. X. (2012). The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters, 516*, 67–71.

Author copy

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649–3658.
Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record, 58*(3), 475–482.
Wiener, S. (2017). Changes in early L2 cue-weighting of non-native speech: Evidence from learners of Mandarin Chinese. In *INTERSPEECH* (pp. 1765–1769).
Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve non-native speakers' Mandarin tone productions? *The Modern Language Journal, 104*(1), 152–168.
Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience, 30*(9), 1048–1060.
Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics, 56*, 38–51.
Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech, 61*(4), 632–656.
Wiener, S., Lee, C. Y., & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning, 69*(3), 527–558.
Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech, 59*(1), 59–82.
Wiener, S, & Lee, C. Y. (2020). Multi-Talker Speech Promotes Greater Knowledge-Based Spoken Mandarin Word Recognition in First and Second Language Listeners. *Frontiers in Psychology, 11*, 214. https://doi.org/10.3389/fpsyg.2020.00214. In this issue.
Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System, 41*(4), 1056–1069.
Wong, P., & Perrachinone, T. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied PsychoLinguistics, 28*(4), 565–585.
Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals, 46*, 680–704.
Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Berlin Heidelberg: Springer.
Yi, H., Shu, H., & Zhang, D. (2001). The characteristics and factors affecting children character output under the dictation task. *Development and Education of Psychology, 1*, 1–5 (in Chinese).
Zhang, J. H. (2013). Analysis of L2 Chinese learners' acquisition of homophonous characters and orthographically similar characters. *Applied Linguistic Research*, (6), 93–94 (in Chinese).
Zhang, H. (2018). *second language Acquisition of Chinese tones—beyond first- language transfer.* Leiden/Boston: Brill.
Zhang, D., Zhang, H., Zhou, X., & Shu, H. (1999). Production of Chinese characters under dictation. *Applied Linguistics, 29*, 67–70 (in Chinese).
Zhao, J. M. (2011). Efficient way of teaching elementary Chinese-Discussion of 'speech before text' pedagogy. *Chinese Teaching in the World, 25*(3), 367–387 (in Chinese).
Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language & Cognitive Processes, 9*(3), 393–422.
Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language & Cognitive Processes, 10*(6), 545–600.
Zhou, H., & Shu, H. (2008). The homophone family size effect and semantic transparency effect in Chinese P-O mapping. *Psychological Science, 31*, 852–855 (in Chinese).
Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science, 11*, 234–238.