2021

Eye-tracking as a window into assembled phonology in native and non-native reading

Katherine I. Martin
Southern Illinois University Carbondale, martinki@siu.edu

Alan Juffs
University of Pittsburgh, juffs@pitt.edu

Eye-tracking as a window into assembled phonology in native and non-native reading

Katherine I. Martin, Southern Illinois University Carbondale
Alan Juffs, University of Pittsburgh

Address for correspondence:
Katherine I. Martin
Department of Linguistics
Southern Illinois University
3226 Faner Hall, Mail Code 4517
1000 Faner Drive
Carbondale, IL 62901

Email: martinki@siu.edu

Eye-tracking as a window into assembled phonology in native and non-native reading

The past 30 years of reading research has confirmed the importance of bottom-up processing. Rather than a psycholinguistic guessing game (Goodman, 1967), reading is dependent on rapid, accurate recognition of written forms. In fluent first language (L1) readers, this is seen in the automatic activation of a word's phonological form, impacting lexical processing (Perfetti & Bell, 1991; Rayner, Sereno, Lesch & Pollatsek, 1995). Although the influence of phonological form is well established, less clear is the extent to which readers are sensitive to the *possible* pronunciations of a word (Lesch & Pollatsek, 1998), derived from the varying consistency of grapheme-to-phoneme correspondences (GPCs) (e.g., although 'great' has only one pronunciation, [ɡɹeɪt], the grapheme <ea> within it has multiple possible pronunciations: [i] in [plit] 'pleat', [ɛ] in [bɹɛθ] 'breath'; Parkin, 1982). Further, little is known about non-native readers' sensitivity to such characteristics. Non-native readers process text differently from L1 readers (Koda & Zehler, 2008; McBride-Chang, Bialystok, Chong & Li, 2004), with implications for understanding L2 reading comprehension (Rayner, Chace, Slattery & Ashby, 2006). The goal of this study was thus to determine whether native and non-native readers are sensitive to the consistency of a word's component GPCs during lexical processing and to compare this sensitivity among readers from different L1s.

1. L1 English Lexical Processing

English speakers initially learn to read decoding letter-by-letter, a process that automatizes over time (Ehri, 1999, 2015) to free up cognitive resources for higher-level processing (Perfetti, 2007; Perfetti & Stafura, 2014). However, bottom-up activation of word forms still occurs in fluent readers and influences lexical processing. For example, shared phonology facilitates picture naming (Alario, De Cara & Ziegler, 2007; Brooks & MacWhinney, 2000) and word recognition (Ashby, 2010; Perfetti & Bell, 1991). Phonological forms also mediate lexical access during silent reading (Folk, 1999; Rayner, 2009) and influence the time course and accuracy of semantic judgments (Luo, Johnson & Gallo, 1998; van Orden, 1987). Although much of this research examines L1 English, the findings extend to other languages (Ferrand & Grainger, 1992; Perfetti, Zhang & Berent, 1992).

Orthographic form, including whether spelling patterns are consistent/regular or inconsistent/irregular (Jared, 2002), also influences automatized lexical processing. Consistency (associated with connectionist approaches to word recognition; Plaut, McClelland, Seidenberg & Patterson, 1996; Seidenberg & McClelland, 1989) refers to whether a letter or letter sequence (grapheme) has the same pronunciation (<ine> as [aɪn] *'mine', 'pine'*) or variable pronunciations (<int> as [aɪnt] *'pint'* or [ɪnt] *'mint'*) across words. Regularity (associated with dual-route approaches; Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) refers to whether a grapheme has the typical pronunciation as defined by a set of GPC rules.[i] In naming and lexical decision, words with inconsistent GPCs typically show slower reaction times (RTs) and lower accuracy than words with consistent GPCs (Jared, McRae & Seidenberg, 1990; Ziegler, Montant & Jacobs, 1997). This pattern is particularly pronounced for low frequency words (Jared & Seidenberg, 1990; Seidenberg, Waters, Barnes & Tanenhaus, 1984; though see Newman, Jared, & Haigh, 2012), lending some support to dual-route models, in which phonological forms of high-frequency words are accessed via a direct, lexical route whereas those of low-frequency words are assembled via an indirect route (Coltheart & Rastle, 1994; Coltheart et al., 2001). Such findings also highlight the need to control frequency in research on phonological activation.

Although most studies of consistency and phonology have used single-word tasks, eye-tracking extends these results to more natural reading (Rayner et al., 1995; Sereno & Rayner, 2000). For example, Pollatsek, Lesch, Morris, and Rayner (1992) found that homophones provided a greater preview benefit for fixation times than visually-matched controls, demonstrating facilitation from shared phonology. Later studies demonstrated that both real words and non-words provide similar phonological priming, suggesting phonology is assembled even in fluent readers (Miellet & Sparrow, 2004; Wheat, Cornelissen, Frost & Hansen, 2010).

Building on this, Lesch and Pollatsek (1998) used a novel approach to examine assembled phonology during lexical processing. Participants saw pairs of words that were semantically associated (e.g., *pillow-bed*) or unrelated (e.g., *pillow-hook*) and were asked to judge their semantic relatedness. Items included pairs containing a homophone of a semantically associated word, such as *sand-beech* (homophone of 'beach'), which were expected to slow RTs and increase error rates. Critically, they had another condition testing 'false homophones': words that could be pronounced as a homophone given the range of possible English GPCs, but were not actually homophones. These included items like *pillow-bead*: <ea> can be pronounced /ɛ/ (e.g., *breath, lead, head*), making <bead> a false homophone to *bed*. In two experiments, including one monitoring eye-gaze, they found that both true homophones (*sand-beech*) and false homophones (*pillow-bead*) had similar effects on accuracy and RTs. These findings provide strong evidence that assembled phonology influences lexical processing, and that provisional phonological codes based on English GPCs impact processing even if they are not ultimately accurate, decay, or are suppressed.

More recently, Ashby and colleagues have used eye-tracking to demonstrate a range of phonological influences on reading. Ashby, Treiman, Kessler, and Rayner (2006) examined whether readers activate vowel phonology parafoveally. Using a boundary paradigm that changes the words displayed parafoveally (in preview) versus foveally (while fixated) (Rayner, 1975), they found that fixations were shorter (indicating facilitated processing) on words when the vowel pronunciation accessed during parafoveal preview was consistent with the target (fixated) word. They also found that participants used the consonant following a vowel to bias their pronunciation expectations, demonstrating that fluent readers do compute probable phonological forms during silent reading. Ashby and colleagues have also demonstrated influence of other phonological components in lexical processing, including syllables, vowel length, and voicing (Ashby & Martin, 2008; Ashby & Rayner, 2004; Ashby, Sanders & Kingston, 2009). Thus, L1 English research demonstrates that even fluent readers not only activate phonological forms during word recognition, but also compute assembled phonology.

2. L2 Lexical Processing

Growing evidence suggests that speakers from different L1s process written words, including their phonology and orthography, differently (Share, 2008). Many of these differences are related to specific L1 orthographic characteristics. For example, readers from alphabetic L1s typically have greater phonological awareness and use phonological information more during reading, whereas readers from non-alphabetic L1s typically have stronger orthographic skills and use visual orthographic information more (McBride-Chang & Ho, 2005; Nassaji & Geva, 1999; Tolchinsky, Levin, Aram & McBride-Chang, 2012). L1-specific text processing approaches also transfer and impact L2 reading (Akamatsu, 2003; Koda & Zehler, 2008). For example, Wang and colleagues (Wang & Koda, 2005; Wang, Koda & Perfetti, 2003) explored English word recognition and decoding in L1 Chinese and L1 Korean speakers using naming, semantic categorization, and spelling. Despite comparable English proficiency, the Chinese speakers relied more on orthographic processing and the Korean speakers relied more on phonological processing. Specifically, the Korean speakers had greater accuracy on phoneme deletion, more false positives when making semantic judgments on homophones, and more regularization errors in naming than the Chinese speakers, who were more sensitive to orthographic form similarity. This suggests that the Korean speakers relied more on assembled phonology whereas the Chinese speakers relied more on orthographic patterns. Wang and Geva (2003) also found that Chinese speakers had greater difficulty spelling to dictation (requiring phonological skills) than on forced-choice (visual) spelling judgments, and Hamada and Koda (2008) found that Korean speakers were faster and more accurate at pseudoword decoding than Chinese speakers.

Although eye-tracking is increasingly popular in L2 research (Godfroid, 2019; Winke, Godfroid & Gass, 2013), relatively little work has used it to examine L1 influence on L2 reading (de León Rodríguez, Buetler, Eggenberger, Laganaro, et al., 2016; Rau, Moll, Snowling & Landerl, 2015). However, some recent studies have used eye-tracking to compare readers of more transparent versus more opaque orthographies. For example, Rau et al. (2015) examined the eye-gaze of German- and English-speaking readers during word recognition. German-English cognates were used as the targets, and eye-gaze behaviors were examined as a function of the words' length and frequency (as well as the orthography of the readers: more consistent German versus more opaque English). They found that, for the same word forms, there were different patterns of length and frequency effects across the two languages, indicating substantial differences between languages in the time course of lexical processing. In another set of studies, de Leon Rodríguez and colleagues (de León Rodríguez, Buetler, Eggenberger, Laganaro, et al., 2016; de León Rodríguez, Buetler, Eggenberger, Preisig, et al., 2016) compared French and German bilinguals and found that the same readers had different eye-gaze behaviors in their two languages, and in their dominant versus non-dominant language. More specifically, readers showed evidence of more local and serial reading strategies in French (the more opaque orthography) and in their L2 (for which they had lower proficiency). Thus, these studies confirm that L1 orthography influences readers' text processing, in L1 and L2, and extend these results to eye-tracking.

Another population of interest in L2 literacy research is L1 Arabic speakers. Anecdotal evidence from English as a second language (ESL) instructors suggests that Arabic speakers have exceptional difficulties reading words in English (Thompson-Panos & Thomas-Ružić, 1983). Indeed, some studies suggest that Arabic speakers have difficulties (compared to L1 English speakers and other ESL learners) detecting letters during visual search, identifying missing letters, and recalling phonemes during immediate recall, particularly for vowels (Hayes-Harb, 2006; Kissling, 2012; Ryan & Meara, 1991). They also show relatively poor word recognition and spelling, again especially affecting vowels (Fender, 2003, 2008; Saigh & Schmitt, 2012), likely due to the influence of Arabic orthography (Fender, 2008). Although there is debate about its classification (compare Brown & Haynes, 1985; Koda, 1990 with Daniels & Bright, 1996; Share, 2008), the Arabic orthography is an abjad, in which consonants are represented with full letters, but many vowels are optionally written as diacritics and are typically omitted; this is permissible because of Arabic's consonant-based morphology (Abu-Rabia, 2002). These features combine to support the development of reading processes focused primarily on consonants. In fact, Abu-Rabia (1999) has stated that reading in Arabic "may be called 'reading consonants and guessing vowels'" (p. 95).

3. The Current Study

Although the influence of assembled phonology during fluent L1 English reading is well established, little is known about such effects in L2 readers. However, extending such work to L2 can inform our understanding of the degree to which L2 readers process text differently from L1 readers. Further, examining L1 Arabic speakers would enhance our limited understanding of their English reading. For example, it is not clear whether Arabic speakers are more similar to true alphabetic L1 speakers (because of their segmental orthography) or to other non-alphabetic L1 speakers. Finally, little work has examined L2 reading processes using a natural reading task, as opposed to processing single letters or words.

Thus, the present study investigates assembled phonological effects during lexical processing in both native and non-native English speakers. Specifically, we examined the effect of GPC consistency ('consistency effect') on lexical processing. Similar to Lesch and Pollatsek (1998), we examined whether readers are affected not by the actual phonological form of a word, but by the *possible* pronunciations that would be available via assembled phonology. To test this, we examined eye-gaze behaviors while participants read English sentences that contained a word whose vowel grapheme had just one common pronunciation ('consistent' words) or more than one common pronunciation ('inconsistent' words)[ii]. Frequency was also manipulated because of prior research documenting an interaction between frequency and consistency (Henderson, Dixon, Petersen, Twilley & Ferreira, 1995; Lee, Binder, Kim, Pollatsek & Rayner, 1999). A range of eye-gaze measures were examined to provide information about both early and late stages of lexical processing (Dussias, 2010; Rayner, 2009).

Two ESL groups (L1 Arabic and L1 Mandarin Chinese, henceforth 'Chinese') were examined to determine not only whether non-native speakers show similar assembled phonology-consistency effects as L1 English readers, but also whether the results differ across

L1s. Arabic speakers were chosen because previous research suggests they struggle with bottom-up English literacy skills, but empirical work is limited. Chinese speakers were chosen because of the evidence that they rely less on phonology for literacy. Based on previous research, we expected English speakers to show robust consistency effects. However, based on the evidence that Arabic speakers are less sensitive to vowels, we predicted that they would show smaller consistency effects than the English speakers. We similarly expected to find reduced consistency effects for the Chinese speakers compared to the English speakers. Regarding frequency, we expected stronger frequency effects in the non-native speaker groups, given previous findings of larger frequency effects in less proficient readers (Ashby, Rayner & Clifton, 2005; Hawelka, Gagl & Wimmer, 2010) or in bilinguals' non-dominant language (Duyck, Vanderelst, Desmet & Hartsuiker, 2008; Gollan et al., 2011).

4. Method

4.1 Participants

Data from 33 English speakers were analyzed; data from five others were excluded because they indicated a language other than English was spoken at home while growing up. Data from 35 Arabic speakers and 28 Chinese speakers were also analyzed. Additional data from four Arabic speakers and five Chinese speakers were excluded due to low comprehension accuracy (<70%), poor eye-tracking calibration, reporting an L1 other than Arabic or Mandarin, or contributing outliers on more than one-third of eye-tracking measures.

All Arabic and Chinese speakers were students in intermediate-level English courses at a large American university. Participants self-rated their English proficiency from 1 (no literacy/fluency) to 10 (high literacy/fluency); results are in Table 1. The Arabic and Chinese speakers had significantly lower self-rated proficiency than the English speakers, *ps*<.01.

Although the Chinese speakers rated themselves as less proficient than the Arabic speakers did (ps $<.05$), all participants were recruited from the same courses (approximately equal to B1 CEFR level), reading comprehension during the study and a post-test vocabulary measure showed no differences between the Arabic and Chinese speakers (ps $>.10$), and on a separate naming task (not analyzed here) the groups did not differ in their word reading speed or accuracy (ps $>.10$). Analyses comparing the Arabic and Chinese speakers' exposure to English (see Table 2) also showed no significant differences (ps $>.10$).[iii]

4.2 Materials

Stimuli were forty target English words,[iv] evenly divided between words with consistent vowel graphemes (e.g., <ee> in *greet*) and inconsistent vowel graphemes (e.g., <ea> in *great*). Consistency was determined from the number and frequency of pronunciation(s) of the vowel and immediately following consonant(s) (Treiman, Mullennix, Bijeljac-Babic & Richmond-Welty, 1995) across all English words and confirmed using empirically-determined consistency statistics (Chateau & Jared, 2003; Ziegler, Stone & Jacobs, 1997)[v]. Consistent and inconsistent words were matched on mean log frequency, imageability, length in letters, phonemes, and syllables; number and frequency of orthographic and phonological neighbors; and bigram sum and frequency (using E-Lexicon, Balota et al., 2007; and MRC Psycholinguistic Database, Wilson, 1988). Frequency was also manipulated; half the consistent words and half the inconsistent words were high frequency and half were low frequency. High and low frequency words differed on log frequency, $t(38)=8.55$, $p<.001$, but matched on mean concreteness, imageability, frequency of orthographic neighbors, and bigram sums and frequencies (ps $>.10$) (see Table 3).

Target words were embedded in sentences normed for naturalness; none of the raters participated in the eye-tracking study. First, 106 native English-speaking undergraduates rated each sentence on how natural it sounded, from 1 (very unnatural) to 7 (very natural). Items with a mean below 3.5 were rejected. Second, similar to Ashby and Clifton (2005), 48 native English-speaking undergraduates read each sentence with the target underlined and rated how natural that word sounded in the sentence, from 1 (very unnatural) to 7 (very natural). Items with a mean below 3.5 were rejected and remaining items were matched on mean naturalness across consistent and inconsistent conditions, $p>.15$. Targets occurred near the middle of each sentence, with no difference in the number of preceding words across conditions (ps>.35). One to two words immediately preceding and following each target were matched across sentences as much as possible. The same words were used in most cases, and the length, frequency, and part-of-speech distribution of pre-target words did not differ by consistency (ps>.30). The part-of-speech distribution of targets also did not differ by consistency ($p>.50$). Targets and their sentence frames are in the Appendix. Five additional practice sentences and 30 filler sentences were used but not analyzed.

4.3 Procedure

A tower-mounted EyeLink 1000 recorded participants' eye movements. The average eye-gaze position accuracy was .05-.25 visual degrees. Viewing was binocular but data were recorded monocularly from the pupil of the right eye at 1000 Hz. The screen resolution was 1024x768 pixels and stimuli were presented in black 20-point Times New Roman font on a white background; three characters occupied an average 1.15 degrees of visual angle.[vi] All sentences were left-justified. Participants were seated 63 cm from the monitor and chin rest and forehead rests were used to minimize head movements. The eye-tracker was calibrated for each participant before the study using a nine-point calibration and validation.

After providing informed consent, participants completed five practice trials. To begin a trial, participants looked at a left-justified calibration point and pressed a button. After a successful calibration check, a sentence appeared on the screen with its beginning at the calibration point. Participants read each sentence silently, then pushed a button to indicate they were finished. A yes/no comprehension question appeared after 20% of sentences (16 questions total), a rate consistent with recent research (e.g., Ashby, Yang, Evans & Rayner, 2012; Hawelka et al., 2010; Whitford et al., 2013). Following this, participants completed another calibration check to begin the next trial. The order of sentences was randomized for each participant. At the end, participants completed a language history questionnaire and a post-test vocabulary measure on which they self-reported whether they knew 147 of the words used in the sentences (including all targets).

4.4 Measures and Analyses

Fixations shorter than 80 ms and within .5 degrees of another fixation were combined with that fixation. Following this, fixations shorter than 80 ms or longer than 1000 ms were removed. Trials in which the participant skipped the trial or did not fixate the target were removed. Trials for which the participant indicated (on the post-test vocabulary measure) that they did not know the target word were also removed. These led to the removal of approximately 14% of the data, with similar amounts removed from each L1 group and item type.

Results are presented for seven measures, chosen to represent commonly-investigated eye-movement behaviors (Pollatsek et al., 1992) that, taken together, provide a broad picture of reading behavior (Sagarra & Seibert Hanson, 2011): skipping rate, first fixation duration, number of first-pass fixations, gaze duration, total number of fixations, target dwell time, and regressions to the target.[vii] Skipping rate indicates the average number of times participants skipped the target during the first pass of reading. First fixation duration is the length of time of only the first fixation on the target. Number of first-pass fixations counts the number of target fixations during the first read-through of a sentence. Gaze duration is the total time participants fixated on the target (first fixation plus any additional fixations) during the first read-through of the sentence. These first-pass measures index initial lexical processing, and thus earlier stages of processing (Inhoff & Radach, 1998; Rayner, 2009). Total number of fixations is a second-pass count of the number of fixations on the targets overall, both first-pass plus regressions. Target dwell time indicates the total fixation time on the targets. Regressions gives the average number of fixations participants made back to the target after reading past it. These second-pass measures index later stages of processing, including the completion of lexical access (Inhoff, 1984; Whitford & Titone, 2012), and are somewhat more global measures (Hawelka et al., 2010).

Linear mixed effects models (LMEs) were used for analyses. Reading time measures were first log-transformed to normalize the data (Baayen & Milin, 2010). Each eye-gaze measure was modeled separately to determine the effects of and interactions among L1, consistency, and frequency. When an interaction with L1 appeared, follow-up analyses examined the effects of consistency and frequency in each L1 separately. Based on Barr, Levy, Scheepers, and Tily (2013) and Jaeger (2009) the maximal random effects structure justified by the data was used for all models: random intercepts for participants and items, plus random slopes by participants for consistency, frequency, and their interaction. Models also omitted correlation parameters; these have little impact on the fixed effects and are irrelevant to the research questions (Barr et al., 2013).

5. Results

5.1 Comprehension Accuracy

The Arabic speakers scored 86.3% correct (*SD*=7.6%) on the comprehension questions and the Chinese speakers scored 85.7% correct (*SD*=7.8%); this difference was not significant, *p*=.98. The English speakers scored 90.5% correct (*SD*=5.3%), significantly higher than either non-native group, *p*s<.05. Thus, all groups read the sentences for meaning and achieved acceptably high comprehension that did not differ between the L2 groups.

5.2 Eye-Gaze Measures

Means and standard deviations for all seven eye-gaze variables, by L1 and item type, are in Table 4. The full models for each variable are in Tables 5 and 6.

For skipping rate the effect of L1 was significant, $\chi^2(df=9)=715.51$, $p<.001$. The Arabic speakers had a marginally lower skipping rate (.05) than the Chinese speakers (.07, *p*=.09), and a significantly lower skipping rate than the English speakers (.19, *p*<.001). The Chinese speakers also had a significantly lower skipping rate than the English speakers (*p*<.001). No other effects or interactions were significant.

For number of first-pass fixations the effect of L1 was significant, $\chi^2(df=8)=68.16$, $p<.001$. The Arabic speakers had more first-pass fixations (1.79) than the Chinese (1.59, *p*<.05) or English speakers (1.14, *p*<.001); the Chinese speakers also had more first-pass fixations than the English speakers (*p*<.001). Consistency was not significant but frequency was (*p*<.001), with more fixations to low frequency words (1.55) than high frequency words (1.46). An interaction between frequency and L1 (*p*<.05) for the Arabic-English comparison indicated that the Arabic speakers showed a frequency effect (a difference of .16 fixations, $\beta=-.19$, $SE=.08$, $p<.05$), whereas the English speakers did not (a difference of .04 fixations, β=-.04, *SE*=.03, *p*=.13). No other interactions were significant.

For first fixation duration the effect of L1 was significant, $\chi^2(df=8)=67.245$, $p<.001$. The Arabic speakers had longer first fixation durations (300 ms) than the Chinese (271 ms, $p<.01$) and English speakers (225 ms, $p<.001$); the Chinese speakers also had longer first fixation durations than the English speakers ($p<.001$). Neither the consistency nor frequency effects nor any interactions were significant.

For gaze duration the effect of L1 was significant, $\chi^2(df=8)=104.68$, $p<.001$. The Arabic speakers had the longest gaze durations (501 ms), followed by the Chinese speakers (416 ms) and English speakers (252 ms). All three groups differed from one another, *p*s<.01. The overall consistency effect was not significant but frequency was ($p<.01$), with longer gaze durations to low frequency (406 ms) than high frequency words (373 ms). A marginally significant two-way interaction between consistency and L1 for the Arabic-English comparison ($p=.08$) indicated that the English speakers showed a much larger consistency effect (21 ms, β=.07, *SE*=.04, $p<.05$) than the Arabic speakers (13 ms, β<.001, *SE*=.05, $p>.50$). No other interactions were significant.

For total number of fixations the effect of L1 was significant, $\chi^2(df=8)=91.19$, $p<.001$. The Arabic (3.11 fixations) and Chinese speakers (3.29 fixations) did not differ, $p>.50$, but each had significantly more fixations than the English speakers (1.42, *p*s<.001). The overall consistency effect was not significant but frequency was, $p<.01$, with more fixations to low frequency words (2.61) than high frequency words (2.45). A two-way interaction between consistency and L1 ($p<.01$) for the Arabic-Chinese comparison showed that although neither group had a significant consistency effect (Arabic, β=.02, *SE*=.22, $p>.50$; Chinese, β=.38, SE=.29, $p=.19$), the difference was numerically much larger for the Chinese speakers (.31 more fixations to inconsistent than consistent words) than the Arabic speakers (.02 fixations). An interaction between frequency and L1 (p<.001) for the Arabic-English comparison showed that the Arabic speakers had a significant frequency effect, with more fixations to low frequency than high frequency words (a difference of .43 fixations, β=-.50, *SE*=.21, p<.05), but the English speakers did not (β=.09, *SE*=.11, p=.43).

For total dwell time, the effect of L1 was significant, $\chi^2(df=8)=98.41$, p<.001. The Arabic (855 ms) and Chinese (824 ms) did not differ, p>.50, but each had longer dwell times than the English speakers (310 ms, *p*s<.001). Neither consistency nor frequency were significant, but this was qualified by an interaction between consistency and L1 for the Arabic-English comparison (p<.001): the English speakers showed a consistency effect (66 ms longer dwell times to inconsistent than consistent words, β=.50, *SE*=.19, p<.05) but the Arabic speakers did not (2 ms *shorter* dwell times to inconsistent than consistent words, β=-.03, *SE*=.11, p>.50). In addition, the three-way interaction was significant for the Arabic-Chinese comparison. Neither group showed a significant consistency effect (Arabic: β=-.02, *SE*=.10, p=.86; Chinese: β=.11, *SE*=.13, p=.42), though it was numerically much larger for the Chinese speakers (62 ms) than the Arabic speakers (-2 ms); the difference for the Chinese speakers was almost as large as in the English speakers (66 ms). In addition, the Arabic speakers showed a marginal frequency effect (dwell times 145 ms longer to low frequency than high frequency words, β=-.20, *SE*=.11, p=.09) and the Chinese speakers showed a significant frequency effect (dwell times 128 ms longer to low frequency than high frequency words, β=-.39, *SE*=.13, p<.01).

Finally, for regressions, the effect of L1 was significant, $\chi^2(df=8)=49.61$, p<.001. The Chinese speakers had more regressions (.61) than the Arabic (.40) or the English speakers (.30, *p*s<.001); the difference between the Arabic and English speakers approached significance (*p*=.11). The consistency effect was marginal, *p*=.095, with more fixations to inconsistent (.46) than consistent (.39) words, but the frequency effect was not significant, *p*=.43. However, this was qualified by a three-way interaction, with a different pattern for each L1 group. The English speakers showed a significant consistency effect, with more regressions to inconsistent (.36) than consistent (.23) words, β=.13, *SE*=.04, *p*<.01, but no frequency effect, β=.06, *SE*=.04, *p*=.18, and no consistency by frequency interaction, β=.01, *SE*=.08, *p*>.50. The Arabic speakers had a marginal consistency by frequency interaction, β=-.22, *SE*=.12, *p*=.08. High frequency words had somewhat more regressions to consistent (.39) than inconsistent (.38) words, whereas low frequency words had substantially more regressions to inconsistent (.51) than consistent (.33) words. The Chinese speakers showed no significant effects: consistency, β=.01, *SE*=.10, *p*>.50; frequency, β=-.08, *SE*=.10, *p*=.40; consistency by frequency, β=.09, *SE*=.19, *p*>.50.

5.3 Summary

Readers' eye-gaze behaviors varied substantially, and were differentially influenced by consistency and frequency, across individuals with different L1s. On first-pass measures Arabic speakers had more fixations and longer fixation durations than Chinese speakers, despite having comparable proficiency levels and comprehension. Regarding global eye-movement behavior, the Chinese speakers had more, shorter fixations, with a much greater propensity for regressions, whereas the Arabic speakers had longer (forward) fixations but relatively fewer regressions.

Frequency effects occurred across both early and late measures, though they were typically stronger for non-native speakers, especially the Arabic speakers. On the other hand, consistency effects were stronger for the English speakers. Additionally, although the Arabic-Chinese comparison was not always significant, numerically the Chinese speakers showed consistency effects of a similar magnitude to the English speakers, thus the lack of significance may be due to greater variability in the non-native data.

6. Discussion

This study used eye-tracking to investigate phonological activation during lexical processing in native and non-native speakers of English. Specifically, following Lesch and Pollatsek (1998), we examined whether readers were sensitive to the *possible* pronunciations of words' vowel graphemes, based on their spelling-pronunciation consistency. In brief, the English speakers demonstrated sensitivity to vowel GPC consistency; however, the Chinese speakers showed greatly reduced sensitivity and the Arabic speakers showed no evidence of sensitivity. The results demonstrate that non-native speakers process a text substantially differently from native speakers, and that there is also variability among L2 speaker groups.

The results are consistent with previous research demonstrating that L1 English speakers rapidly and automatically activate phonological information while reading (Folk, 1999; Inhoff & Topolski, 1994). They also corroborate research showing that readers are sensitive to GPCs during sentence-level reading (Ashby et al., 2006; Lesch & Pollatsek, 1998): despite the fact that targets in this study only contained ambiguous graphemes (rather than actually having ambiguous pronunciations), English speakers had more fixations and regressions to and longer gaze durations on inconsistent than consistent words. This provides strong evidence for the use of assembled phonology during silent reading, and is consistent both with our predictions and the literature demonstrating the prevalence and influence of phonological activation in reading (Henderson et al., 1995; Sereno & Rayner, 2000).

Another key finding from this study was the differences between the non-native speaker groups. The Chinese speakers showed more sensitivity to consistency than the Arabic speakers, particularly on total number of fixations and total dwell time. Thus, these speakers showed some evidence of assembled phonology during silent reading, though less robustly than the native speakers. In contrast, Arabic speakers' results suggest they did not assemble the various possible pronunciations of the graphemes in the targets. This is consistent with work demonstrating that Arabic speakers have difficulty processing written vowels in English (Fender, 2008; Ryan & Meara, 1991; Saigh & Schmitt, 2012). When consistency effects did appear, they were somewhat unexpectedly found on global rather than first-pass measures (where such findings are typically reported; Ashby & Clifton, 2005; Rayner, 1998). We draw two tentative conclusions from this finding. First, it suggests that non-native speakers (at this proficiency level) are less sensitive to spelling consistency than L1 English speakers, perhaps because they are less skilled in computing rapidly assembled phonological forms during reading. This conclusion is consistent with research demonstrating phonological difficulties in L2 English, particularly for learners from non-alphabetic L1s (Hayes-Harb, 2006; McBride-Chang et al., 2004; Wang & Koda, 2005). Second, the presence of consistency effects mostly on later eye-gaze measures suggests that non-native readers' phonological activation is quite delayed relative to native speakers. This pattern can be understood in the context of eye-tracking research with young readers, persons with dyslexia, and bilinguals. Typically, phonological codes begin being activated parafoveally, so foveal fixations reflect lexical processing that has been ongoing since parafoveal preview (Ashby & Rayner, 2004). However, less-skilled readers (including L2 readers) have smaller perceptual spans (Leung, Sugiura, Abe & Yoshikawa, 2014; Whitford & Titone, 2015) and are less able to take advantage of parafoveal preview to begin lexical processing (Ashby et al., 2012; Chace, Rayner & Well, 2005). Thus, the time-course of their lexical processing is delayed relative to native readers. Other work has demonstrated that eye-gaze behaviors are slowed in less-skilled readers (Hawelka et al., 2010; Rayner et al., 2010; Whitford & Titone, 2012), likely contributing further to delayed consistency effects.

The influence of L1 may help explain the differences between the Chinese and Arabic speakers. Because of its morphological and orthographic structure, Arabic readers develop lexical processing strategies focusing on consonants rather than vowels (Abu-Rabia, 1999). This likely results in a learned inattention to vowels that impacts L2 word recognition (Koda, 1989, 1990). Note that a learned inattention to vowels does not imply that Arabic speakers ignore vowels; rather, their automatic attentional processes prioritize consonants over vowels. The fact that both Arabic and English use segment-based orthographies likely encourages such L1 influence on L2 reading (Prior, 2012). In contrast, the Chinese morphosyllabary cannot support such a strategy. Because Chinese characters are visually complex, readers have no opportunity to develop learned inattention to specific components of a written form. Additionally, in Chinese, vowels (particularly their tone) provides much of the semantic content of a word, further discouraging a learned inattention to vowels. Thus, Chinese speakers' relative lack of segmental orthography experience becomes a relative advantage because they do not develop a bias to consonants over vowels.

The vowel inventories of each language may also contribute to the current results. Both Arabic and Mandarin have fewer vowels than English (Chao, 1968; Duanmu, 2006; Holes, 2004)[viii], and in some cases vowel graphemes (e.g., <ea>) may have multiple pronunciations that are not contrastive in Arabic or Mandarin. Sound discrimination abilities in non-native readers clearly influence their processing of written forms, particularly in cases where orthographic distinctions correspond to non-native phonological contrasts (Darcy et al., 2012; Ota, Hartsuiker & Haywood, 2009). It is therefore possible that these smaller vowel inventories, which may not contain the full range of pronunciations associated with inconsistent vowels, may also attenuate individuals' sensitivity to consistency. Although this issue deserves further research, a lack of L1 vowel distinctions is unlikely to explain the current results. First, a variety of inconsistent vowel graphemes with a wide range of possible pronunciations were used, making it unlikely that a few missing contrasts could drive the pattern of results. Second, the non-native speakers were matched on English reading proficiency (via comprehension, target pronunciation accuracy, and reading speed). They thus likely had similar levels of sound discrimination skill, which should have led to more consistent findings across the groups.

Two additional findings deserve attention. First, significant three-way interactions were found for total dwell time and regressions, revealing that frequency does have some influence on the consistency effect. Specifically, the Chinese speakers were more sensitive to consistency than the Arabic speakers, but only for high frequency words. This moderation of the consistency effect by frequency – in a pattern opposite of what is typically found with native readers (Jared et al., 1990; Seidenberg et al., 1984) – suggests that non-native readers must have a minimum amount of exposure to words before they are able to rapidly and automatically activate their phonological information. Second, the Chinese speakers had fewer fixations and shorter eye-gaze durations but more regressions than the Arabic speakers. Correlations revealed no evidence for a trade-off between number and durations of fixations. Although not conclusive, these differences suggest that the two groups may show different patterns in their overall eye-movement behaviors, which merits further investigation (Luegi, Costa & Faria, 2011; Roberts & Siyanova-Chanturia, 2013).

We conclude that although phonological activation may be a universal component of reading (Perfetti, 2003), the elaboration and time course of phonological activation during lexical processing may vary substantially based on the orthographic characteristics of readers' L1. Findings from this type of research lead to a better understanding of L2 reading, with direct applications to pedagogy. For example, a more fine-grained understanding of L2 reading may reveal differing challenges across L1s groups, with the potential for pedagogical interventions targeted to these L1s (Field, 2008). For example, reduced sensitivity to vowels in Arabic speakers suggests a focus on phonics and reading aloud may be particularly helpful. Such exercises may improve the quality of their lexical representations (Perfetti & Hart, 2002) by strengthening the phonological component of word knowledge. Future work examining consistency effects via eye-tracking may provide additional insight into the time course of phonological processing and the characteristics of visual span in readers with different L1 backgrounds (see also Dussias, 2010; Perfetti & Tan, 1998). Eye-tracking is an ideal methodology for this work, given its excellent temporal resolution, provision of rich data on a range of reading processes, and increasing popularity in L2 research (Winke et al., 2013). This work will continue to develop our understanding of this crucial area for successful L2 literacy outcomes.

Appendix

Target stimuli

Words with Inconsistent GPCs	Words with Consistent GPCs
binds	bakes
eases	bells
eight	belts
export	better
follow	expect
forms	feeds
fuels	films
great	germs
lands	green
latter	hiking
leads	melts
loving	mends
meats	plane
peace	sends
seals	spring
shook	stock
strong	sweep
sweat	tight
teams	yellow

Target sentences and comprehension questions (answers)

The new English teacher binds all of his ideas together in his story.

The French teacher eases his class into speaking more each day.
Does the class read more each day? (No)

The musician knows about the eight spaces where he can practice.

In business it is important to export knowledge to new countries and new companies.

In the movie the little girls follow puppets and other toys through their secret world.

The new employee saw a lot of new forms at the office that he had to fill out.

My friend thinks that the movie fuels a lot of discussion about important issues.

The girl always goes to great places when she really needs a vacation.

The pilot wearing a hat lands his airplane on the ground very smoothly.
Does the pilot land his airplane well? (Yes)

The man believes he should use that latter size envelope instead of the former one.

The nice woman leads her book group in their discussion.

The student is loving the new friends that he has made.

The boy likes restaurant meats better than his mother's cooking.

There are a lot of very fat meats which are not very healthy to eat.
Are the fat meats healthy? (No)

The government prefers a peace that is international rather than just local.
Does the government way international peace? (Yes)

The worker in the church seals the old windows so they do not leak.

Many of the boys shook some of the water out of their ears after swimming.

The man really likes strong beans because they have the best flavor.
Do strong beans taste bad? (No)

The soccer player knows that he is going to sweat a lot during the game.

There are a lot of teams that are very hard to beat.

The French teacher bakes his class some of his famous cookies.

The little child played the church bells the best he could last weekend.
Did the child play the church bells? (Yes)

There are things called fat belts which are supposed to help someone lose weight.

The man knows he should use that better size envelope to send the important letter.

The teacher begins to expect knowledge and discussion in his class by the second week.

The main character in the movie feeds a lot of children who are very hungry.

The group of students saw a lot of new films at the big festival they went to.

There are a lot of germs that are very dangerous.
Are the germs dangerous? (Yes)

The boy always goes to green fields to pick some flowers for his girlfriend.
Does the boy pick flowers for his mother? (No)

The student is hiking the new forest path with his class.

The ice cream in the restaurant melts very quickly because it is warm.

When you cook beef the fat melts off it so that it becomes healthier.

The nice woman mends her clothes when they get holes.

Travelers really prefer a plane that is large rather than the one that is small.

The boy wearing a hat sends his airplane and some other toys to his friend.

The woman really likes spring beans because they remind her of nice weather.

Those groups of boys stock some of their favorite candy in a good hiding place.

The waiter knows that he is going to sweep the floor before the customers arrive.

The painter knows about the tight spaces where he must paint carefully.
Does the painter need to be careful? (Yes)

The child has a lot of yellow puppets and dolls because she loves the color yellow.

References

Abu-Rabia, S. (1999). The effect of Arabic vowels on the reading comprehension of second- and sixth-grade native Arab children. *Journal of Psycholinguistic Research, 28*(1), 93-101. doi:10.1023/a:1023291620997

Abu-Rabia, S. (2002). Reading in a root–based–morphology language: The case of Arabic. *Journal of Research in Reading, 25*(3), 299-309. doi:10.1111/1467-9817.00177

Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning, 53*(2), 207-231. doi:10.1111/1467-9922.00216

Alario, F. X., De Cara, B., & Ziegler, J. C. (2007). Automatic activation of phonology in silent reading is parallel: Evidence from beginning and skilled readers. *Journal of Experimental Child Psychology, 97*(3), 205-219. doi:http://dx.doi.org/10.1016/j.jecp.2007.02.001

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review, 17*(1), 95-100. doi:10.3758/pbr.17.1.95

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*(3), B89-B100. doi:10.1016/j.cognition.2004.12.006

Ashby, J., Dix, H., Bontrager, M., Dey, R., & Archer, A. (2013). Phonemic awareness contributes to text reading fluency: Evidence from eye movements. *School Psychology Review, 42*(2), 157-170.

Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 224-236. doi:10.1037/0096-1523.34.1.224

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*, *19*(3), 391-426. doi:10.1080/01690960344000233

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065-1086. doi:10.1080/02724980443000476

Ashby, J., Sanders, L. D., & Kingston, J. (2009). Skilled readers begin processing sub-phonemic features by 80ms during visual word recognition: Evidence from ERPs. *Biological Psychology*, *80*(1), 84-94. doi:http://dx.doi.org/10.1016/j.biopsycho.2008.03.009

Ashby, J., Treiman, R., Kessler, B., & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 416-424. doi:10.1037/0278-7393.32.2.416

Ashby, J., Yang, J., Evans, K. H. C., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception, & Psychophysics*, *74*(4), 634-640. doi:10.3758/s13414-012-0277-0

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28. doi:http://dx.doi.org/10.21500/20112084.807

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445-459.

Barber, H. A., Otten, L. J., Kousta, S.-T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*(1), 47-53. doi:http://dx.doi.org/10.1016/j.bandl.2013.01.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:http://dx.doi.org/10.1016/j.jml.2012.11.001

Brooks, P. J., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language, 27*(2), 335-366. doi:undefined

Brown, T. L., & Haynes, M. (1985). Literacy background and reading development in a second language. In H. Carr (Ed.), *The development of reading skills* (pp. 19-34). San Francisco, CA.

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: Effects of reading skill. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(3), 209-217. doi:10.1037/h0087476

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chateau, D., & Jared, D. (2003). Spelling–sound consistency effects in disyllabic word naming. *Journal of Memory and Language, 48*(2), 255-280. doi:10.1016/S0749-596X(02)00521-1

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 589-608. doi:10.1037/0033-295X.100.4.589

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1197-1211. doi:http://dx.doi.org/10.1037/0096-1523.20.6.1197

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256. doi:10.1037/0033-295x.108.1.204

Daniels, P. T., & Bright, W. (1996). *The world's writing systems*. New York: Oxford University Press.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. (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*(1), 5-40. doi:10.1177/0267658311423455

de León Rodríguez, D., Buetler, K. A., Eggenberger, N., Laganaro, M., Nyffeler, T., Annoni, J.-M., & Müri, R. M. (2016). The impact of language opacity and proficiency on reading strategies in bilinguals: An eye movement study. *Frontiers in Psychology, 7,* 649. doi:10.3389/fpsyg.2016.00649

de León Rodríguez, D., Buetler, K. A., Eggenberger, N., Preisig, B. C., Schumacher, R., Laganaro, M., . . . Müri, R. M. (2016). The modulation of reading strategies by language opacity in early bilinguals: an eye movement study. *Bilingualism: Language and Cognition, 19*(3), 567-577. doi:10.1017/S1366728915000310

Duanmu, S. (2006). Chinese (Mandarin): Phonology. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd Edition ed., pp. 351-355): Elsevier.

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics, 30,* 149-166. doi:10.1017/S026719051000005X

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15*(4), 850-855. doi:10.3758/pbr.15.4.850

Ehri, L. C. (1999). Phases of development in learning to read words. In J. O. R. Beard (Ed.), *Reading development and the teaching of reading: A psychological perspective* (pp. 79-108). Oxford, England: Blackwell Science.

Ehri, L. C. (2015). How children learn to read words. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading* (pp. 293-310). New York: Oxford University Press.

Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics, 24*(02), 289-315. doi:10.1017/S014271640300016X

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language, 20*(1), 19-42.

Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology Section A, 45*(3), 353-372. doi:10.1080/02724989208250619

Field, J. (2008). Face to face with the ghost in the machine: Psycholinguistics and TESOL. *TESOL Quarterly, 42*(3), 361-374. doi:10.1002/j.1545-7249.2008.tb00136.x

Folk, J. R. (1999). Phonological codes are used to access the lexicon during silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 892-906. doi:10.1037/0278-7393.25.4.892

Godfroid, A. (2019). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York: Routledge.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140*(2), 186-209. doi:10.1037/a0022256

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6*(4), 126-135. doi:10.1080/19388076709556976

Grant, A., Gottardo, A., & Geva, E. (2012). Measures of reading comprehension: Do they measure different skills for children learning English as a second language? *Reading and Writing, 25*(8), 1899-1928. doi:10.1007/s11145-012-9370-y

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58*(1), 1-31. doi:10.1111/j.1467-9922.2007.00433.x

Hawelka, S., Gagl, B., & Wimmer, H. (2010). A dual-route perspective on eye movements of dyslexic readers. *Cognition, 115*(3), 367-379. doi:http://dx.doi.org/10.1016/j.cognition.2009.11.004

Hayes-Harb, R. (2006). Native speakers of Arabic and ESL texts: Evidence for the transfer of written word identification processes. *TESOL Quarterly*, *40*(2), 321-339.

Henderson, J. M., Dixon, P., Petersen, A., Twilley, L. C., & Ferreira, F. (1995). Evidence for the use of phonological representations during transsaccadic word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 82.

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*: Georgetown University Press.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior, 23*(5), 612-624. doi:https://doi.org/10.1016/S0022-5371(84)90382-7

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-54). Oxford: Elsevier Science Ltd.

Inhoff, A. W., & Topolski, R. (1994). Use of phonological codes during eye fixations in reading and in on-line and delayed naming tasks. *Journal of Memory and Language, 33*(5), 689-713. doi:http://dx.doi.org/10.1006/jmla.1994.1033

Jaeger, T. F. (2009). Random effect: Should I stay or should I go? Retrieved from https://hlplab.wordpress.com/2009/05/14/random-effect-structure/

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46*(4), 723-750. doi:http://dx.doi.org/10.1006/jmla.2001.2827

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29*(6), 687-715. doi:http://dx.doi.org/10.1016/0749-596X(90)90044-Z

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 92-105. doi:10.1037/0096-1523.16.1.92

Kissling, E. M. (2012). Cross-linguistic differences in the immediate serial recall of consonants versus vowels. *Applied Psycholinguistics, 33*(03), 605-621. doi:doi:10.1017/S014271641100049X

Koda, K. (1989). Effects of L1 orthographic representation on L2 phonological coding strategies. *Journal of Psycholinguistic Research, 18*(2), 201-222. doi:10.1007/bf01067782

Koda, K. (1990). The use of L1 reading strategies in L2 reading: Effects of L1 orthographic structures on L2 phonological recoding strategies. *Studies in Second Language Acquisition, 12*(04), 393-410. doi:10.1017/S0272263100009499

Koda, K., & Zehler, A. M. (2008). *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development*. New York: Routledge.

Lee, Y.-A., Binder, K. S., Kim, J.-O., Pollatsek, A., & Rayner, K. (1999). Activation of phonological codes during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 948-964. doi:10.1037/0096-1523.25.4.948

Lesch, M. F., & Pollatsek, A. (1998). Evidence for the use of assembled phonology in accessing the meaning of printed words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 573.

Leung, C. Y., Sugiura, M., Abe, D., & Yoshikawa, L. (2014). The perceptual span in second language reading: An eye-tracking study using a gaze-contingent moving window paradigm. *Open Journal of Modern Linguistics, 4*(05), 585.

Luegi, P., Costa, A., & Faria, I. (2011). Using eye-tracking to detect reading difficulties. *Journal of Eye Tracking, Visual Cognition and Emotion, 1*(1), 41-49.

Luo, C. R., Johnson, R. A., & Gallo, D. A. (1998). Automatic activation of phonological information in reading: Evidence from the semantic relatedness decision task. *Memory & Cognition, 26*(4), 833-843. doi:10.3758/bf03211402

McBride-Chang, C., Bialystok, E., Chong, K. K. Y., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89*(2), 93-111. doi:http://dx.doi.org/10.1016/j.jecp.2004.05.001

McBride-Chang, C., & Ho, C. S.-H. (2005). Predictors of beginning reading in Chinese and English: A 2-year longitudinal study of Chinese kindergartners. *Scientific Studies of Reading, 9*(2), 117-144. doi:10.1207/s1532799xssr0902_2

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Visual Cognition, 13*(1), 89-98. doi:10.1080/13506280500143367

Miellet, S., & Sparrow, L. (2004). Phonological codes are assembled before word fixation: Evidence from boundary paradigm in sentence reading. *Brain and Language, 90*(1), 299-310. doi:https://doi.org/10.1016/S0093-934X(03)00442-5

Nassaji, H., & Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics, 20*(02), 241-267.

Newman, R. L., Jared, D., & Haigh, C. A. (2012). Does phonology play a role when skilled readers read high-frequency words? Evidence from ERPs. *Language and Cognitive Processes, 27*(9), 1361-1384. doi:10.1080/01690965.2011.603932

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition, 111*(2), 263-269. doi:http://dx.doi.org/10.1016/j.cognition.2008.12.007

Parkin, A. (1982). Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend on how regularity is defined. *Memory & Cognition, 10*(1), 43-53. doi:10.3758/bf03197624

Pasquarella, A., Gottardo, A., & Grant, A. (2012). Comparing factors related to reading comprehension in adolescents who speak English as a first (L1) or second (L2) language. *Scientific Studies of Reading, 16*(6), 475-503. doi:10.1080/10888438.2011.593066

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*(1), 3-24. doi:10.1207/s1532799xssr0701_02

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. doi:10.1080/10888430701530730

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30*(4), 473-485. doi:http://dx.doi.org/10.1016/0749-596X(91)90017-E

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Philadelphia, PA: John Benjamins.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22-37. doi:10.1080/10888438.2013.827687

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 101-118. doi:10.1037/0278-7393.24.1.101

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.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 148.

Prior, A. (2012). Reading in more than one language: Behavior and brain perspectives. In D. Molfese, Z. Breznitz, V. W. Berninger, & O. Rubinsten (Eds.), *Listening to many voices: Reading, writing, mathematics and the brain* (pp. 131-156). Dordrecht: Springer.

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology, 130*, 92-105. doi:https://doi.org/10.1016/j.jecp.2014.09.012

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*(1), 65-81. doi:https://doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506. doi:10.1080/17470210902816461

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*(3), 241-255. doi:10.1207/s1532799xssr1003_3

Rayner, K., Sereno, S. C., Lesch, M. F., & Pollatsek, A. (1995). Phonological codes are automatically activated during reading: Evidence from an eye movement priming paradigm. *Psychological Science, 6*(1), 26-32. doi:10.1111/j.1467-9280.1995.tb00300.x

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review, 17*(6), 834-839. doi:10.3758/pbr.17.6.834

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*(02), 213-235. doi:doi:10.1017/S0272263112000861

Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language, 7*(2), 531-540.

Sagarra, N., & Seibert Hanson, A. (2011). Eyetracking methodology: A user's guide for linguistic research *Studies in Hispanic and Lusophone Linguistics* (Vol. 4, pp. 543).

Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System, 40*(1), 24-36. doi:10.1016/j.system.2012.01.005

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568. doi:10.1037/0033-295X.96.4.523

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*(3), 383-404. doi:10.1016/s0022-5371(84)90270-6

Sereno, S. C., & Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Perception & Psychophysics, 62*(2), 402-409. doi:10.3758/bf03205559

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*(4), 584-615. doi:10.1037/0033-2909.134.4.584

Slattery, T. J., & Rayner, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology, 24*(8), 1129-1148. doi:doi:10.1002/acp.1623

Thompson-Panos, K., & Thomas-Ružić, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly, 17*(4), 609-623.

Tolchinsky, L., Levin, I., Aram, D., & McBride-Chang, C. (2012). Building literacy in alphabetic, abjad and morphosyllabic systems. *Reading and Writing, 25*(7), 1573-1598. doi:10.1007/s11145-011-9334-7

Tolentino, L. C., & Tokowicz, N. (2009). Are pumpkins better than heaven? An ERP investigation of order effects in the concrete-word advantage. *Brain and Language, 110*(1), 12-22.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*(2), 107-136. doi:10.1037/0096-3445.124.2.107

Tuninetti, A., Warren, T., & Tokowicz, N. (2015). Cue strength in second-language processing: An eye-tracking study. *The Quarterly Journal of Experimental Psychology, 68*(3), 568-584. doi:10.1080/17470218.2014.961934

van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*(3), 181-198. doi:10.3758/bf03197716

Wang, M., & Geva, E. (2003). Spelling performance of Chinese children using English as a second language: Lexical and visual–orthographic processes. *Applied Psycholinguistics, 24*(01), 1-25. doi:10.1017/S0142716403000018

Wang, M., & Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning, 55*(1), 71-98. doi:10.1111/j.0023-8333.2005.00290.x

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition, 87*(2), 129-149. doi:10.1016/s0010-0277(02)00232-9

Wheat, K. L., Cornelissen, P. L., Frost, S. J., & Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: Evidence from magnetoencephalography. *The Journal of Neuroscience, 30*(15), 5229-5233. doi:10.1523/jneurosci.4448-09.2010

Whitford, V., O'driscoll, G. A., Pack, C. C., Joober, R., Malla, A., & Titone, D. (2013). Reading impairments in schizophrenia relate to individual differences in phonological processing and oculomotor control: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: General, 142*(1), 57.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review, 19*(1), 73-80. doi:10.3758/s13423-011-0179-5

Whitford, V., & Titone, D. (2015). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1118-1129. doi:10.1037/xlm0000093

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, Version 2. *Behavior Research Methods, 20*(1), 6-11.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue: Eye-movement recordings in second language research. *Studies in Second Language Acquisition, 35*(02), 205-212. doi:doi:10.1017/S027226311200085X

Wolf, M. K., Farnsworth, T., & Herman, J. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment, 13*(2-3), 80-107. doi:10.1080/10627190802394222

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3-29. doi:10.1037/0033-2909.131.1.3

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*(4), 533-554. doi:http://dx.doi.org/10.1006/jmla.1997.2525

Ziegler, J. C., Stone, G., & Jacobs, A. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers, 29*(4), 600-618. doi:10.3758/bf03210615

Table 1. *Participants' English language skills*.

Language skill	English (*n*=33)	Arabic (*n*=35)	Chinese (*n*=28)
Reading[a]	9.48 (0.80)	6.14 (1.38)	5.29 (1.46)
Writing[a]	9.39 (0.93)	5.66 (1.51)	4.57 (1.64)
Conversational fluency[a]	9.67 (0.69)	6.57 (1.86)	4.96 (1.62)
Spoken language comprehension[a]	9.79 (0.48)	6.74 (1.56)	4.79 (1.71)
Vocabulary post-test score[b]	146.99 (0.17)	136.79 (7.32)	136.43 (9.65)
Comprehension score[c]	90.48 (5.26)	86.33 (7.63)	85.71 (7.20)

Note. Standard deviations are in parentheses.
[a]Scores are for self-rated ability. Scale is from 1 (no literacy/fluency) to 10 (full literacy/fluency). [b]Maximum score is 147. [c]Percent correct.

Table 2. *Participant characteristics*.

Participant characteristic	Arabic (*n*=35)	Chinese (*n*=28)
Age (years)	24.63 (3.20)	25.21 (5.40)
Age when first began learning English (years)	13.12 (4.47)	12.44 (2.00)
Time spent studying English (years)	9.49 (5.30)	9.13 (3.70)
Time in the U.S. (months)	8.40 (4.68)	6.14 (6.25)

Note. Standard deviations are in parentheses.

Table 3. *Stimulus characteristics*.

	Inconsistent		Consistent	
	High frequency	Low frequency	High frequency	Low frequency
Log frequency	10.44 (.92)	7.41 (1.20)	10.57 (.95)	6.92 (1.71)
Length	5.40 (.52)	5.10 (.32)	5.44 (.53)	5.09 (.30)
Concreteness	400.29 (117.72)	506.00 (123.20)	514.17 (63.26)	470.17 (97.21)
Imageability	454.88 (100.37)	524.00 (93.60)	525.57 (117.06)	468.13 (76.48)
Number of orthographic neighbors	6.60 (3.37)	8.10 (3.45)	5.56 (2.79)	9.91 (4.70)
Frequency of orthographic neighbors	6.69 (1.11)	7.97 (1.30)	7.83 (1.99)	7.26 (1.25)
Bigram sum	16546.20 (10267.56)	14588.20 (7666.89)	16801.89 (9037.13)	13299.73 (6395.01)
Bigram mean	3684.15 (2025.53)	3502.90 (1655.62)	3741.58 (1784.78)	3260.53 (1288.80)

Note. Standard deviations are in parentheses.

Table 4. *Eye-gaze measures by L1 group and stimulus type.*

	English				Arabic				Chinese			
	High Frequency		Low Frequency		High Frequency		Low Frequency		High Frequency		Low Frequency	
	Cons	Incons	Cons	Incons	Cons	Incons	Cons	Incons	Cons	Incons	Cons	Incons
Skipping rate	.22 (.42)	.19 (.39)	.19 (.40)	.14 (.35)	.05 (.22)	.06 (.24)	.05 (.22)	.05 (.22)	.08 (.28)	.07 (.25)	.06 (.24)	.09 (.28)
Number of first-pass fixations	1.07 (.28)	1.15 (.39)	1.15 (.40)	1.17 (.46)	1.65 (.88)	1.77 (1.04)	1.87 (1.08)	1.88 (1.17)	1.43 (.67)	1.59 (.70)	1.59 (.79)	1.75 (1.10)
First fixation duration	217.62 (73.04)	223.90 (91.40)	221.35 (84.75)	236.67 (91.63)	299.70 (127.60)	298.05 (136.34)	305.06 (134.31)	297.16 (128.19)	265.60 (111.52)	259.19 (105.37)	296.87 (143.00)	261.41 (120.61)
Gaze duration	231.64 (93.81)	254.89 (127.16)	248.33 (115.78)	268.39 (130.82)	458.33 (266.28)	486.32 (313.55)	530.82 (318.39)	534.07 (354.09)	340.31 (198.18)	402.00 (218.19)	441.24 (240.22)	451.25 (309.63)
Total number of fixations	1.31 (1.09)	1.60 (1.06)	1.26 (.84)	1.51 (.99)	2.91 (2.00)	2.91 (2.17)	3.30 (2.37)	3.39 (2.30)	2.85 (2.16)	3.28 (2.26)	3.41 (2.04)	3.62 (2.51)
Total dwell time	284.57 (254.89)	348.63 (249.79)	270.63 (194.42)	337.87 (228.93)	784.49 (579.15)	785.73 (616.70)	925.25 (716.26)	935.78 (643.31)	699.17 (499.91)	821.54 (615.95)	887.73 (537.30)	895.81 (640.30)
Regressions	.26 (.44)	.39 (.52)	.21 (.42)	.33 (.50)	.39 (.61)	.38 (.66)	.33 (.61)	.51 (.78)	.54 (.81)	.59 (.81)	.68 (.80)	.62 (.85)

Note. Cons=consistent, Incons=inconsistent. Standard deviations are in parentheses. First fixation duration, gaze duration, and total dwell time are in milliseconds (ms).

Table 5. *Linear mixed effects model details for early eye-gaze measures.*

	Skipping rate			Number of first-pass fixations			First fixation duration			Gaze duration		
	β	*SE*	*z*	*β*	*SE*	*t*	*β*	*SE*	*t*	*β*	*SE*	*t*
Intercept	-3.32	.22	-15.07***	1.81	.06	31.49***	5.61	.02	250.51***	6.06	.04	161.91***
Fixed effects												
L1 (Chinese)	.48	.29	1.67†	-.22	.08	-2.59*	-.11	.03	-3.23**	-.18	.05	-3.37**
L1 (English)	1.66	.26	6.39***	-.69	.08	-8.62***	-.26	.03	-8.37***	-.63	.05	-12.42***
Consistency	.13	.29	.46	.06	.05	1.06	-.02	.03	-.59	.01	.04	.19
Frequency	.13	.29	.46	-.19	.05	-3.47***	-.01	.02	-.39	-.11	.04	-2.80**
L1-Chinese*Consistency	-.06	.36	-.16	.10	.07	1.60	-.05	.04	-1.35	.02	.04	.45
L1-English*Consistency	-.44	.30	-1.44	-.004	.06	-.06	.06	.04	1.59	.07	.04	1.75†
L1-Chinese*Frequency	-.09	.37	-.24	.004	.07	.06	-.04	.03	-1.08	-.02	.05	-.33
L1-English*Frequency	.17	.30	.58	.15	.06	2.26*	-.02	.03	-.72	.06	.05	1.36
Consistency*Frequency	.19	.58	.33	.13	.11	1.18	.002	.05	.04	.06	.08	.73
L1-Chinese*Cons*Freq	-.79	.73	-1.08	-.13	.14	-.95	.08	.07	1.18	.06	.09	.68
L1-English*Cons*Freq	.03	.60	.05	-.07	.13	-.52	-.05	.06	-.80	-.05	.09	-.61

	Variance Component	*SD*	Variance Component	*SD*	Variance Component	*SD*	Variance Component	*SD*
Random effects								
Participants	.62	.79	.09	.30	.01	.11	.04	.19
Intercept								
Consistency Slope	<.001	.01	<.001	<.001	.004	.06	<.001	<.001
Frequency Slope	<.001	.005	.01	.08	<.001	<.001	.01	.10
Cons*Freq Slope	.001	.02	.03	.18	<.001	<.001	.02	.13
Items Intercept	.14	.37	.01	.10	.001	.03	.01	.07
Residual	--	--	.55	.74	.15	.39	.23	.47

Note. Reference group is L1 Arabic. †p<.10, *p<.05, **p<.01, ***p<.001.

Table 6. *Linear mixed effects model details for total eye-gaze measures.*

	Total number of fixations			Target dwell time			Regressions		
	β	*SE*	*t*	*β*	*SE*	*t*	*β*	*SE*	*t*
Intercept	3.18	.17	18.48***	6.29	.12	54.27***	.40	.04	9.21***
Fixed effects									
L1-Chinese	.16	.23	.70	.001	.16	.01	.21	.06	6.31***
L1-English	-1.75	.22	-7.84***	-1.38	.15	-9.24***	-.11	.05	-2.02*
Consistency	.06	.17	.36	-.02	.14	-.13	.10	.06	1.69†
Frequency	-.51	.17	-2.98**	-.22	.14	-1.62	-.05	.06	-.79
L1-Chinese*Consistency	.31	.14	2.23*	.13	.15	.85	-.09	.05	-1.70†
L1-English*Consistency	.21	.13	1.64	.52	.14	3.84****	.03	.05	.64
L1-Chinese*Frequency	-.002	.14	-.01	-.17	.15	-1.17	-.04	.06	-.76
L1-English*Frequency	.59	.13	4.40***	.10	.14	.74	.11	.05	2.00*
Consistency*Frequency	-.10	.34	-.31	-.19	.28	-.68	-.21	.11	-1.88†
L1-Chinese*Cons*Freq	.25	.28	.91	.64	.31	2.07*	.31	.11	2.88**
L1-English*Cons*Freq	.15	.25	.58	.14	.29	.49	.21	.10	2.09*

	Variance Component	*SD*	Variance Component	*SD*	Variance Component	*SD*
Random effects						
Participants Intercept	.78	.88	.30	.55	.04	.20
Consistency Slope	<.001	<.001	<.001	<.001	<.001	<.001
Frequency Slope	.03	.16	<.001	<.001	.004	.06
Cons*Freq Slope	<.001	<.001	.12	.35	<.001	<.001
Items Intercept	.20	.45	.10	.31	.02	.14
Residual	2.57	1.60	2.95	1.72	.38	.61

Note. Reference group is L1 Arabic. †p<.10, *p<.05, **p<.01, ***p<.001.

---

[i] Although both consistency and regularity are examined in the literature, we use 'consistency' here following the inspiration of this study (Lesch & Pollatsek, 1998).
[ii] The original stimuli included two sets of inconsistent words. However, close inspection of their lexical properties revealed that one set was not well controlled on concreteness and imageability, which impact lexical processing (Barber, Otten, Kousta & Vigliocco, 2013; Tolentino & Tokowicz, 2009). Therefore results are reported only for the set of inconsistent words that were well matched to the consistent words.
[iii] Due to privacy restrictions and time constraints, it was not possible to obtain a reading proficiency score from a standardized assessment (e.g., Woodcock Johnson, Nelson-Denny). Although such assessments are often used in L2 studies, they have typically not been validated or normed with non-native-speaker populations, raising questions about the validity and interpretation of such scores (see also Grant, Gottardo & Geva, 2012; Pasquarella, Gottardo & Grant, 2012; Wolf, Farnsworth & Herman, 2008). It was thus determined that for this study, more direct, objective measures of reading proficiency (e.g., comprehension accuracy, reading rate, exposure to English, etc.) would be used.
[iv] Due to an oversight during stimulus selection two items were repeated, leaving 19 unique pairs. There were no differences in results when including vs. excluding repeated items, therefore, they were included in reported analyses.
[v] The VC unit was chosen for the focus of the consistency manipulation because research has repeatedly shown that it is the most relevant orthographic unit for English word recognition (e.g., Treiman et al., 1995; Ziegler & Goswami, 2005). However, we have also confirmed that the consistency of the CV (body) unit did not differ between consistent and inconsistent words.
[vi] Monospaced fonts are more common in eye-tracking research (Rayner, Slattery & Bélanger, 2010), but proportionally-spaced fonts like Times New Roman are also used (Ashby, Dix, Bontrager, Dey & Archer, 2013; Tuninetti, Warren & Tokowicz, 2015). Research comparing monospaced vs. proportionally-spaced fonts in eye-tracking suggests there is little impact (Rayner et al., 2010), and any influence appears primarily on measures not considered here (fixation location, skipping probability; McDonald, 2006). Further, reading in a familiar font, such as Times New Roman, facilitates reading compared to less-familiar fonts (Slattery & Rayner, 2010), an important consideration for non-native readers.
[vii] Word skipping was not used here because targets were rarely skipped, especially by the non-native speakers; this is consistent with previous findings that less-skilled readers have smaller parafoveal previews and often fixate words multiple times (Ashby et al., 2012; Rayner et al., 2010; Whitford & Titone, 2015).
[viii] Note that although Chinese has a larger number of 'finals' (phonological rimes) and diphthongs than standard Arabic, the number of singleton vowel phonemes is comparable (Chao, 1968).