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

Follow this and additional works at: [https://opensiuc.lib.siu.edu/ling_articles](https://opensiuc.lib.siu.edu/ling_articles)

**Recommended Citation**

This Article is brought to you for free and open access by the Department of Linguistics at OpenSIUC. It has been accepted for inclusion in Articles by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.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
Assembled phonology in native and non-native readers’ lexical processing

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
Assembled phonology in native and non-native readers’ lexical processing

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.¹ 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). 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.\textsuperscript{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. 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, $ps<.05$. 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, $\beta=-.04$, $SE=.03$, $p=.13$). 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, $ps<.01$. 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, $\beta=.07$, $SE=.04$, $p<.05$) than the Arabic speakers (13 ms, $\beta<.001$, $SE=.05$, $p>.50$). 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, $ps<.001$). 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, $\beta=.02$, $SE=.22$, $p>.50$; Chinese, $\beta=.38$, $SE=.29$, $p=.19$), the difference was numerically much larger for the Chinese speakers (.31 more
Assembled phonology in native and non-native readers’ lexical processing

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, χ²(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, ps<.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, χ²(df=8)=49.61, p<.001. The Chinese speakers had more regressions (.61) than the Arabic (.40) or the English speakers (.30, ps<.001); the difference between the Arabic and English speakers approached significance
Assembled phonology in native and non-native readers’ lexical processing

The consistency effect was marginal, $p=.11$, 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, $\beta=.13, SE=.04, p<.01$, but no frequency effect, $\beta=.06, SE=.04, p=.18$, and no consistency by frequency interaction, $\beta=.01, SE=.08, p>.50$. The Arabic speakers had a marginal consistency by frequency interaction, $\beta=-.22, SE=.12, p=.08$. 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, $\beta=.01, SE=.10, p>.50$; frequency, $\beta=-.08, SE=.10, p=.40$; consistency by frequency, $\beta=.09, SE=.19, p>.50$.

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,
Assembled phonology in native and non-native readers’ lexical processing

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), 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
Assembled phonology in native and non-native readers’ lexical processing

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

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

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


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](http://dx.doi.org/10.1016/0749-596X(90)90044-Z)


Assembled phonology in native and non-native readers’ lexical processing


Assembled phonology in native and non-native readers’ lexical processing


Assembled phonology in native and non-native readers’ lexical processing


Assembled phonology in native and non-native readers’ lexical processing


doi: [https://doi.org/10.1016/j.jecp.2014.09.012](https://doi.org/10.1016/j.jecp.2014.09.012)


doi:10.1080/17470210902816461


Table 1. *Participants’ English language skills.*

<table>
<thead>
<tr>
<th>Language skill</th>
<th>English (n=33)</th>
<th>Arabic (n=35)</th>
<th>Chinese (n=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.48 (0.80)</td>
<td>6.14 (1.38)</td>
<td>5.29 (1.46)</td>
</tr>
<tr>
<td>Writing&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.39 (0.93)</td>
<td>5.66 (1.51)</td>
<td>4.57 (1.64)</td>
</tr>
<tr>
<td>Conversational fluency&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.67 (0.69)</td>
<td>6.57 (1.86)</td>
<td>4.96 (1.62)</td>
</tr>
<tr>
<td>Spoken language comprehension&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.79 (0.48)</td>
<td>6.74 (1.56)</td>
<td>4.79 (1.71)</td>
</tr>
<tr>
<td>Vocabulary post-test score&lt;sup&gt;b&lt;/sup&gt;</td>
<td>146.99 (0.17)</td>
<td>136.79 (7.32)</td>
<td>136.43 (9.65)</td>
</tr>
<tr>
<td>Comprehension score&lt;sup&gt;c&lt;/sup&gt;</td>
<td>90.48 (5.26)</td>
<td>86.33 (7.63)</td>
<td>85.71 (7.20)</td>
</tr>
</tbody>
</table>

*Note.* Standard deviations are in parentheses.  
<sup>a</sup>Scores are for self-rated ability. Scale is from 1 (no literacy/fluency) to 10 (full literacy/fluency).  
<sup>b</sup>Maximum score is 147.  
<sup>c</sup>Percent correct.
Table 2. *Participant characteristics.*

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

*Note.* Standard deviations are in parentheses.
Table 3. *Stimulus characteristics.*

<table>
<thead>
<tr>
<th></th>
<th>Inconsistent</th>
<th>Consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High frequency</td>
<td>Low frequency</td>
</tr>
<tr>
<td>Log frequency</td>
<td>10.44 (.92)</td>
<td>7.41 (1.20)</td>
</tr>
<tr>
<td>Length</td>
<td>5.40 (.52)</td>
<td>5.10 (.32)</td>
</tr>
<tr>
<td>Concreteness</td>
<td>400.29 (117.72)</td>
<td>506.00 (123.20)</td>
</tr>
<tr>
<td>Imageability</td>
<td>454.88 (100.37)</td>
<td>524.00 (93.60)</td>
</tr>
<tr>
<td>Number of orthographic neighbors</td>
<td>6.60 (3.37)</td>
<td>8.10 (3.45)</td>
</tr>
<tr>
<td>Frequency of orthographic neighbors</td>
<td>6.69 (1.11)</td>
<td>7.97 (1.30)</td>
</tr>
<tr>
<td>Bigram sum</td>
<td>16546.20 (10267.56)</td>
<td>14588.20 (7666.89)</td>
</tr>
<tr>
<td>Bigram mean</td>
<td>3684.15 (2025.53)</td>
<td>3502.90 (1655.62)</td>
</tr>
</tbody>
</table>

*Note.* Standard deviations are in parentheses.
Table 4. Eye-gaze measures by L1 group and stimulus type.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Arabic</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Frequency</td>
<td>Low Frequency</td>
<td>High Frequency</td>
</tr>
<tr>
<td>Skipping rate</td>
<td>Cons</td>
<td>Incons</td>
<td>Cons</td>
</tr>
<tr>
<td></td>
<td>.22</td>
<td>.19</td>
<td>.19</td>
</tr>
<tr>
<td>Number of fixations</td>
<td>Cons</td>
<td>Incons</td>
<td>Cons</td>
</tr>
<tr>
<td>First fixation</td>
<td>1.07</td>
<td>1.15</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>(.28)</td>
<td>(.39)</td>
<td>(.40)</td>
</tr>
<tr>
<td>First fixation</td>
<td>217.62</td>
<td>223.90</td>
<td>221.35</td>
</tr>
<tr>
<td>duration</td>
<td>(73.04)</td>
<td>(91.40)</td>
<td>(84.75)</td>
</tr>
<tr>
<td>Gaze duration</td>
<td>231.64</td>
<td>254.89</td>
<td>248.33</td>
</tr>
<tr>
<td></td>
<td>(93.81)</td>
<td>(127.16)</td>
<td>(115.78)</td>
</tr>
<tr>
<td>Total number of</td>
<td>1.31</td>
<td>1.60</td>
<td>1.26</td>
</tr>
<tr>
<td>fixations</td>
<td>(1.09)</td>
<td>(1.06)</td>
<td>(.84)</td>
</tr>
<tr>
<td>Total dwell time</td>
<td>284.57</td>
<td>348.63</td>
<td>270.63</td>
</tr>
<tr>
<td></td>
<td>(254.89)</td>
<td>(249.79)</td>
<td>(194.42)</td>
</tr>
<tr>
<td>Regressions</td>
<td>.26</td>
<td>.39</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>(.44)</td>
<td>(.52)</td>
<td>(.42)</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>Skipping rate</th>
<th>Number of first-pass fixations</th>
<th>First fixation duration</th>
<th>Gaze duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>( SE )</td>
<td>( z )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.32</td>
<td>.22</td>
<td>-15.07***</td>
<td>1.81</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1 (Chinese)</td>
<td>.48</td>
<td>.29</td>
<td>1.67†</td>
<td>-.22</td>
</tr>
<tr>
<td>L1 (English)</td>
<td>1.66</td>
<td>.26</td>
<td>6.39***</td>
<td>-.69</td>
</tr>
<tr>
<td>Consistency</td>
<td>.13</td>
<td>.29</td>
<td>.46</td>
<td>.06</td>
</tr>
<tr>
<td>Frequency</td>
<td>.13</td>
<td>.29</td>
<td>.46</td>
<td>-.19</td>
</tr>
<tr>
<td>L1-Chinese*Consistency</td>
<td>-.06</td>
<td>.36</td>
<td>-.16</td>
<td>.10</td>
</tr>
<tr>
<td>L1-English*Consistency</td>
<td>-.44</td>
<td>.30</td>
<td>-1.44</td>
<td>-.004</td>
</tr>
<tr>
<td>L1-Chinese*Frequency</td>
<td>-.09</td>
<td>.37</td>
<td>-2.4</td>
<td>.004</td>
</tr>
<tr>
<td>L1-English*Frequency</td>
<td>.17</td>
<td>.30</td>
<td>.58</td>
<td>.15</td>
</tr>
<tr>
<td>Consistency*Frequency</td>
<td>.19</td>
<td>.58</td>
<td>.33</td>
<td>.13</td>
</tr>
<tr>
<td>L1-Chinese<em>Cons</em>Freq</td>
<td>-.79</td>
<td>.73</td>
<td>-1.08</td>
<td>-.13</td>
</tr>
<tr>
<td>L1-English<em>Cons</em>Freq</td>
<td>.03</td>
<td>.60</td>
<td>.05</td>
<td>-.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Variance Component</th>
<th>SD</th>
<th>Variance Component</th>
<th>SD</th>
<th>Variance Component</th>
<th>SD</th>
<th>Variance Component</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>.62</td>
<td>.79</td>
<td>.09</td>
<td>.30</td>
<td>.01</td>
<td>.11</td>
<td>.04</td>
<td>.19</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td>&lt;.001</td>
<td>.01</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.01</td>
<td>.06</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Frequency</td>
<td>&lt;.001</td>
<td>.005</td>
<td>.01</td>
<td>.08</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.01</td>
<td>.10</td>
</tr>
<tr>
<td>Cons*Freq</td>
<td>.001</td>
<td>.02</td>
<td>.03</td>
<td>.18</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.02</td>
<td>.13</td>
</tr>
<tr>
<td>Items Intercept</td>
<td>.14</td>
<td>.37</td>
<td>.01</td>
<td>.10</td>
<td>.001</td>
<td>.03</td>
<td>.01</td>
<td>.07</td>
</tr>
<tr>
<td>Residual</td>
<td>--</td>
<td>--</td>
<td>.55</td>
<td>.74</td>
<td>.15</td>
<td>.39</td>
<td>.23</td>
<td>.47</td>
</tr>
</tbody>
</table>

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.*

<table>
<thead>
<tr>
<th></th>
<th>Total number of fixations</th>
<th>Target dwell time</th>
<th>Regressions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>SE</td>
<td>t</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.18</td>
<td>.17</td>
<td>18.48***</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1-Chinese</td>
<td>.16</td>
<td>.23</td>
<td>.70</td>
</tr>
<tr>
<td>L1-English</td>
<td>-1.75</td>
<td>.22</td>
<td>-7.84***</td>
</tr>
<tr>
<td>Consistency</td>
<td>.06</td>
<td>.17</td>
<td>.36</td>
</tr>
<tr>
<td>Frequency</td>
<td>-.51</td>
<td>.17</td>
<td>-2.98**</td>
</tr>
<tr>
<td>L1-Chinese*Consistency</td>
<td>.31</td>
<td>.14</td>
<td>2.23*</td>
</tr>
<tr>
<td>L1-English*Consistency</td>
<td>.21</td>
<td>.13</td>
<td>1.64</td>
</tr>
<tr>
<td>L1-Chinese*Frequency</td>
<td>-.002</td>
<td>.14</td>
<td>-.01</td>
</tr>
<tr>
<td>L1-English*Frequency</td>
<td>.59</td>
<td>.13</td>
<td>4.40***</td>
</tr>
<tr>
<td>Consistency*Frequency</td>
<td>-.10</td>
<td>.34</td>
<td>-.31</td>
</tr>
<tr>
<td>L1-Chinese<em>Cons</em>Freq</td>
<td>.25</td>
<td>.28</td>
<td>.91</td>
</tr>
<tr>
<td>L1-English<em>Cons</em>Freq</td>
<td>.15</td>
<td>.25</td>
<td>.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Variance Component</th>
<th>SD</th>
<th>Variance Component</th>
<th>SD</th>
<th>Variance Component</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants Intercept</td>
<td>.78</td>
<td>.88</td>
<td>.30</td>
<td>.55</td>
<td>.04</td>
<td>.20</td>
</tr>
<tr>
<td>Consistency Slope</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Frequency Slope</td>
<td>.03</td>
<td>.16</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.004</td>
<td>.06</td>
</tr>
<tr>
<td>Cons*Freq Slope</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.12</td>
<td>.35</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Items Intercept</td>
<td>.20</td>
<td>.45</td>
<td>.10</td>
<td>.31</td>
<td>.02</td>
<td>.14</td>
</tr>
<tr>
<td>Residual</td>
<td>2.57</td>
<td>1.60</td>
<td>2.95</td>
<td>1.72</td>
<td>.38</td>
<td>.61</td>
</tr>
</tbody>
</table>

*Note.* Reference group is L1 Arabic. †p<.10, *p<.05, **p<.01, ***p<.001.
Although both consistency and regularity are examined in the literature, we use ‘consistency’ here following the inspiration of this study (Lesch & Pollatsek, 1998).

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.

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.

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.

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.

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.

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).

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).