Journal of Experimental Child Psychology, in press

Does visual letter similarity modulate masked form priming in young readers of Arabic?

Manuel Perea ¹², Reem Abu Mallouh ², Ahmed Mohammed ², Batoul Khalifa ³, and Manuel Carreiras ²⁴

¹ Universitat de València, Valencia, Spain

² Basque Center on Cognition, Brain, and Language, Donostia, Spain

³ Qatar University, Doha, Qatar

⁴ Ikerbasque. Basque Foundation for Science, Bilbao, Spain

Short title: visual letter similarity in Arabic

Word count: 3995 (main text + references)

Correspondence:

Manuel Perea Departamento de Metodología Universitat de València Av. Blasco Ibáñez, 21 46010-Valencia (Spain) Email: mperea@uv.es

Author's notes: This report was made possible by a NPRP award [Grant NPRP No. 6-378-5-035] from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

Abstract

We carried out a masked priming lexical decision experiment to study whether visual letter similarity plays a role during the initial phases of word processing in young readers of Arabic (5th Graders). Arabic is ideally suited to test these effects because most Arabic letters share their basic shape with at least one other letter and only differ in the number/position of diacritical points.

(e.g., روز - روز

Key words: Lexical access; Masked priming; Developing readers; Visual similarity; Lexical decision Most theorists would agree that, in a mature reading system, the abstract letter units that drive the process of lexical access are activated in the initial phases of processing (see Grainger, Dufau, & Ziegler, 2016). Indeed, in the masked priming paradigm, a visually dissimilar lowercase-uppercase identical pair such as arte-ARTE produces as much masked repetition priming—relative to an unrelated control—as a visually similar lowercase-uppercase identical pair such as kiss-KISS (Bowers, Vigliocco, & Haan, 1998; see also Jacobs, Grainger, & Ferrand, 1995). Importantly, the mapping from visual letter forms to abstract letter representations also appears to occur early in processing with developing readers (see Perea, Jiménez, & Gomez, 2015, for a replication of the Bowers et al., 1998, and the Jacobs et al., 1995, experiments with Grade 5 children). However, most developmental research on visual word recognition and letter processing has been conducted in English or other Indo-European languages that employ the Roman alphabet (e.g., Grainger, Lété, Bertrand, Dufau, & Ziegler, 2012; Nation, 2009). In the present study, we examined the role of visual letter similarity during the initial stages of word processing—using the masked priming technique—in young readers of an underrepresented language (namely, Arabic) that has two distinctive characteristics in terms of morphology and visual features.

First, as in other Semitic languages (e.g., Hebrew), words in Arabic are typically created by putting together a three-letter consonantal root that denotes the general meaning, and a word pattern that indicates the specific inflectional/derived form. For instance, the consonantal root بسبح, which means to swim (the Buckwalter transliteration is sbH), can be used to form a number of different words such as سباحة (swimming [sbAHp]; word pattern: CCACp, where the Cs denote the consonants from the root), حسبح swimmer [sbAH] word pattern: CCAC), or مسبح (swimming pool [msbH]; word pattern: mCCC). The consonantal root appears to play a prevalent role during lexical

access in Semitic languages. In fact, it has been claimed that lexical space in Semitic languages is organized morphologically, whereas it is organized orthographically in Indo-European languages (Frost, 2009). A key finding supporting this dissociation is that while masked morphological priming in lexical decision is sizeable in Hebrew or Arabic adult readers (e.g., جميل - جميل [gmyl-gmAl] lovely—beauty; the root is [gml] in both cases), masked orthographic priming effects with word/nonword primes tend to be null or minimal (as in the word pair غيمة - خيمة [xymp-gymp] tent-cloud; the roots are [xym] and عنيم [gym]) (see Frost, Kugler, Deutsch, & Forster, 2005, for evidence in Hebrew and Arabic; see also Velan & Frost, 2009). In contrast, masked orthographic priming is a highly replicable finding in Indo-European languages (Forster et al., 2007; see also Castles et al., 2007; Comesaña, Soares, Marcet, & Perea, 2016, for evidence with young readers). Nonetheless, recent studies with adult Arabic readers reported significant masked orthographic effects in Arabic under some circumstances (e.g., using the go/no-go variant of lexical decision task; see Perea, Abu Mallouh, & Carreiras, 2013).

Second, Arabic is written in a right-to-left semi-cursive script in which some letters may be connected to the previous letter, thus forming subwords (e.g., the word [comfort rAHp] is composed of three subwords: [r-A-Hp]). Importantly, most Arabic letters share their basic shape with at least one other letter and only differ in the number/position of diacritical points

(e.g., رض - ص; ظ - ط; غ - ع; ث - ت - ن - ب; ذ - د; خ - ح - ج; ق - ف; ش - س ; ر - س . As a result, many Arabic words look physically the same, except for the presence/location of diacritical points (e.g., بحار [sailor; bHAr] and بخار [steam; bxAr]). The evidence concerning the role of visual similarity effect in the early stages of word processing with young readers in Arabic is very scarce. Perea, Abu Mallouh, and Carreiras (2013)

examined whether the ligation pattern of Arabic words plays a role during the early moments of word recognition with Grade 3 and Grade 6 children. They employed a masked priming lexical decision task in which the target words could be preceded by a morphologically-related, substituted-letter prime that kept the same ligation pattern (e.g., كتاب - كت

A more direct approach to examine visual similarity effects during word processing in Arabic is not in terms of the ligation pattern of the words, but in terms of visual letter similarity (e.g., nevtral and document may be initially processed as neutral and document; see Marcet & Perea, 2017, in press, for evidence with the Roman alphabet). As indicated above, many Arabic letters differ from other letters solely in the number/position of diacritical marks. In a masked priming lexical decision experiment with adult readers, Perea, Abu Mallouh, Mohammed, Khalifa, and Carreiras (2016) employed substituted-letter primes that kept the same shape as the replaced letter and only differed in the number/position of the diacritical points (e.g., صحفیة — صحفیة — صحفیة — صحفیة — SHfyp-Sxfyp]) and substituted-letter primes that had a different shape from the replaced

letter (e.g., صحفیة – سكفیة [SHfyp - Skfyp]). (In all cases, the replaced letter was a letter from the consonantal root.) As a control, an identity priming condition was employed (e.g., صحفیة - صحفیة - صحفیة - صحفیة [SHfyp - SHfyp]). Results showed remarkably similar word identification times in the two replaced-letter conditions, which in turn produced slower word identification times than the identity condition. Perea et al. (2016) concluded that adult Arabic readers are able to process the letter's diacritical points very quickly (see Wiley, Wilson, & Rapp, 2016, for a pivotal role of diacritical points with Arabic readers in a same-different task with pairs of letters). A potential limitation of this study was the lack of an unrelated condition. That is, it could be argued that there could have been a floor effect due to the fact that the replaced-letter pairs did not share the consonantal root with the target word. Furthermore, one important remaining question is whether the effects of visual letter similarity on masked orthographic priming can be obtained with normally developing young readers of Arabic.

processing the letters' diacritical marks in the early moments of word processing (i.e., could be processed as خدمة), one would expect an advantage of the same-shape substituted-letter condition over the different-shape substituted-letter condition. Alternatively, if developing readers are able to quickly process the diacritical marks, the visually similar prime حدمة would not be more effective than the visually dissimilar prime فدمة at activating the word خدمة. Furthermore, we used two control conditions. One was the identity condition. This allowed us to examine the degree of activation of the orthographically related conditions relative to the identity condition (e.g., خدمة – سكاب , (footnote_1). The other was the unrelated condition (e.g., خدمة [skAb-xdmp]). This allowed us to examine the degree of activation of the orthographically related conditions relative to the unrelated condition (i.e., the effect of masked orthographic priming). If, as argued by Frost (2009), lexical space in Semitic languages is organized morphologically via the consonantal root (e.g., the letters xdm in the word xdmp [service]), orthographic (but not morphological) letter primes (e.g., *Hdmp*) should not be more effective at activating the target word than an unrelated prime (e.g., skAb). Alternatively, if the difference in word processing between Semitic and Indo-European languages is more quantitative (e.g., more dense lexical space in Semitic languages) than qualitative, a processing advantage of the orthographically related conditions over the unrelated condition could be expected

We employed the go/no-go variant of the lexical decision task because it produces shorter word identification times and higher accuracy rates than the two-choice variant in developing readers (see Perea, Comesaña, & Soares, 2013).

Furthermore, Perea et al. (2013) found sizeable masked orthographic priming effects with adult readers in Arabic using this procedure. Finally, as Arabic does not have a lowercase/uppercase distinction and to minimize visual continuity between primes and

targets, the primes were presented in a smaller size than the targets (see Frost et al., 1997; Perea et al., 2013, 2017, for the same procedure).

Method

Participants. Thirty-two Grade 5 children of 11-12 years, all boys, from a public school in Qatar participated in the experiment—the experiment took place at the end of the academic year. All were native speakers of Arabic and had normal/corrected-tonormal vision. None of them had any learning or reading difficulties. Each of the children' parents signed an informed consent form before the experiment. **Materials**. We extracted 180 Arabic five-letter words from textbooks aimed at primary school in Qatar. To verify that Grade 5 children knew the words, we asked four children from Grade 4 in Qatar (i.e., a lower grade than the participants in the experiment) to read and tell the meaning of the words. None of them had any problems indicating the meaning of these words. We also presented the list of words to two primary school teachers to verify that Grade 5 children would know the words. The average frequency per million of these words in the Modern Standard Arabic database (Aralex; Boudelaa & Marslen-Wilson, 2010) was 26.9 (range: 0.03-360.4). Each selected word had critical pairs of letters with the same basic shape (e.g., ψ - ψ ; $\dot{\psi}$ - $\dot{\psi}$; $\dot{\psi}$ - $\dot{\psi}$). The primetarget conditions were as follows: a) the prime was the same as the target (e.g., [xdmp-xdmp] service; identity condition; the root is خدمة - خدمة prime was a pseudoword created by substituting a letter from the consonantal root that kept the same shape (e.g., خدمة – حدمة [Hdmp-xdmp]): same-shape replaced-letter condition); b) the prime was a pseudoword created by substituting a letter from the consonantal root that had a different shape, but kept the ligation pattern (e.g.,

[fdmp-xdmp]; different-shape substituted-letter condition); or c) the prime was an unrelated pseudoword (e.g., خدمة – سكاب [skAb-xdmp]; unrelated condition). (Given the difficulty of creating the pseudoword primes, three of the primes were actually infrequent Arabic words probably unknown to the children—excluding these trials did not affect the pattern of findings.) We rotated the four prime-target conditions across each list in a Latin square manner. Eight participants were assigned to each of the four lists—they received 45 items per condition (4320 data points overall). For the purposes of the lexical decision task, we also created 180 five-letter pseudowords (e.g., that contained the same critical pairs of letters as the word stimuli. For these حمهر, مائس pseudoword trials, we created prime stimuli with the same characteristics as those for word trials. The entire set of stimuli is presented in the supplemental materials. **Procedure.** The experiment took place individually in a silent room equipped with a computer running DMDX (Forster & Forster, 2003). Each trial had the following sequence: 1) A 500-ms pattern mask (i.e., a series of #'s); 2) a 50-ms prime stimulus in DejaVu Sans Mono 14-pt; and 3) a target stimulus in DejaVu Sans Mono 28-pt that was displayed until the participant's response or 2 seconds had elapsed. All stimuli werepresented centered on the computer monitor. Participants were told to press the "yes" button if the letter string was an Arabic word and to withhold the response otherwise. Both speed and accuracy were stressed in the instructions. The 360 experimental trials were preceded by a short practice phase consisting of 16 trials. To minimize tiredness, we included short breaks every 90 trials. Each participant received a random ordering of the stimuli. The session lasted for around 20 min.

Results

Very fast lexical decision responses (less than 250 ms; 11 data points) and incorrect responses were removed from the response time analyses. Table 1 displays the mean correct lexical decision times (in ms) and accuracy (in proportion) for word targets in each experimental condition (identity condition; same-shape substituted-letter priming condition; different-shape substituted-letter priming condition; unrelated condition); for nonword targets (i.e., no-go trials), we indicated the accuracy in each condition.

To answer the research questions posed in the Introduction, we conducted three orthogonal contrasts using linear mixed effects models with Prime-target relationship as a fixed factor in the design: 1) is there an advantage of the identity condition over the form-related conditions? (identity condition vs. substituted-letter conditions); 2) is there a visual letter-similarity effect for the orthographically-related primes? (same-shape substituted-letter prime condition vs. different-shape substituted-letter condition); 3) is there an orthographic priming effect? (substituted-letter conditions vs. unrelated condition). The number of data points in the RT analyses was 5388. To maintain the normality assumption of linear mixed effects models, RTs were inverse-transformed (-1000/RT). The model included Prime-target as a fixed factor with the above-cited orthogonal contrasts with a maximal random effect structure of subjects and items.

These analyses were conducted in R using the packages *lme4* (Bates, Maechler, Bolker, & Walker 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2016).

Question 1: Identity vs. orthographically-related conditions. Word recognition times were, on average, 16.5 ms faster when preceded by the identity prime than when preceded by form-related prime, t = -3.67, b = -0.06, SE = 0.02, p < .001. Question 2: Visual-similarity effects in orthographically-related conditions. On average, lexical decision times were virtually the same for those words preceded by a replaced-

letter visually similar prime and for those words preceded by a replaced-letter visually dissimilar prime (667 vs. 668 ms, respectively), t < 1.1, p > .29.

Question 3: Orthographically-related vs. unrelated conditions. Word identification times were, on average, 21.5 ms faster when preceded by a, orthographically related prime than when preceded by an unrelated prime, t = 5.42, b = 0.067, SE = 0.01, p < .001.

The statistical analyses of the accuracy data on word trials and nonword trials were similar to those described above, except that we employed the *glmer* function in R. None of the contrasts approached significance, all ps > .21.

Discussion

The main findings of the current masked priming experiment with Grade 5 children in Arabic are as follows. First, we found a sizeable 21.5 ms masked orthographic priming effect (e.g., خدمة مسكاني [fdmp-xdmp] is recognized faster than خدمة مسكاني [skAb-xdmp]), thus showing that the consonantal root is not critical for the initial contact to lexical entries in developing readers of Arabic. Second, word identification times were similar for those target words that were preceded by a same-shape substituted-letter prime (خدمة مدمة [Hdmp-xdmp]) and by a different-shape substituted-letter prime (خدمة مدمة [fdmp-xdmp]), thus showing that participants had access to abstract letter/word representations not mediated by visual letter similarity—at least not in terms of letter shape. Third, the two orthographically related conditions produced longer word response times than the identity condition, thus showing that the cognitive system was sensitive to the full match of all the letters from the prime stimuli.

As indicated in the Introduction, a common claim in the literature on visual word recognition and reading is that lexical space of Semitic languages is organized

differently from that of Indo-European languages (see Frost, 2009). A benchmark finding supporting this view is that, unlike Indo-European languages, masked orthographic priming is absent—or minimal— in Hebrew and Arabic (Frost, 2009; Frost et al., 1997). However, we found a statistically robust 21.5 ms masked orthographic priming effect with developing readers in Arabic—this extends the findings reported by Perea et al. (2014) with adult readers in Arabic. Furthermore, the magnitude of the effect is similar to that obtained in prior orthographic priming experiments with developing readers in Indo-European languages (e.g., see Comesaña et al., 2016). Therefore, sharing the root letters is not necessary to access whole-word units in Arabic. Taken together, the data from adults and developing children suggest that any potential differences between masked priming effects in Semitic and Indo-European words are not qualitative. Indeed, there is evidence suggesting that the mechanisms underlying visual word recognition are also similar in other writing systems (e.g., the hiragana and katakana scripts of Japanese; see Okano, Grainger, & Holcomb, 2013).

Another key question of the current experiment was whether masked orthographic priming was modulated by visual letter similarity in developing readers. To shed some light on this issue, we employed two orthographically related conditions, one in which the replaced letter kept the same shape as the original letter (i.e., the only difference was the number/position of diacritical points) and another condition in which the replaced letter was visually different (e.g., خدمة – خدمة – عدمة – كدمة – خدمة – خدمة – خدمة (e.g., خدمة – خد

points would produce a large amount of uncertainty in the word recognition system. Further research should examine whether this pattern holds for beginning readers of Arabic or for adult readers of an Indo-European language acquiring Arabic as a second language. Importantly, Carreiras, Perea, and Abu Mallouh (2012) and Carreiras, Perea, Gil-López, Abu Mallouh, and Salillas (2012, 2013) investigated whether visual form influenced letter processing in the Arabic alphabet, by using masked priming paradigms with isolated letters that had a different letter shape depending on the position within a word. While Carreiras et al. (2012) showed masked repetition priming effects of the same magnitude for letter pairs with similar and with dissimilar visual features, Carreiras et al. (2013) showed an early transient effect of visual similarity in early components (P/N150), followed by an effect of abstract letter priming in a later component (P300) in adult skilled readers. Therefore, it is possible that effects of visual letter similarity can occur in skilled Arabic readers at early stages of processing, but they can be weaker at later stages of processing or when processing becomes more complex (e.g., during the processing of whole words as compared to isolated letters). A masked priming experiment in which the participants' event-related potentials are recorded would be necessary to test this hypothesis.

To sum up, we found that masked orthographic priming can be readily obtained in Arabic with developing readers, hence suggesting that the word identification stream in Semitic and Indo-European languages is more similar than previously thought.

Furthermore, typically developing Arabic readers are able to quickly and effectively process the diacritical marks of Arabic letters. Additional longitudinal research with children/adults learning to read Arabic is necessary to establish the mechanisms underlying the processing of diacritical points during letter and word recognition.

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
- Boudelaa, S., & Marslen-Wilson, W.D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods*, 42, 481–487. doi:10.3758/brm.42.2.481
- Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1705–1719. doi:10.1037/0096-1523.24.6.1705
- Carreiras, M., Perea, M., & Abu Mallouh, R. (2012). Priming of abstract letter representations may be universal: The case of Arabic. *Psychonomic Bulletin and Review*, 19, 685–690. doi:10.3758/s13423-012-0260-8
- Carreiras, M., Perea, M., Gil-López, C., Abu Mallouh, R., & Salillas, E. (2013). Neural correlates of visual vs. abstract letter processing in Roman and Arabic scripts. *Journal of Cognitive Neuroscience*, 25, 1975–1985. doi:10.1162/jocn_a_00438
- Forster, K.I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*, 116–124. doi:10.3758/bf03195503
- Forster, K.I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *Quarterly Journal of Experimental Psychology*, *39*, 211–251. doi:10.1080/14640748708401785.
- Frost, R., Kugler, T., Deutsch, A., & Forster, K.I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language.

 *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31, 1293–1326. doi:10.1037/0278-7393.31.6.1293.

- Frost, R. (2009). Reading in Hebrew vs. Reading in English: Is there a qualitative difference? In K. Pugh & P. McCradle (Eds.), *How children learn to read:*Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice (pp. 235-254). New York: Psychology Press.
- Grainger, J., Dufau, S., & Ziegler, J.C. (2016). A vision of reading. *Trends in Cognitive Sciences*, 20, 171–179. doi:10.1016/j.tics.2015.12.008
- Grainger, J., Lété, B., Bertrand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, 123, 280–292. doi:10.1016/j.cognition.2012.01.003
- Jacobs, A.M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics*, 57, 1101–1110. doi:10.3758/BF03208367
- Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2016). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package): R package version 2.0-30.
- Marcet, A., & Perea, M. (2017). Is nevtral NEUTRAL? Visual similarity effects in the early phases of written-word recognition. *Psychonomic Bulletin and Review*, 24, 1180–1185. doi:10.3758/s13423-016-1180-9
- Marcet, A., & Perea, M. (in press). Can I order a burger at rnacdonalds.com? Visual similarity effects of multi-letter combinations at the early stages of word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. doi:10.1037/xlm0000477
- Nation, K. (2009). Form-meaning links in the development of visual word recognition.

 *Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 3665–3674. doi:10.1098/rstb.2009.0119
- Okano, K., Grainger, J., & Holcomb, P.J. (2013). An ERP investigation of visual word recognition in syllabary scripts. *Cognitive, Affective & Behavioral Neuroscience*, 13, 390–404. doi:10.3758/s13415-013-0149-7

- Perea, M., Abu Mallouh, R., & Carreiras, M. (2013). Early access to abstract representations in developing readers: Evidence from masked priming.

 Developmental Science, 16, 564–573. doi:10.1111/desc.12052
- Perea, M., Abu Mallouh, & Carreiras, M. (2014). Are root letters compulsory for lexical access in Semitic languages? The case of masked form-priming in Arabic.

 *Cognition, 132, 491–500. doi:10.1016/j.cognition.2014.05.008
- Perea, M., Abu Mallouh, R., Mohammed, A., Khalifa, B., & Carreiras, M. (2016). Do diacritical marks play a role at the early stages of word recognition in Arabic? *Frontiers in Psychology*, 7, 1255. doi: 10.3389/fpsyg.2016.01255
- Perea, M., Jiménez, M., & Gómez, P. (2015). Do young readers have fast access to abstract lexical representations? Evidence from masked priming. *Journal of Experimental Child Psychology*, 129, 140–147. doi:10.1016/j.jecp.2014.09.005
- Perea, M., Soares, A.P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word-identification times in young readers. *Journal of Experimental Child Psychology*, 116, 37–44. doi:10.1016/j.jecp.2012.10.014
- Velan, H., & Frost, R. (2009). Letter-transposition effects are not universal: The impact of transposing letters in Hebrew. *Journal of Memory and Language*, 61, 285–302. doi:10.1016/j.jml.2009.05.003
- Wiley, R.W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1186–1203. doi:10.1037/xhp0000213

Footnotes

1. For word trials, identity primes were words whereas the orthographically related primes were pseudowords—the number of word pairs such as بخار and بخار is very small. Of note, prior research in the Roman alphabet has reported slightly faster word response times for word targets when preceded by an identity prime than when preceded by a visually similar pseudoword prime (e.g., document-DOCUMENT vs. document-DOCUMENT; Marcet & Perea, 2017, in press).

Table 1. Mean response times (RTs; in ms) and accuracy for words and nonwords in the four prime-target conditions of the experiment. The standard error or the mean are presented in parentheses.

| | Prime-target Relationship | | | | |
|----------|---------------------------|-------------------------------|------------------------------------|-------------|--|
| | <u>Identity</u> | Same shape Substituted-letter | Different shape Substituted-letter | Unrelated | |
| Words | | | | | |
| RT | 651 (25) | 667 (29) | 668 (26) | 689 (24) | |
| Accuracy | .931 (.009) | .940 (.010) | .942 (.006) | .937 (.007) | |
| Nonwords | | | | | |
| RT | | | | | |
| Accuracy | .913 (.012) | .910 (.014) | .919 (.011) | .915 (.009) | |
| | | | | | |