

ARTICLE

Does phonological distance impact quality of phonological representations? Evidence from Arabic diglossia

Elinor SAIEGH-HADDAD^{1*} and Lina HAJ^{1,2}

¹Bar-Ilan University, Israel and ²Western Galilee Hospital, Israel

*Corresponding author. Department of English Literature and Linguistics, Bar-Ilan University, Ramat-Gan 52900, Israel. E-mail: Elinor.Saiegh-Haddad@biu.ac.il

(Received 27 November 2017; revised 6 May 2018; accepted 3 July 2018;
first published online 13 August 2018)

Abstract

The study tested the impact of the phonological distance between Spoken Arabic (SpA) and Standard Arabic (StA) on quality of phonological representations among kindergarten, first-, second-, and sixth-grade Arabic-speaking children (N = 120). A pronunciation accuracy judgment task targeted three types of StA words that varied in extent of phonological distance from their form in SpA: (a) IDENTICAL WORDS, with an identical lexical–phonological form in StA and SpA; (b) COGNATE WORDS, with partially overlapping phonological forms; items in this category varied in degree of phonological distance too; and (c) UNIQUE WORDS with entirely different lexical–phonological forms. Multilevel Regression analysis showed that phonological distance had a significant impact on quality of phonological representations across all grades. Growth in quality of phonological representations was mainly noted between the three younger groups and the sixth-graders. Implications for the impact of phonological distance on phonological representations and on language and literacy development are discussed.

Keywords: Arabic; diglossia; literacy; phonological distance; phonological representations

Introduction

Phonological representations play a fundamental role in word reading development (Perfetti, 2007; Perfetti & Hart, 2001; Perfetti & Stafura, 2014), and in the development of other critical academic skills, such as phonological processing, word retrieval, word learning, word decoding, and reading comprehension (e.g., Gathercole & Baddeley, 1990; Gray, 2005; Nouwens, Groen, Kleemans, & Verhoeven, 2018; Perfetti, Wlotko, & Hart, 2005; Verhoeven, Leeuwe, Irausquin, & Segers, 2016). This warrants an in-depth understanding of the linguistic factors that impact the establishment and development of phonological representations. Also warranted is an investigation of the role of sociolinguistic factors that pertain to language exposure and use on phonological representations, especially in bilingual and bilingual contexts (Gibson, Summers, Pena, Dedore, Gillam, & Bohman, 2015; Saiegh-Haddad &

Ghawi-Dakwar, 2017; Themistocleous, 2017). The current study investigates phonological representations in Arabic diglossia. In this context, children simultaneously develop two linguistic systems: Spoken Arabic (SpA) for everyday informal speech and Standard Arabic (StA) for formal speech and reading/writing. Importantly, the StA phonological form of the word may vary in extent of distance from (alternatively proximity or overlap with) its form in SpA, and this distance can range from almost identical to entirely different forms. The study tests the impact of the phonological distance between the SpA and StA forms on quality of phonological representations in children at different developmental stages from kindergarten to sixth grade. By studying the impact of the phonological distance between StA and SpA on the establishment and development of phonological representations in StA, this investigation will shed light on the factors that impact phonological representations in diglossic Arabic, in particular, and in analogous bilingual and bilingual contexts.

Phonological representations

Lexical Quality (Perfetti, 2007; Perfetti & Hart, 2001) refers to the quality of stored phonological, orthographic, and semantic representations of lexical items, as well as the retrieval efficiency of these representations. High-quality lexical representations are thought to develop over time and to allow for efficient retrieval of words. Phonological representations, which capture the range of phonological information about words that language users store in long-term memory (Elbro & Jensen, 2005), are a fundamental component of the lexical representations of words and are essential to language and literacy development (e.g., Elbro, Borström, & Petersen, 1998; Nouwens *et al.*, 2018; Verhoeven *et al.*, 2016).

Research on phonological representations concerned the underlying nature of this construct and the linguistic underpinnings that underlie its establishment and development. In this respect, it has been demonstrated that phonological representations are multilayered and allow speakers to represent and process phonological information at different levels (e.g., the syllable, the onset-rime, the phoneme). At the same time, phonological representations are dynamic and developmental and they undergo enormous change during the first few years of a child's speech and language development (Fowler, 1991). The LEXICAL RESTRUCTURING MODEL (Metsala & Walley, 1998) captures the factors that influence the development of phonological representations. According to this model, because the vocabulary store of young children is small, there is no need to represent words in a detailed manner, and therefore early word representations may be holistic. As vocabulary grows, however, these holistic representations are gradually restructured, so that smaller units of sound such as syllables, onset-rimes, and, ultimately, phonemes are encoded (Metsala, 1997; Metsala & Walley, 1998; Walley, 1993).

Phonological representations are also word-specific, and as a result sensitive to the phonological properties of the word, such as the size of the phonological family (or phonological neighborhood density); As such, the increasing number of similarly sounding words in the child's mental lexicon creates developmental pressure to develop representations that encode smaller units of sound for words in dense phonological neighborhoods (De Cara & Goswami, 2003). Research shows that functional properties of words, such as familiarity status, indexed by word frequency and age of acquisition, also impact phonological representational quality (Metsala,

1997; Storkel, Armbrüster, & Hogan, 2006). Finally, phonological representations are affected by letter knowledge and by exposure to literacy. As such, understanding that a particular letter corresponds to a particular phoneme might precipitate the restructuring of the lexicon (Carroll & Snowling, 2001; Dollaghan & Campbell, 1998; Gray & Brinkley, 2011; Walley, 1993), especially when the orthography is transparent (de Gelder & Vroomen, 1991; Goswami, Schneider, & Scheurich, 1999).

Phonological representations capture abstract knowledge of phonological probabilities and phonological properties of words, and this may be tested using pseudo-word stimuli too. In fact, a commonly used task of phonological memory, and one that is argued to be strongly affected by phonological representations, is a task that requires the repetition of nonwords (Gathercole, 1995). Nonword repetition was found to be affected by language-specific lexical factors, or *WORDLIKENESS*, operationalized as the extent to which a nonword is similar to a real word in lexical parameters: phonotactic probability, syllabic length, morphological structure, and stress patterns (Gathercole, 1995; Gathercole & Baddeley, 1990). These findings endorse the role of phonological representations in verbal memory tasks and are in keeping with Baddeley's (2003) multi-componential model of working memory which posits a positive influence of phonological representational quality in long-term memory on phonological processing in working memory.

Because phonological representations are extracted from the lexicon, they show a significant correlation with vocabulary size, both in L1 (e.g., Conti-Ramsden, 2003; Hoff, Core, & Bridges, 2008; Rispen & Baker, 2012) and in L2 (Dufva & Voeten, 1999; Masoura & Gathercole, 1999; Schwartz, Yeh, & Shaw, 2008). Relatedly, children who struggle with vocabulary acquisition, such as children with developmental language disorders (DLD), show a strong nonword repetition deficit (Gathercole, 1995) and are more strongly affected by *WORDLIKENESS* than typically developing children, even when degree of language exposure is similar (Armon-Lotem, 2017). This suggests a deficit in phonological representations in these children. Children with developmental reading disorder also show a phonological representations deficit (Hansen & Bowey, 1994; Katz, 1986; Snowling, Stackhouse, & Rack 1986; Stackhouse & Wells, 1997). It has been argued by several researchers that lack of distinctness and/or segmental specificity in the representations supporting spoken word recognition and production might underlie the phonological processing and reading difficulties observed in individuals with reading disability (Adlard & Hazan, 1998; Elbro, 1998; Goswami, 2000; Snowling, 2000). Evidence in support of this argument is offered by Elbro *et al.* (1998), who show that a measure of the distinctness of phonological representations taken in kindergarten is a particularly sensitive predictor of poor phoneme awareness two years later in children of dyslexic parents. In the same way, Swan and Goswami (1997a) demonstrate a picture naming deficit in developmental dyslexia, implying difficulty in the underlying phonological representations of words. Moreover, they show that when phonological representational quality is taken into account some of the phonological awareness deficits observed in dyslexic children disappear (Swan & Goswami, 1997b).

Despite significant progress in our understanding of the nature of phonological representations and their relationship with language and reading development, research in this domain remains lacking in two ways. First, it focuses mainly on impaired populations. Second, it addresses structural linguistic factors to the exclusion of sociolinguistic factors. In effect, two issues remain unaddressed. The first is the nature and development of phonological representations among typically

developing children. The second is the role of sociolinguistic factors that pertain to language exposure and use on phonological representations. Arabic diglossia, in which children simultaneously acquire and use two linguistic systems that vary in their phonological structure – SpA for everyday speech and StA for formal speech and writing – provides a natural setting for testing these questions. The current study explores phonological representations in Arabic diglossia; it tests the impact of the phonological distance between the two language varieties, and of related differences in extent of exposure and use, on phonological representational quality for StA words. This question has important theoretical implications. Furthermore, because young Arabic-speaking children first learn to read in StA, and before they have fully acquired the phonological system of this language variety, the impact of the phonological distance between SpA and StA on phonological representations may turn out to be fundamental for understanding some of the underpinnings of language and reading development and difficulties in Arabic (Saiegh-Haddad, 2017a; Saiegh-Haddad & Everatt, 2017).

Diglossia and phonological processing in Arabic

All Arabic native-speaking children are born into a dual linguistic context called ‘diglossia’ (Ferguson, 1959). In diglossic Arabic, children grow up speaking a local dialect of SpA that they hear at home and in the neighborhood. Yet, at school, around the age of six, they are formally and extensively exposed to (Modern) Standard Arabic as the language of reading and writing, while Spoken Arabic remains the language of informal speech, even inside the classroom (Amara, 1995). Outside the school context, there is a stable co-existence of the two major varieties, each functioning for distinct spheres of social communication: SpA is used by all native speakers – young and old, educated and uneducated – for informal and intimate verbal interaction in the home, at work, and in the community. StA is at least expected to be used for formal oral interactions, such as delivering a speech or a lecture, and is the only variety considered appropriate for writing (however, see Al-Khatib & Sabbah, 2008, Haggan, 2007, and Mostari, 2009, on use of SpA in electronic writing in Arabic). Notwithstanding the fact that extensive exposure to StA begins with the inception of formal schooling and reading instruction in the first grade, Arabic-speaking children are simultaneously exposed to StA: they pray in StA, watch TV programs and dubbed series in StA, read stories (or are read to) in StA, and do all their written school work in StA. Thus, linguistic proficiency in Arabic requires proficiency in both SpA and StA.

StA is a largely uniform code (Holes, 2004). In contrast, spoken Arabic vernaculars are mainly regional varieties and they vary from one country to another and from one city, town, or village to another. Despite the vast linguistic differences between the different vernaculars of SpA (which might lead to linguistic unintelligibility, as between Western and Eastern dialects of Arabic), they are all structurally related to StA (Maamouri, 1998). At the same time, a contrastive linguistic comparison of StA with any specific dialect of SpA always reveals differences in all domains of language, including the phonological, morphological, morphosyntactic, and lexical-semantic domains (Saiegh-Haddad & Henkin-Roitfarb, 2014).

Given the linguistic distance between SpA and StA, any given linguistic unit in Arabic may have any one of three linguistic affiliations: only-StA, only-SpA, or StA-and-SpA. This trichotomy may be applied to all domains of language. In the

domain of phonology, for instance, Arabic phonemes may belong (for speakers of a given vernacular) to one of the following categories: (a) only-SpA phonemes, which are used in a specific SpA vernacular but are not within the phonemic inventory of StA (e.g., the voiceless affricate /č/ used in some rural dialects of Palestinian Arabic used in Israel); (b) only-StA phonemes, which are only standard and not within a given spoken vernacular (e.g., interdental fricatives: voiced /ð/ and voiceless /θ/ in most urban dialects of Palestinian Arabic spoken in the north of Israel); and (c) StA-and-SpA phonemes, which are phonemes used in both SpA and StA. In the domain of phonology, this last category comprises by far the largest set of Arabic phonemes.

Applying the same three-way classification delineated above to the lexical domain, a recent study analyzed a corpus of 4,408 word types derived from a total of 17,499 word tokens collected from five-year-old speakers of Palestinian Arabic in Central Israel (Saiegh-Haddad & Spolsky, 2014). This analysis revealed the following distribution of lexical items: (a) IDENTICAL WORDS, words that have an identical phonological form in SpA and StA (e.g., /*na:m*/ 'sleep', /*ljanu:b*/ 'south', or /*daftar*/ 'notebook'), made up 21.2% of the word types in the children's lexicon – note that StA case and mood inflections were not regarded in this analysis; (b) COGNATE WORDS, words that are phonologically related yet show only partially overlapping phonological forms when used in each of the two varieties (e.g., SpA /*dahab*/ versus StA /*ḏahab*/ 'gold'; SpA /*sama*/ versus StA /*sama:ʔ*/ 'sky'), made up another 40.6%; and (c) UNIQUE SPA WORDS, words that have a unique form in SpA that is completely different from the one used in StA (e.g., SpA /*juzda:n*/ versus StA /*ḥaqi:bah*/ 'bag'; SpA /*hat*/ versus StA /*waḏaʔ*/ 'he put'), made up 38.2% of the total number of word types in the children's lexicon. This study also showed that the majority of cognate words (~42%) depicted one phonological distance parameter between their form in SpA and StA, such as consonant substitution, glottal stop deletion, vowel change, and vowel insertion. This was followed by cognates depicting two phonological distance parameters (~24%) and three phonological distance parameters (~11%), with all other word types depicting more than three phonological distance parameters each making up less than 5% of the total number of cognates.

The results reported above have serious implications for StA language and literacy acquisition in Arabic. This is because they show that only about 20% of the word types in the SpA lexicon of a five-year-old child, who will soon embark on the reading acquisition journey, have identical phonological forms in StA, the language of literacy, whereas almost 80% of the total number of word types in their lexicon have different forms (either completely different – unique word – or partially different – cognate words). Given established evidence for the important role of oral language skills, and of lexical quality in language and reading development, such a massive orality–literacy gap is expected to have a strong impact on children's ability to develop language and literacy skills in StA (Perfetti & Stafura, 2014; Verhoeven *et al.*, 2016).

The study of the impact of the linguistic distance between SpA and StA on language and literacy development in diglossic Arabic is scarce. With a focus on reading, and on phonological skills in particular, Saiegh-Haddad and colleagues (Saiegh-Haddad, 2003, 2004, 2005, 2007a; Saiegh-Haddad, Levin, Hende, & Ziv, 2011; Saiegh-Haddad & Schiff, 2016; Schiff & Saiegh-Haddad, 2017, 2018) tested the impact of the phonological distance between SpA and StA on the acquisition of literacy-related skills in StA, including phonological awareness, pseudo-word decoding, and word reading accuracy

and fluency. These studies showed that the acquisition of reading, and of related phonological processing skills, is impacted by the phonological distance between SpA and StA, both in impaired and in typically developing readers. Research has also endorsed the debilitating impact of phonological distance on letter naming (Asaad & Eviatar, 2013) and on phonological memory in typically developing and in children with developmental language disorder (Saiegh-Haddad & Ghawi-Dakwar, 2017). Altogether, the results from these studies suggest difficulty among Arabic-speaking children in developing high-quality phonological representations for StA words.

To sum up, research shows that phonological representations are gradient, and that their construction and acquisition are affected by structural linguistic factors (e.g., neighborhood density) as well as by functional factors (e.g., frequency). Moreover, phonological representations affect the acquisition of language and literacy skills. These findings are fundamental when conceived within the framework of COMPARATIVE LINGUALITY, which situates multilectalism and its impact on language and literacy development on a gradient scale (Grohmann & Kambanaros, 2016; Rowe & Grohmann, 2013). The current study tests the role of phonological distance between the two forms of a word: a dialectal spoken form and a standardized mainly written form, on the development of phonological representations in a bilectal context. The study addresses this question in diglossic Arabic. It tests the impact of the phonological distance between SpA and StA on phonological representations for StA words by asking children in senior kindergarten (one year before first grade), and first-, second-, and sixth-grade children to make pronunciation accuracy judgments on StA words that vary in phonological distance from their forms in SpA. By manipulating the phonological distance/overlap between the form of the word in StA and in SpA, the study allows an exploration of whether the construction of phonological representations for StA words in Arabic is affected by the phonological distance between StA and SpA. Moreover, by testing children at different grade levels, and particularly at the point of transition from kindergarten to the first grade and into the second grade, it is possible to test the developmental trajectory of phonological representations and the extent to which they benefit from extensive exposure to StA and from learning to read in StA. In this context, sixth-graders are included as a comparison group.

It was hypothesized that the larger the distance between the SpA and the StA forms the more difficult it would be for children to make a correct pronunciation accuracy judgment; hence, identical words were expected to yield the highest accuracy rates, followed by cognate words and then by unique words. It was also hypothesized that cognate words would yield variations in representational accuracy that would be commensurate with the degree of phonological distance. Finally, it was predicted that the older the participant the more accurate phonological representations would become, with sixth-grade children faring the highest among all four groups and probably showing ceiling levels of performance.

Method

Participants

The sample for the study consisted of 120 children enrolled in a Christian private school in northern Israel. All children were native speakers of an urban dialect of Palestinian Arabic spoken in the north of Israel and were from a middle socioeconomic

background. No child had reported hearing, language, psychological, behavioral, or neurological problems. Four age-groups were targeted: Senior Kindergarten (N = 30; Mean age = 5.9 years; 15 females); First grade (N = 30; Mean age = 7.2 years; 17 females); Second grade (N = 30; Mean age = 7.7 years; 14 females); and Sixth grade (N = 30; Mean age = 11.4 years, 15 females).

Materials

Item words for the study were selected from a Spoken Arabic corpus collected from five-year-old native speakers of Palestinian Arabic in northern and central Israel as well as from elementary school textbooks (Saiegh-Haddad, 2007b; Saiegh-Haddad & Spolsky, 2014). The following criteria were followed in the selection of words. All items were concrete nouns that could be easily represented in a picture, and they were between 1 and 3 syllables long. All words were within the receptive vocabulary of all children. In order to verify this, a picture-naming vocabulary pre-test was conducted. In this pre-test, each of the candidate StA words was embedded within a simple sentence in StA and was orally presented to the children along with four pictures (correct picture plus 3 distractors). Participants were instructed to point to the picture that best represented the target sentence. For instance, the word *ʔabahl* 'ball' was inserted within the sentence 'the girl is playing with a ball' and was presented together with four pictures, a girl playing with a ball and three other distractors: a girl dancing, a girl playing with blocks, and a girl playing with toys. Recognizing the correct sentence required that children recognize the target word 'ball' which distinguished the target picture from the three distractors. Only those words that were recognized by all children were included in the experiment.

A total of 108 words were selected that belonged to one of the three following word categories: (1) IDENTICAL WORDS (N items = 18, e.g., SpA-StA */baʔʔi:x/* 'watermelon'); (2) UNIQUE STA WORDS (N items = 18, e.g., StA */miðallah/* 'umbrella'; and (3) COGNATE WORDS (N items = 72). Cognate words were classified into four subclasses that varied in degree of phonological distance/overlap between their SpA and StA forms (N items = 18 per category): (a) 1-phoneme distant cognates; within this category, words were further divided into two types: 1-vowel distant cognates (e.g., StA */fuʔu:r/*; SpA */ftu:r/* 'breakfast'; vowel deletion) and 1-consonant distant cognates (e.g., StA */qalam/*; SpA */ʔalam/* 'pen'; consonant substitution); (b) 2-phoneme distant cognates: a vowel and a consonant (e.g., StA */miqlama/*, SpA */miʔlami/* 'pencil bag'; (d) more than 2-phoneme distant cognates (e.g., StA */ʔa:ʔiral/*; SpA */ʔayya:ra/* 'airplane'). Modelled after similar methodological procedures (Brenders, Hell, & Dijkstra, 2011; Cristoffaninia, Kirsnera, & Milecha, 1986), half of the stimuli were used in their intact form (N items = 54) while the second half were subjected to a phoneme substitution procedure yielding a pseudo-word (N items = 54). Four criteria were considered in selecting the location of the phonological substitution procedure and the phonological profile of the two phonemes to be exchanged: (a) the target phoneme to be replaced was not a phonological distance parameter in the word, and therefore replacing it by another phoneme did not alter the phonological distance of the StA word from its SpA form. For instance, if the StA word varied from its SpA form in one consonant (e.g., StA */muði:ʔ/*; SpA */muzi:ʔ/* 'newscaster'), consonant substitution could not apply to the variable consonant /ð/ but to another consonant within the word; (b) the target phoneme to be replaced was a consonant, not a vowel; (c) the target phoneme to be replaced was a medial consonant; not the initial

nor the final; and (d) the phoneme which would replace the target original phoneme was also a consonant and was different from the original phoneme in only one feature of articulation: voicing, place of articulation, or manner of articulation. It is important to point out that the nouns selected for the study were not deverbal nouns (not derived by a root), so affiliation of the consonant with the root morpheme of the word was not relevant for the phoneme substitution procedure. It is also important to add that, whereas words in the first two categories of cognates (1- and 2-phoneme distant cognates) involved paradigmatic phonological differences between SpA and StA (e.g., vowel insertion, vowel change, consonant substitution, word-final glottal stop deletion, word-initial glottal stop deletion, etc.), words in the third category (more than 2-phoneme distant cognates) involved mainly words that depicted word-specific phonological differences. Because all words targeted in the study were within the receptive vocabulary of children, and because the phoneme substitution procedure did not alter the phonological distance of the StA word from its SpA form, these words too were deemed appropriate for inclusion in this study. All pseudo-words that resulted from the phoneme substitution procedure were pronounceable strings of sound that abided by the phonotactics of StA. See 'Appendix' for examples of items.

Procedure

The experiment used a computerized, picture-supported pronunciation accuracy judgment task (Anthony, Williams, Aghara, Dunkelberger, Novak, & Mukherjee, 2010). The task was designed using the Opensesame software (Mathôt, Schreij, & Theeuwes, 2012). Each of the target StA words was presented on a computer screen together with a picture illustrating the object noun it represented, and a sound file of its correct pronunciation if it was a real word and its twisted form if it was a pseudo-word. Participants were asked to decide if the pronunciation of the target StA word was correct or not. If correct, s/he was asked to press a green bar. If incorrect, s/he was asked to press a red bar. A picture of a broadcaster holding a microphone always appeared on the left-hand top corner of the screen to remind participants that the experiment involved StA words. The participant was told that this broadcaster (pointing to the picture on the left-hand top corner of the computer screen) did not speak StA properly, and therefore her/his task was to judge if the pronunciation of the word represented in the picture was accurate or not.

Data was collected by the second author, a graduate student of speech and language pathology and a native speaker of the dialect targeted in this study. Authorization for data collection was obtained from the office of the chief scientist of the Ministry of Education. Written parental consent was obtained from all children participating in the study. The child received one score if s/he pressed the correct bar – green for correct and red for incorrect – and a zero score if s/he pressed the wrong bar. ALPHA CRONBACH reliability across all tested words was $\alpha = .80$.

Results

The study tested the effect of the phonological distance between SpA and StA on phonological representations by targeting three types of StA words – identical, cognate, and unique – in four grade levels: kindergarten, first, second, and sixth grade. Furthermore, we tested phonological representations for different types of

cognate words that varied in degree of phonological distance between their forms in StA and SpA. Table 1 presents summary statistics of pronunciation accuracy judgments by phonological distance and grade.

In order to assess the word's phonological distance effect and the individual child's effect on each pronunciation accuracy judgment, we used a mixed-effects model within a cross-classification approach with subjects and items as crossed random effects (Baayen, Davidson, & Bates, 2008). This multilevel analysis assumes that children differ from each other over all pronunciation accuracy judgments they perform, and that words differ from each other across the children who performed the judgment tasks across all words. The smallest unit of measurement in this analysis is a single pronunciation accuracy judgment. The pool of data for the study consisted of 12,960 measurements (120 children \times 108 words). To determine the grade-level effect and the word-type effect as independent variables, Multilevel Binary Logistic Regression analysis was used. Our first modeling attempt was to assess intra-class correlation (ICC) indices. Two ICC measurements were calculated, one for the children's level (subjects), and the other for the word's level variance (words).

Table 2 provides the ICC modeling results. The second panel of Table 2 shows an inter-class correlation of .05 for the children's level variance (ICC subjects) and .33 for the word level variance (ICC words). This means significant variation across words and children in the propensity to make a correct judgment. Model 1 is the unconditional model, that is the intercept not explained by word type and grade. Model 2 is the fixed-effects part which shows the overall grade and word effects compared to a reference category. Model 2 was performed twice, once with identical words as a reference category and once with unique words as a reference category. In both model runs, the reference-grade category is kindergarten. In addition, the variance component section in Table 2 (second panel) shows to what extent children differ from each other and to what extent words differ from each other. Model 2 shows significant variation across children and words (0.14, $p < .001$; 1.34, $p < .001$, respectively).

As Table 2 shows, two main effects were examined to understand differences in successful performance: grade (L2: Grade level) and word type (L2: Word level). Effects (the unstandardized b coefficients) are the log odds, or the log ratio of the probability of successful judgment by a specific category versus that probability of the reference category. The results for Model 2 represented in Table 2 regarding grade effect show that while first-graders showed no difference in pronunciation accuracy judgments in comparison to kindergarten children ($b = 1.14$, $p > .05$), second-graders and sixth-graders did ($b = 0.39$, $p < .01$; $b = 0.87$, $p < .001$, respectively). The results of the word-type effect show that identical words were correctly judged more easily than unique words, 1-consonant distant cognates (category 3), 2-phoneme distant cognates (category 4), and more than 2-phoneme distant cognates (category 5) ($b = -1.82$, $p < .001$; $b = -1.19$, $p < .01$; $b = -1.51$, $p < .01$; $b = -1.67$, $p < .001$, respectively). However, identical words were not more significantly easily judged than 1-vowel distant cognates (category 2). At the same time, unique words were judged significantly more easily than identical words ($b = 1.82$, $p < .001$), and 1-vowel distant cognates (category 2) ($b = 1.45$, $p < .001$). Yet, they were not significantly different in pronunciation accuracy judgment from any of the other three types of cognates (category 3, 4, 5), not even those which were distant by just one consonantal phoneme.

Next, we asked whether differences in phonological representations for the different types of words were different in the different grades. To achieve this, we expanded the

Table 1. Summary Statistics of Pronunciation Accuracy Judgments by Phonological Distance and Grade (N = 120)

Word Type	Grade				
	Kindergarten N = 30	Gr. 1 N = 30	Gr. 2 N = 30	Gr.6 N = 30	All N = 120
Identical words	89.81 (1.04)	92.40 (1.04)	92.96 (1.04)	95.18 (1.04)	92.59 (0.52)
Cognate words: 1-phoneme distant cognates: vowel	88.51 (1.29)	87.40 (1.29)	90.55 (1.29)	90.55 (1.29)	89.26 (0.65)
Cognate words: 1-phoneme distant cognates: consonant	77.22 (1.55)	78.51 (1.55)	83.71 (1.55)	89.63 (1.55)	85.27 (0.77)
Cognate words: 2-phoneme distant cognates	67.40 (1.60)	67.29 (1.60)	75.55 (1.60)	86.48 (1.60)	76.43 (0.80)
Cognate words: More than 2-phoneme distant cognates	72.59 (1.97)	70.11 (1.97)	75.92 (1.97)	82.40 (1.97)	75.26 (0.98)
Cognate words: all	76.43 (6.91)	78.08 (5.88)	81.43 (6.09)	86.26 (5.08)	81.55 (7.26)
Unique words	67.22 (2.05)	69.63 (2.05)	73.33 (2.05)	76.66 (2.05)	71.71 (1.02)
Total	77.13 (1.00)	79.06 (1.00)	82.00 (1.00)	86.82 (1.00)	81.25 (0.50)

Table 2. Cross-Classified Multilevel Binary Logistic Regression: Testing Pronunciation Accuracy Judgment by Grade and Word Level

		Model 1	Model 2
Fixed effects			
	Intercept	1.93*** (.14)	2.69*** (.31)
L2: grade level	G1 vs. Kindergarten		1.14 (.12)
	G2 vs. Kindergarten		0.39** (.12)
	G6 vs. Kindergarten		0.87*** (.12)
L2: word level	Category 1: Unique vs. Identical		-1.82*** (.41)
	Category 2: Cognate-vowel vs. Identical		-0.37 (.42)
	Category 3: Cognate-consonant vs. Identical		-1.19** (.41)
	Category 4: Cognate-2-phoneme vs. Identical		-1.51*** (.41)
	Category 5: Cognate-more than 2-phonemes vs. Identical		-1.67*** (.41)
	Category 1: Identical vs. Unique		1.82*** (.41)
	Category 2: Cognate-vowel vs. Unique		1.45*** (.40)
	Category 3: Cognate-consonant vs. Unique		0.63 (.410)
	Category 4: Cognate-2-phoneme vs. Unique		0.31 (.40)
	Category 5: Cognate-more than 2-phonemes vs. Unique		0.15 (.40)
Variance components			
Level 2 – grades	Grade intercept $\sigma_{\text{subject}}^2$	0.24*** (.49)	0.14*** (.14)
Level 2 – words	Word intercept σ_{word}^2	1.75*** (1.32)	1.34*** (1.16)
ICC subjects		0.05	–
ICC words		0.33	–
Pseudo R ²		–	0.10

Notes. Standard errors are in parentheses for fixed effects and Standard deviation for random parameters; ~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

analysis to test all word categories one against the other, for each grade separately. Results are shown in Table 3.

As can be seen from Table 3, within the group of kindergarteners, identical words were easier to judge than unique words ($b = -1.68$, $p < .001$), and they were easier than all other types of cognates, except 1-vowel distant cognates (category 2) ($b = -1.11$, $p < .01$, $b = -1.64$, $p < .001$, $b = -1.44$, $p < .001$). At the same time, 1-vowel distant cognates were easier than unique words ($b = 1.52$, $p < .001$), but no other significant differences were found between unique words and the other word categories. The same pattern was replicated among first- and second-graders, but it changed among sixth-graders. Within the sixth-grade group, cognates distant by one

Table 3. Cross-Classified Multilevel Binary Logistic Regression Testing Pronunciation Accuracy Judgment by Word Level for Each Grade Separately

Kindergarten	Intercept	2.55*** (.32)	
	Category 1: Unique vs. Identical	-1.68*** (.42)	
	Category 2: Cognate-vowel vs. Identical	-0.16 (.44)	
	Category 3: Cognate-consonant vs. Identical	-1.11** (.42)	
	Category 4: Cognate-2-phonemes vs. Identical	-1.64*** (.42)	
	Category 5: Cognate-more than 2-phonemes vs. Identical	-1.44*** (.42)	
	Intercept	0.87** (.28)	
	Category 1: Identical vs. Unique	1.68*** (.42)	
	Category 2: Cognate-vowel vs. Unique	1.52*** (.41)	
	Category 3: Cognate-consonant vs. Unique	0.57 (.40)	
	Category 4: Cognate-2-phonemes vs. Unique	0.04 (.39)	
	Category 5: Cognate-more than 2-phonemes vs. Unique	0.24 (.39)	
	Grade 1	Intercept	2.82*** (.32)
		Category 1: Unique vs. Identical	-1.80*** (.42)
		Category 2: Cognate-vowel vs. Identical	-0.64 (.43)
Category 3: Cognate-consonant vs. Identical		-1.36** (.42)	
Category 4: Cognate-2-phonemes vs. Identical		-1.39** (.42)	
Category 5: Cognate-more than 2-phonemes vs. Identical		-1.79*** (.42)	
Intercept		1.02*** (.28)	
Category 1: Identical vs. Unique		1.80*** (.42)	
Category 2: Cognate-vowel vs. Unique		1.16** (.40)	
Category 3: Cognate-consonant vs. Unique		0.45 (.39)	
Category 4: Cognate-2-phonemes vs. Unique		0.41 (.39)	
Category 5: Cognate-more than 2-phonemes vs. Unique		0.01 (.38)	
Grade 2		Intercept	3.01*** (.39)
		Category 1: Unique vs. Identical	-1.68** (.51)
		Category 2: Cognate-vowel vs. Identical	-0.15 (.54)
	Category 3: Cognate-consonant vs. Identical	-1.08* (.52)	
	Category 4: Cognate-2-phonemes vs. Identical	-1.53** (.52)	
	Category 5: Cognate-more than 2-phonemes vs. Identical	-1.62** (.51)	
	Intercept	1.33*** (.35)	
	Category 1: Identical vs. Unique	1.68** (.51)	
	Category 2: Cognate-vowel vs. Unique	1.53** (.51)	
	Category 3: Cognate-consonant vs. Unique	0.60 (.49)	
	Category 4: Cognate-2-phonemes vs. Unique	0.16 (.49)	

	Category 5: Cognate-more than 2-phonemes vs. Unique	0.06 (.48)
Grade 6	Intercept	3.59*** (.44)
	Category 1: Unique vs. Identical	-2.02*** (.56)
	Category 2: Cognate-vowel vs. Identical	-0.73 (.58)
	Category 3: Cognate-consonant vs. Identical	-0.84 (.58)
	Category 4: Cognate-2-phonemes vs. Identical	-1.04 (.57)
	Category 5: Cognate-more than 2-phonemes vs. Identical	-1.54** (.56)
	Intercept	1.57*** (.37)
	Category 1: Identical vs. Unique	2.02*** (.56)
	Category 2: Cognate-vowel vs. Unique	1.28* (.53)
	Category 3: Cognate-consonant vs. Unique	1.17* (.53)
	Category 4: Cognate-2-phonemes vs. Unique	0.98 (.52)
	Category 5: Cognate-more than 2-phonemes vs. Unique	0.47 (.51)

Notes. Standard errors are in parentheses for fixed effects and Standard deviation for random parameters; p-value signs: * $p < .05$, ** $p < .01$, *** $p < .001$.

phoneme (vowel or consonant) and by two phonemes yielded similar accuracy scores to identical words (category 2, 3, 4). However, unique words and more than 2-phoneme distant cognates were significantly more difficult than identical words ($b = -2.02$, $p < .001$; $b = -1.54$, $p < .01$, respectively). At the same time, only 1-phoneme distant cognates (1 vowel or 1 consonant) were significantly easier than unique words ($b = 1.17$, $p < .05$). The rest of the cognates were not significantly different from unique words even in this group of sixth-graders.

Discussion

The current study had a two-fold objective. First, to investigate whether phonological distance between word forms in SpA and StA, which is a predominant phenomenon in Arabic diglossia (Ferguson, 1959) and a prominent feature of child Arabic (Saiegh-Haddad & Spolsky, 2014), impacts the ability of young children to establish and develop accurate phonological representations for StA words, the language of literacy. Second, to probe whether schooling, and the concomitant exposure to Standard Arabic and to literacy, results in a significant refinement of phonological representations among children. The study targeted three classes of StA words. The first is identical words, which are the closest or the least distant from their form in SpA because they keep an almost identical form in SpA and StA (e.g., /na:m/ 'sleep' or /daftar/ 'notebook'). The second category is cognate words, which are used in both SpA and StA but which have a different surface phonological form in each of them. Words in this class are very heterogeneous and vary in degree of distance from SpA. Four categories of cognate words were targeted: 1-phoneme distant cognates; these were further classified into 1-vowel distant cognates and 1-consonant distant cognates, and 2-phoneme distant cognates (a consonant and a vowel). In both cases, the distance between the StA and the SpA forms of cognates was paradigmatic and computable via rule-based phonological transformations, such as

vowel insertion, vowel change, glottal stop deletion, consonant substitution, etc. The third class of cognates was more than 2-phoneme distant cognates, and included words depicting word-specific alterations (e.g., StA /riʒl/; SpA /ʔiʒir 'leg'). The last category was unique words, whose StA phonological form was unrelated and completely different from its form in SpA (StA /miðallah/; SpA /ʃamsiyyih/ 'umbrella').

The first question addressed in this study was whether words that vary in degree of phonological distance/overlap also vary in quality of phonological representation, despite being equal in familiarity; note that all words used in the study were within the receptive vocabulary of children. Using children's ability to judge if the pronunciation of a picture-supported StA word was accurate or not, the results showed that phonological distance had a significant and a very strong impact on phonological representational quality. Moreover, this effect was not limited to young kindergarten children, but extended to older first- and second-graders, and was significant even among the sixth-grade comparison group. This finding is compelling because it shows that, even though children were familiar with the word, they still found it difficult to judge if pronunciation was accurate when the phonological form of the word was distant or only partially overlapping with its form in their SpA dialect. Furthermore, it was found that, in general, the more distant the word was, the more difficult the task turned out to be, with identical words producing the highest accuracy scores, unique words producing the lowest scores, and cognate words mostly faring in between. Similarly, the more similar and overlapping the StA word was with the children's SpA dialectal form, the easier it was for them to correctly judge if it was accurately pronounced or not. These findings imply that phonological distance between the StA of the word and its form in the spoken dialect of children interferes with children's ability to encode an accurate phonological representation, even when the word is within the receptive vocabulary of the child. This finding supports the separability of lemma activation processes (including activation of the semantic features of the word) from phonological encoding (Levelt, 1989), and it suggests that phonological distance may be a specific constraint on the phonological encoding process. Alternatively, phonological overlap between the StA form of the word and its form in the spoken dialect appears to bootstrap children's phonological encoding, resulting in higher pronunciation accuracy judgment performance.

These results have important theoretical implications for phonological representations in general, and for the factors that impact lexical specificity and children's ability to establish and develop high-quality phonological representations. One important factor that the current results demonstrate is phonological distance/overlap between the form of the word in the spoken language, or dialect, and its form in the language of formal discourse and literacy. This factor is fundamental when conceived within the framework of COMPARATIVE LINGUALITY, and its effect on language development and metalinguistic skills in bilingual and bicultural children (Grohmann & Kambanaros, 2016; Rowe & Grohmann, 2013). This is because it implies that linguistic distance is not a monolithic construct but is a gradable one, and that its impact on processing is commensurate with degree of distance. Moreover, these results show that the impact of linguistic distance on processing is quantifiable and may, hence, be reliably measured (Kambanaros, Michaelides, & Grohmann, 2017).

The impact of the phonological distance between SpA and StA on phonological representational quality that this study has demonstrated supports earlier evidence

endorsing the impact of phonological distance on Arabic-speaking children's phonological awareness (Saiegh-Haddad, 2003, 2004, 2007a; Saiegh-Haddad *et al.*, 2011), letter-naming (Assad & Eviatar, 2013), pseudo-word decoding (Saiegh-Haddad, 2003, 2005), and word reading (Saiegh-Haddad & Schiff, 2016; Schiff & Saiegh-Haddad, 2018). It also aligns with evidence for the effect of phonological distance on word decoding observed among disabled readers (Schiff & Saiegh-Haddad, 2017) and on typically developing and language impaired children's phonological processing in working memory (Saiegh-Haddad & Ghawi-Dakwar, 2017). Given that phonological processing and reading development are predicated on phonological representations (Elbro, 1998; Goswami, 2000; Katz, 1986; Swan & Goswami, 1997a, 1997b), the current results suggest that the phonological processing and reading difficulties encountered by Arabic native-speaking children may be, at least partly, attributed to difficulty in constructing accurate phonological representations for StA lexical items in long-term memory. This idea is captured by MAWRID, a Model of Arabic Word Reading in Development (Saiegh-Haddad, 2017a), which argues that diglossia and the linguistic distance between SpA and StA governs the acquisition of basic reading skills across development (Saiegh-Haddad, 2017b).

The observed impact of phonological distance/overlap on phonological representations has important educational implications for language and reading instruction too, as well as for assessment and material development for Arabic-speaking learners. If the phonological distance between the StA form of the word, which children encounter in reading, and its form in SpA, the form children use in everyday speech, impacts phonological representations, and in turn probably also phonological processing and decoding, and if the more distant the word is the more difficult it is for her to establish an accurate representation for it, then reading instruction as well as assessment should take this factor into account (Saiegh-Haddad & Everatt, 2017). This principle of gradability and control over the 'what' (which words to use) and 'when' (when to introduce different words) questions in reading curriculum development constitutes a core principle of the Exposure-through-Reading Program (Saiegh-Haddad & Spolsky, 2014), which is proposed as a way of bridging the gap between the language of the child and the language of the book in teaching reading in diglossic Arabic (Saiegh-Haddad & Everatt, 2017).

For instance, instruction in phonemic awareness and decoding should probably start with StA words that keep an identical phonological form in the SpA dialect spoken by the children, and then progress gradually to non-overlapping words. This is because phonological awareness and phonological recoding involve operations on a phonological representation, and the extent to which this representation is high in quality would impact phonological analysis (Saiegh-Haddad, 2017c). Furthermore, instruction should devote particular attention to cognate words. Cognate words are less accurately represented than identical words because they are only semi-familiar, and their processing might involve an activation and competition of two phonological forms – a standard and a spoken form – hence the lower pronunciation accuracy judgments. Yet cognate words are easier than unique words, probably because they are only semi-novel and the overlap between the two forms of the word might bootstrap children's ability to encode an accurate phonological representation. Future research should further explore the role of distance/proximity in children's phonological representation and processing to shed light on whether partially overlapping cognate words would always yield a less accurate representation than completely novel unique words. It might be reasonable to argue that a unique StA

word that does not encode a StA phoneme might be easier to process than a StA cognate word that has a competing SpA form in the lexicon of children. Note that the unique words used in the current study were a combination of unique words encoding StA phonemes and those not encoding StA phonemes. Earlier research has shown that unique words that do not encode StA phonemes are easier to process in memory (using a nonword repetition task) than unique words that do not encode StA phonemes (Saiegh-Haddad & Ghawi-Dakwar, 2017). This factor may turn out to be important in clarifying some of the underlying differences in phonological representation between cognates and different types of unique words, and is a question is for future research to pursue.

Another question that the current study tested pertained to the developmental trajectory of phonological representations for the different types of words, and specifically at the transitional point from preliteracy in kindergarten to literacy in the first grade. The results of the study showed that the impact of phonological distance on phonological representations is not transient, and neither is it a characteristic of the lexicons of young preliterate children. Rather, it is persistent and long-standing and it extends to literate children in elementary school, as well as to older children in middle school. At the same time, the results support a developmental progression in phonological representations with second- and sixth-graders performing generally higher than kindergarteners and first-graders, yet no developmental difference was noted between the kindergarten and the first-grade sample. These findings endorse a beneficial effect of oral exposure to StA and of literacy in StA at school on quality of phonological representations. At the same time, the finding that first-grade children were not different from kindergarten children is puzzling, especially when evaluated in light of earlier findings in other languages which indicate an upsurge in phonological representational accuracy with the inception of literacy in the first grade (de Gelder & Vroomen, 1991; Goswami *et al.*, 1999; Swan & Goswami, 1997a). This unexpected finding might reflect cross-linguistic differences in the challenges that confront children in constructing phonological representations in different languages. As such, it might imply that the challenge which diglossia poses to the construction of high-quality phonological representations might supersede the impact of other challenges undermining this undertaking in other languages, such as the encoding of consonantal clusters, for instance. Two features of Arabic are in order: the simple syllabic structure of Arabic and the transparent orthography of vowel Arabic, the orthography children learn to read in (Saiegh-Haddad & Henkin-Roitfarb, 2014). It may be argued that, given the simple syllabic structure of StA, with the absence of complex consonantal clusters, the contribution of exposure to the shallow orthography of vowel Arabic to phonological representations in the first grade is significantly reduced, giving way to extended oral exposure to StA as the main factor. To address this conjecture, future research should address the question of the relative effect of oral exposure to StA vis-à-vis exposure to reading in shallow vowel Arabic orthography on quality of phonological representations.

Besides a general effect of phonological distance on phonological representations reflected as greatest facility with pronunciation accuracy judgment for identical words, followed by cognates and then unique words, the results of the study showed an interesting pattern of results: 1-vowel distant cognates were indistinguishable from identical words, whereas the rest of the cognates, namely categories 1–3 (1-consonant distant, 2-phoneme distant, and more than 2-phoneme distant cognates) were all indistinguishable from unique words (Table 2). This pattern of results was unaltered

in the three younger groups: kindergarten, first grade, and second grade (Table 3). This finding is outstanding and it shows that cognate words, which are usually regarded as easy to recognize by children because they are half similar and their forms in StA and SpA are overlapping, are in fact not as accurately represented and may therefore not be as easily accessible or retrievable by children as they are often assumed to be. This has important implications for the acquisition of decoding skills in young children (Verhoeven *et al.*, 2016) and of reading comprehension later on (Nouwens *et al.*, 2018). Note that cognate words that vary in just one or two paradigmatic rule-based phonological change parameters, such as a consonant substitution (StA /*dahab*/; SpA /*dahab*/ 'gold') or word-final glottal stop deletion plus preceding vowel reduction (e.g., StA /*sama:ʔ*/; SpA /*samal*/ 'sky'), were more difficult to judge than identical words and, unexpectedly, not significantly easier than unique words. This finding should be taken very seriously in Arabic language and reading instruction and assessment, especially in light of widespread folk wisdom usually endorsed by educational policy-makers, according to whom, if a StA word also exists in SpA (i.e., is a cognate) or "has its origin in StA" as some put it, it should not cause any difficulty for children. This finding should be equally heeded in Arabic psycholinguistic research, especially when the research design relies heavily on word-based manipulations.

The finding that 1-vowel distant cognates aligned with identical words and were significantly different from 1-consonant distant cognates might also imply a different status of vowels versus consonants in the Arabic lexicon. The results showed that when StA cognate words varied from their SpA form in just one vowel, they were indistinguishable from identical words. This was not the case for a StA cognate that varied from its SpA form in one consonant. This suggests greater salience of Arabic consonants over vowels, and may be attributed to the fact that Arabic is a consonantal root-based language in which the meaning of the word is carried mainly on its consonantal structure (Saiegh-Haddad & Henkin-Roitfarb, 2014). In contrast, vowels are mostly part of the word pattern template, which encodes the word's grammatical properties, and these tend to vary in different dialects, and even within StA for variants of the same template. Evidence in support of this hypothesis comes from research showing that awareness of the consonantal root morpheme in Arabic develops early, and before awareness of the word-pattern morpheme (Taha & Saiegh-Haddad, 2017). Moreover, research shows that morphological processing using the consonantal root is observed in spelling and reading in young children (Saiegh-Haddad & Taha, 2017), and is more common among young children than morphological processing using the word-pattern (Saiegh-Haddad, 2013). To directly test the hypothesis that consonants may be psycholinguistically more salient than vowels, future research should directly compare phonological awareness for Arabic vowels versus consonants (Saiegh-Haddad, 2017b).

Thus, cognates were found to behave in different ways depending on the type of distance, with 1-vowel distant cognates being indistinguishable from identical words and with the rest of the cognates indistinguishable from unique words. This pattern was replicated in the three younger groups of children and was only different in the sixth grade (Table 3). In the latter group, all cognate words except category 4 (more than 2-phoneme distant cognates) aligned with identical words. However, category 4 of cognates aligned with unique words. This observation is compelling, especially in light of the fact that all other cognates (category 1–3) depict a phonological distance between StA and SpA that is regular, paradigmatic, and rule-based, i.e., that the SpA

form may be computed from the StA form by applying a set of phonological processes, whereas category 4 (more than 2-phoneme distant) depict word-specific phonological alterations. This finding implies that children might find it easier to develop accurate phonological representation for StA words whose distance from SpA is paradigmatic and whose StA forms are regularly related to their form in SpA, yet they only appear to show this advantage in the sixth grade. In contrast, younger children find all cognate words, even those that regularly alternate with their SpA form in just one consonant, indistinguishable from unique words. This advantage observed in sixth-graders as against all the other three younger groups might reflect the combined effect of extended exposure to StA (6 years as against 1–3 years in kindergarten and second grade, respectively), as well as extended experience with literacy in the transparent orthography of Arabic. Extended exposure to StA enhances children's ability to extract statistical patterns from the language, including patterns that govern the relationship between StA forms and their paradigmatically related SpA parallel forms, and this could in turn lead to greater grasp and awareness of these transformations as reflected in higher pronunciation accuracy judgments. Research supports a fast increase in linguistic awareness – phonological and morphological – during the first six years of literacy instruction among Arabic native-speaking children (Saiegh-Haddad & Taha, 2017). One outcome of enhanced phonological and morphological awareness may be increased awareness of the phonological and morphophonological differences and similarities between lexical items as they are used in SpA and StA, and this might bootstrap phonological encoding in memory, as reflected in performance on our task. A second factor could be extended literacy experience with the orthography of Arabic in general and with unvoiced Arabic in particular. Children are taught how to read using the voiced Arabic orthography, and they only start reading in unvoiced Arabic approximately around the fourth grade. Developing reading fluency in unvoiced Arabic appears to be particularly dependent on morphological processing skills, in addition to phonological skills (Saiegh-Haddad & Taha, 2017; Schiff & Saiegh-Haddad, 2018). This is because the unvoiced word is morphologically transparent and because morphological awareness (especially pattern awareness) enables the reader to recover the missing vocalic information that is mapped by the diacritics (Saiegh-Haddad, 2017a). It may be argued that enhanced morphological processing in reading unvoiced Arabic contributes to enhanced awareness of the linguistic relatedness of SpA and StA lexical items. This may be especially so as phonological distance/overlap may be obscured in speech yet preserved in the spellings of words. It is noteworthy in this respect that Arabic spelling is loyal to morphological structure (root and word pattern) and, given strong velarization assimilation processes operating on the phonological form of the word, yet not represented in its spelling (Saiegh-Haddad, 2013), morphological processing might be particularly conducive to awareness of the linguistic distance/overlap between SpA and StA forms of words. This question is for future research to explore. Future research should also test children's knowledge of SpA–StA transformations and their ability to apply these transformations to unfamiliar or pseudo-word stimuli, as well as the relationship between this ability, phonological representations, and statistical learning. This avenue may turn out to be critical for understanding individual variations in children's ability to develop a range of language and literacy skills in Arabic diglossia (Saiegh-Haddad, 2017b). Future research should also test the role of awareness of cognates and of cognate training in bootstrapping vocabulary acquisition. Also,

whether cognate training would have the same impact on phonological representations in bilingual contexts (e.g., Kambanaros *et al.*, 2017).

Conclusion and limitations

The results of the current study demonstrate the psychological reality of phonological distance in Arabic diglossia, reflected as differences in the phonological representational quality for different StA words in the lexicons of Arabic native-speaking children. The results show that the accuracy with which the phonological form of StA words is represented may be predicted by the relative distance or overlap of the StA word from its form in SpA, with the least distant words, the identical words which preserve an identical phonological form in SpA and StA, showing the most accurate representations. In contrast, cognate words, including those that are distant by a single consonant or more, show lower level of representational accuracy, which may be as low as the representation of unique words which have two completely different forms in SpA and StA. Interestingly, we also found that enhancement of phonological representations in Arabic is very slow and not evident in the early school grades, and certainly not by the transition from orality to literacy in the first grade. These results have significant and far-reaching implications for language and literacy acquisition and education in diglossic Arabic.

Two critical limitations on the generalizability of the results of this study are warranted: sample size and socio-cultural context. The findings we report in this study are based on a sample of 120 children. Future research should aim for a larger sample. Moreover, the results are based on Arabic native-speaking children living in Israel. These children are monolingual native speakers of Arabic and they attend Arabic-medium instruction schools, in which Hebrew is taught as a second language starting in the third grade only. Nonetheless, the presence of Hebrew in the linguistic landscape of Israel and in the linguistic repertoire of young children in Israel is not to be overlooked. Future research is needed that replicates the current study among speakers of Arabic living in other regions in the Arabic-speaking world.

References

- Adlard, A., & Hazan, V. (1998). Speech perception in children with specific reading difficulties (dyslexia). *Quarterly Journal of Experimental Psychology A*, 51, 153–77.
- Al-Khatib, M., & Sabbah, E. (2008). Language choice in mobile text messages among Jordanian university students. *SKY Journal of Linguistics*, 21, 1–27.
- Amara, M. H. (1995). Arabic diglossia in the classroom: assumptions and reality. In S. Izra'el & R. Drory (Eds.), *Israel oriental studies*, vol. 15, *language and culture in the Near East* (pp. 131–42). Leiden: E.J. Brill.
- Anthony, J., Williams, J., Aghara, R., Dunkelberger, M., Novak, B., & Mukherjee, A. (2010). Assessment of individual differences in phonological representation. *Reading and Writing*, 23(8), 969–94.
- Armon-Lotem, S. (2017). Disentangling bilingualism from SLI: dissociating exposure and input. *Bilingualism: Language and Cognition*, 20(1), 33–4.
- Asaad, H., & Eviatar, Z. (2013). The effects of orthographic complexity and diglossia on letter naming in Arabic: developmental changes. *Writing Systems Research*, 5, 156–68.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189–208.
- Brenders, P., Hell, J. G. v., & Dijkstra, T. (2011). Word recognition in child second language learners: evidence from cognates and false friends. *Journal of Experimental Child Psychology*, 109(4), 383–96.

- Carroll, J. M., & Snowling, M. J.** (2001). The effects of global similarity between stimuli on children's judgment of rime and alliteration. *Applied Psycholinguistics*, 22, 327–42.
- Conti-Ramsden, G.** (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, 46, 1029–37.
- Cristoffaninia, P., Kirsner, K., & Milecha, D.** (1986). Bilingual lexical representation: the status of Spanish–English cognates. *Quarterly Journal of Experimental Psychology*, 38(3), 367–93.
- De Cara, B., & Goswami, U.** (2003). Phonological neighborhood density: effects in a rhyme awareness task in five-year-old children. *Journal of Child Language*, 30, 695–710.
- de Gelder, B., & Vroomen, J.** (1991). Phonological deficits: beneath the surface of reading-acquisition problems. *Psychological Research*, 53, 88–97.
- Dollaghan, C., & Campbell, T. F.** (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1136–46.
- Dufva, M., & Voeten, M. J. M.** (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, 20, 329–48.
- Elbro, C.** (1998). When reading is 'readn' or somthn: distinctness of phonological representations of lexical items in normal and disabled readers. *Scandinavian Journal of Psychology*, 39, 149–53.
- Elbro, C., Borström, I., & Petersen, D. K.** (1998). Predicting dyslexia from kindergarten: the importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly*, 33, 36–60.
- Elbro, C., & Jensen, M. N.** (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology*, 46, 375–84.
- Ferguson, C. A.** (1959). Diglossia. *Word*, 14, 47–56.
- Fowler, A. E.** (1991). How early phonological development might set the stage for phoneme awareness. In S. A. B. D. P. Shankweiler (Ed.), *Phonological processes in literacy: a tribute to Isabelle Y. Liberman* (pp. 97–117). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gathercole, S. E.** (1995). Is nonword repetition a test of phonological working memory or long term knowledge? It all depends on the nonwords. *Memory and Cognition*, 23, 83–94.
- Gathercole, S. E., & Baddeley, A. D.** (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29, 336–60.
- Gibson, T. A., Summers, C., Pena, G. D., Dedore, L. M., Gillam, R. B., & Bohman, T. H.** (2015). The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition*, 18, 551–60.
- Goswami, U.** (2000). Phonological representations, reading development and dyslexia: towards a cross-linguistic theoretical framework. *Dyslexia*, 6, 133–51.
- Goswami, U., Schneider, W., & Scheurich, B.** (1999). Picture naming deficits in developmental dyslexia in German. *Developmental Science*, 2, 53–8.
- Gray, S.** (2005). Word learning by preschoolers with specific language impairment: effect of phonological and semantic cues. *Journal of Speech, Language, and Hearing Research*, 48, 1452–67.
- Gray, S., & Brinkley, S.** (2011). Fast mapping and word learning by preschoolers with specific language impairment in a supported learning context: effect of encoding cues, phonotactic probability, and object familiarity. *Journal of Speech, Language, and Hearing Research*, 54, 870–84.
- Grohmann, K. K., & Kambanaros, M.** (2016). The gradience of multilingualism in typical and impaired language development: positioning bilingualism within comparative bilingualism. *Frontiers in Psychology*, 7, 37. doi: 10.3389/fpsyg.2016.00037.
- Hagan, M.** (2007). Text messaging in Kuwait: Is the medium the message? *Multilingua*, 26, 427–49.
- Hansen, J., & Bowey, J. A.** (1994). Phonological analysis skills, verbal working memory, and reading ability in second-grade children. *Child Development*, 65, 938–50.
- Hoff, E., Core, C., & Bridges, K.** (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, 35, 903–16.
- Holes, C.** (2004). *Modern Arabic: structures, functions & varieties*. London: Longman.
- Kambanaros, M., Michaelides, M., & Grohmann, K. K.** (2017). Cross-linguistic transfer effects after phonologically based cognate therapy in a case of multilingual specific language impairment (SLI). *International Journal of Language & Communication Disorders*, 52(3), 270–84.
- Katz, R. B.** (1986). Phonological deficiencies in children with reading disability: evidence from an object-naming task. *Cognition*, 22, 225–57.
- Levett, W. J. M.** (1989). *Speaking*. Cambridge, MA: MIT Press.

- Maamouri, M.** (1998). *Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region*. Paper presented at the The Mediterranean Development Forum, Washington, DC.
- Masoura, V. E., & Gathercole, S. E.** (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology*, 3, 383–8.
- Mathôt, S., Schreij, D., & Theeuwes, J.** (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44, 314–24.
- Metsala, J. L.** (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory and Cognition*, 25, 47–56.
- Metsala, J. L., & Walley, A. C.** (Eds.) (1998). *Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mostari, H. A.** (2009). What do mobiles speak in Algeria? Evidence from language. *Current Issues in Language Planning*, 10, 377–86.
- Nouwens, S., Groen, M. A., Klemans, T., & Verhoeven, L.** (2018). The role of semantic retrieval in children's reading comprehension development in the upper primary grades. *Journal of Research in Reading*, 41(3), 597–614.
- Perfetti, C. A.** (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–83.
- Perfetti, C. A., & Hart, L.** (2001). The lexical bases of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: perspectives on resolving lexical ambiguity* (pp. 67–86). Washington, DC: American Psychological Association.
- Perfetti, C., & Stafura, J.** (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18, 22–37.
- Perfetti, C. A., Wlotko, E. W., & Hart, L. A.** (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1281–92.
- Rispens, J., & Baker, A.** (2012). Nonword repetition: the relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, 55, 683–94.
- Rowe, C., & Grohmann, K. K.** (2013). Discrete bilingualism: towards co-overt prestige and diglossic shift in Cyprus. *International Journal of the Sociology of Language*, 224, 119–42.
- Saiegh-Haddad, E.** (2003). Linguistic distance and initial reading acquisition: the case of Arabic diglossia. *Applied Psycholinguistics*, 24, 431–51.
- Saiegh-Haddad, E.** (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudowords in a diglossic context. *Applied Psycholinguistics*, 25, 495–512.
- Saiegh-Haddad, E.** (2005). Correlates of reading fluency in Arabic: diglossic and orthographic factors. *Reading and Writing: An Interdisciplinary Journal*, 18, 559–82.
- Saiegh-Haddad, E.** (2007a). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics*, 28, 605–25.
- Saiegh-Haddad, E.** (2007b). *Enhancing mother tongue literacy in Arabic. Final report and data base submitted to Had Hanadiv Foundation*. CET (Centre for Educational Technology) research project.
- Saiegh-Haddad, E.** (2013). A tale of one letter: morphological processing in early Arabic spelling. *Writing Systems Research*, 5, 169–88.
- Saiegh-Haddad, E.** (2017a). MAWRID: A Model of Arabic Word Reading in Development. *Journal of Learning Disabilities*. doi:0022219417720460.
- Saiegh-Haddad, E.** (2017b). Learning to read in Arabic. In L. Verhoeven & C. Perfetti (Eds.), *Reading acquisition across languages and writing systems: an international handbook* (pp. 127–54). Cambridge University Press.
- Saiegh-Haddad, E.** (2017c). What is phonological awareness in L2? *Journal of Neurolinguistics*. doi.org/10.1016/j.jneuroling.2017.11.001.
- Saiegh-Haddad, E., & Everatt, J.** (2017). Literacy education in Arabic. In N. Kucirkova, C. Snow, V. Grover, & C. McBride-Chang (Eds.), *The Routledge international handbook of early literacy education* (pp. 185–99). London: Routledge.
- Saiegh-Haddad, E., & Ghawi-Dakwar, O.** (2017). The impact of diglossia on word and nonword repetition in TD and SLI children. *Frontiers in Psychology*. doi:10.3389/fpsyg.2017.02010.

- Saiegh-Haddad, E., & Henkin-Roitfarb, R.** (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad & M. Joshi (Eds.), *Handbook of Arabic literacy: insights and perspectives* (pp. 3–28). Springer: Dordrecht.
- Saiegh Haddad, E., Levin, I., Hende, N., & Ziv, M.** (2011). The Linguistic Affiliation Constraint and phoneme recognition in diglossic Arabic. *Journal of Child Language*, 38, 297–315.
- Saiegh-Haddad, E., & Schiff, R.** (2016). The impact of diglossia on vowelized and unvowelized word reading in Arabic: a developmental study from childhood to adolescence. *Scientific Studies of Reading*, 20, 311–24.
- Saiegh-Haddad, E., & Spolsky, B.** (2014). Acquiring literacy in a diglossic context: problems and prospects. In E. Saiegh-Haddad & M. Joshi (Eds.), *Handbook of Arabic literacy: insights and perspectives* (pp. 225–40). Dordrecht: Springer.
- Saiegh-Haddad, E., & Taha, T.** (2017). The role of phonological and morphological awareness in the early development of word reading and spelling in typical and disabled Arabic readers. *Dyslexia*, 23, 345–71.
- Schiff, R., & Saiegh-Haddad, E.** (2017). When diglossia meets dyslexia: the effect of diglossia on vowelized and unvowelized word reading among native Arabic-speaking dyslexic children. *Reading and Writing*, 30, 1089–113.
- Schiff, R., & Saiegh-Haddad, E.** (2018). Development and relationships between phonological awareness, morphological awareness and word reading in spoken and standard Arabic. *Frontiers in Psychology*. doi.org/10.3389/fpsyg.2018.00356.
- Schwartz, A. I., Yeh, L.-H., & Shaw, M. P.** (2008). Lexical representation of second language words: implications for second language vocabulary acquisition and use. *The Mental Lexicon*, 3, 309–24.
- Snowling, M. J.** (2000). *Dyslexia*, 2nd ed. Oxford: Blackwell.
- Snowling, M., Stackhouse, J., & Rack, J.** (1986). Phonological dyslexia and dysgraphia – a developmental analysis. *Cognitive Neuropsychology*, 3(3), 309–39.
- Stackhouse, J., & Wells, B.** (1997). *Children's speech and literacy difficulties 1: a psycholinguistic framework*. London: Whurr.
- Storkel, H. L., Armbrüster, J., & Hogan, T. P.** (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, 49, 1175–92.
- Swan, D., & Goswami, U.** (1997a). Picture naming deficits in developmental dyslexia: the phonological representations hypothesis. *Brain and Language*, 56, 334–53.
- Swan, D., & Goswami, U.** (1997b). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, 66, 18–41.
- Taha, H., & Saiegh-Haddad, E.** (2017). Morphology and spelling in Arabic: development and interface. *Journal of Psycholinguistic Research*, 46, 27–38.
- Themistocleous, C.** (2017). Effects of two linguistically proximal varieties on the spectral and coarticulatory properties of fricatives: evidence from Athenian Greek and Cypriot Greek. *Frontiers in Psychology*, 13. doi.org/10.3389/fpsyg.2017.01945.
- Verhoeven, L., van Leeuwel, J., Irausquin, R., & Segers, E.** (2016). The unique role of lexical accessibility in predicting kindergarten emergent literacy. *Reading & Writing: An Interdisciplinary Journal*, 29, 591–608.
- Walley, A. C.** (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286–350.

Appendix

Examples of stimulus items by word category (Identical, Cognate, Unique) and condition (real, pseudo). Bold symbols indicate location and identity of consonant substitution.

Identical words

Real word			Pseudo-word				Meaning
Stimulus in Arabic	Real word	Meaning	Stimulus in Arabic	Phonological change	Pseudo-word	Target word	
حمام	/ħamma:m/	bathroom	خريطة	Voicing	/ħari:da/	/ħari:ta/	map
جرس	/ʒaras/	bell	جمل	Place of articulation	/ʒanal/	/ʒamal/	camel

Cognate words

One-vowel distant cognates

وحش	/waʃh/	monster	نفر	Manner of articulation	/naʃr/	/naħr/	river
جمار	/ħima:r/	donkey	محلّة	Place of articulation	/maħla/	/naħla/	bee

One-consonant distant cognates

ثعلب	/θaʕʕlab/	fox	سَيِّت	Voicing	/sayt/	/zayt/	oil
سيف	/sayf/	sword	بنيع	Manner of articulation	/buði: :ʕ/	/muðiʕ /	broadcaster

Two-phoneme-distant cognates: one vowel and one consonant

سقف	/saqf/	roof	قنح	Manner of articulation	/qanħ/	/qamħ/	wheat
صحراء	/ʕaħra:ʔ/	desert	فل	Place of articulation	/faʔl/	/faʔr/	mouse

More than two-phoneme distant cognates

حائط	/ħa:ʔit/	Wall	فاكهة	Place of articulation	/faħeħa/	/faħeħa/	fruit
شاطئ	/ʃa:teʔ/	beach	زارق	Voicing	/za:req/	/sa:req/	thief

Unique words

سرير	/sari:r/	bed	جران	Place of articulation	/ħiza:n/	/ħiza:m/	belt
قبعة	/qubbaʕa	hat	ثُعْدان	Place of articulation	/θuʕdan/	/θuʕban/	snake

Cite this article: Saiegh-Haddad E, Haj L (2018). Does phonological distance impact quality of phonological representations? Evidence from Arabic diglossia. *Journal of Child Language* 45, 1377–1399. <https://doi.org/10.1017/S0305000918000302>