Reading and Spelling in Arabic: Linguistic and Orthographic Complexity

Haitham Y. Taha
Safra Brain Research for Learning Disabilities and the Learning Disabilities Department, University of Haifa, Israel; Special Education Department, Sakhnin College, Sakhnin, Israel

Abstract—This paper presents a discussion in an attempt to understand how reading and spelling processes in Arabic are influenced by the orthographic and linguistic uniqueness of the Arabic language, and what are the implications of this uniqueness on the acquisition of literacy skills from cognitive and developmental psycholinguistic points of view. From the psycholinguistic point of view, the paper deals with the unique developmental trajectories of phonological processing skills within the diglossic context of Arabic beside to the contribution of the orthographic exposure to the establishment of the phonological representations among native Arab readers. Other psycholinguistic and cognitive processes such as morphological and visual processing and their role within context of reading and word recognition were discussed. Comprehensive conclusions are suggested for the applied and research fields.

Index Terms—reading acquisition, Arabic orthography, diglossia, spelling, phonological awareness, morphological awareness

I. INTRODUCTION

Reading and spelling acquisition in general is a challenging process for children, and requires sufficient development of different cognitive and psycholinguistic abilities. Within different language systems, this process may be even more complex and challenging for children with low readiness due to the linguistic and the orthographic complexity of their specific written language. However, the main research findings about reading and spelling have come from English-language research (Share, 2008). This body of research was focused mainly on the acquisition of reading and spelling in the English language, which differs in its linguistic and orthographic features from many other alphabetic orthographies. In the last decade we have witnessed an intensity of research that examines the impact of specific linguistic and orthographic features of different language systems on the process of acquiring reading and writing in those systems. One of those interesting cases in terms of orthographic and linguistic features could be the Arabic language (Abu Rabia & Taha, 2006a).

II. LINGUISTIC UNIQUENESS AND PSYCHOLINGUISTIC ABILITIES

Arabic belongs to corpus of languages named Semitic languages. Different from other Semitic languages, Arabic Spoken language has different vernaculars which vary from one geographic area to another. The difference between those vernaculars can be expressed on all linguistic domains; phonological, morphological, syntactic and semantic levels. For example, we may find within one specific vernacular in a particular geographical area the existence of certain phonemes, when in turn, those phonemes could be absent in another vernacular of another geographical region. In addition, the meaning of certain words may vary from vernacular to another, beside the fact that verbal expressions that indicate a specific meaning can vary between different regions. In any case, all variations of different spoken vernaculars are different from the Modern Standard Arabic (MSA), which is considered universal literary language for Arabic speakers. The literary language is characterized by specific grammatical rules, with specific semantic and phonological systems differentiated from all spoken variations. Usually, the first time that Arabic-speaking children are exposed to the written standard Arabic is when they begin to read and write. Within this situation, and specifically at the point in time of learning about the Arabic alphabetic system and letter-sound correspondences, the children may find themselves exposed to a new linguistic system featured by phonological and syllabic structures which are different from those of their own spoken vernacular linguistic system. This linguistic distance between the written and the spoken language is a classic example of a specific linguistic situation called “Diglossia” (Ayari, 1996; Ferguson, 1959), and already has an effect on the acquisition of phonological awareness, reading and writing among Arabic-speaking children (Abu Rabia & Taha, 2006a; al Mannai & Everatt, 2007; Saiegh-Haddad, 2003, 2004). Following from this situation, and within the phonological level, phonemes that only exist in MSA and not within the spoken vernacular are considered as diglossic phonemes.

In the last decade, several psycholinguistic studies in the field of reading acquisition in Arabic have been conducted. Those studies dealt intensively in investigating the effect of this special linguistic situation on the process of reading acquisition and related domains such as spelling (Abu Rabia & Taha, 2004, 2006a, 2006b; Saiegh-Haddad, 2003, 2004,
According to Saiegh-Haddad (2003, 2004, 2007a), the pre-school linguistic experience and exposure determine the phonological representations and the access quality to those phonological representations. Saiegh-Haddad supported this postulation by several studies. In one of those studies (Saiegh-Haddad, 2003), kindergarten and first grade native Arabic-speaking children from north of Israel were tested with two phoneme isolation tasks, one for the initial phoneme and the other for the final phoneme. These phonemes were divided into two linguistic categories, spoken phonemes and standard phonemes. The children were asked to isolate the phonemes either from heard words or syllables. The syllables themselves were divided into two categories, spoken syllables and standard syllables. The findings of this research confirmed the previous assumption of Saiegh-Haddad about the linguistic constraints for the phonological representation among native Arabic children. It was found that the access to standard phonemes within the isolation process were more difficult than with the spoken phonemes. In the same spirit, the isolation from spoken syllables was easier and more effective than that from standard syllables. Within the same study, Saiegh-Haddad tested the decoding ability among the children using pseudowords decoding task. The findings showed that the large numbers of decoding errors were measured for pseudowords with MSA phonemes and syllabic structures.

According to Saiegh-Haddad, the linguistic distance between spoken and Standard Arabic, which has a great effect on the phonological representations among native Arab speakers, can be the main reason for the phonological awareness delay in development among Arab children when we consider the contribution of the phonological awareness to reading and spelling acquisition within alphabetic orthographies. Accordingly, Saiegh-Haddad (2005) found that the fluency of the decoding abilities, especially for new words in Arabic, is related to the degree of awareness to the standard phonemes. Such awareness increases throughout the exposure to reading and writing and continues to develop throughout elementary school (see Figure 1). This development is a direct result of the intensive exposure to the grapheme-phoneme correspondences of the written Standard words and makes the postulation about the contribution of the exposure to print to the development of the phonological representations of the words. According to Taha (in preparation b), the main contributor to the development of grain size of phonological awareness abilities among native Arabic-speaking children for the MSA words is the familiarity with the orthographic representations of such words. Taha postulates that the surface orthographic features of the written vowelized Arabic and the one to one phoneme-grapheme correspondences in spelling of the majority of the written Arabic words enable the child to use his knowledge about the orthographic pattern of the word as mechanism of feedback while performing phonological awareness tasks like the segmentation one. Accordingly, Taha found that the phonological segmentation of pseudowords in Arabic was less accurate than the segmentation of real words due to the fact that the children are not familiar with the orthographic patterns of those pseudo words (Figure 1). Also, Taha found that the phonological segmentation of words with surface correspondences between phonemes and graphemes were easier than the segmentation of words with deep correspondences. Accordingly, familiarity with the orthographic patterns of written Arabic words strengthens the phonological representation of such words. This postulation could be explained by the notion that the transparent relationship between the orthography and the phonology makes the familiarity with the orthographic pattern of the written words a main contributor to the phonological representation of such words. Accordingly, because of the surface features of the orthographic patterns of written Arabic words, and specifically the vowelized patterns, the exposure to reading and writing could be the main factor in establishing intact phonological representations of written words in Arabic. Anyway, the mutual relationship between the lexical orthographic representations and the phonological abilities was postulated by Taha (2006) suggests that phonological superiority could lead to lexical processing superiority.

Beyond the effect of the presence of diglossic phonemes, other phonemes considered as emphatic. Emphatic phonemes are those similar to other “soft” phonemes in Arabic but pronounced emphatically using the same articulation parts (for example the emphatic phoneme /ض/ = ḍˤ considered as emphatic to the soft phoneme /ذ/ = d). Within some vernaculars, some of those emphatic phonemes are absent within the specific phonological system of such vernacular
(for example the emphatic /ḍˤ/ does not exist within some spoken vernaculars). This phonological similarity between the emphatic and the non emphatic phonemes produces difficulties in spelling words that include one emphatic phoneme or more. A result of this difficulty can be inaccuracy in spelling. This spelling inaccuracy can manifest as phonetic errors (the subject writes down a pseudo-homophone instead of writing the correct orthographic pattern of the word e.g. مَفْعَلَة instead of مَفْعُولَة). Therefore, spelling words that contain those emphatics requires a specific familiarity with the word's orthographic pattern and demands intensive cognitive resources. Research on Spelling error analysis in Arabic found that the phonetic errors were developmentally the dominant spelling errors types from Grades 1 to 9 (Abu Rabia & Taha, 2006b). Abu Rabia and Taha postulated that the continued establishment of specific orthographic lexicon and reliance on direct phoneme-grapheme encoding while the full orthographic knowledge about the specific word does not available are the main causes of the phonetic errors. As mentioned above, this kind of errors emerged as the predominate type among normal spellers through different ages were found to be the predominate type among native Arabic dyslexic readers (Abu Rabia & Taha, 2004). Accordingly, words including emphatic phonemes can be spelled into two homophone orthographic patterns or more, and just one of those patterns could be the true one according to the Arabic orthographic conventions. As a result, the spelling of emphatic words in Arabic demands proficient spelling abilities and a strong establishment of the mental orthographic lexicon. This may explain why phonetic errors become the predominate errors along the development of orthographic knowledge, simply because there are always new emphatic words which the speller must be exposed to, while he/she does not yet have any stored orthographic patterns of those words. According to Abu Rabia and Taha (2004), this situation of producing the phonetic errors suggests the importance of the existence of lexical routes in the processes of spelling in Arabic for overcoming the production of such errors. Abu Rabia and Taha proposed a dual model for spelling in Arabic that illustrated the trajectories of the spelling process for familiar and non-familiar words. Based on the basis of the dual-route cascaded (DRC) model (Coltheart, Perry, Langdon & Ziegler, 2001; Coltheart, 2005), this model proposed phonological route (mapping the phonemes into graphemes directly) for spelling new unfamiliar words and lexical route for spelling familiar words and specifically words with specific orthographic features that cannot be written down only by reliance on phoneme-to-grapheme correspondences. Abu Rabia and Taha postulated that this lexical route contains the specific spelling rules in Arabic and the morpho-orthographic patterns beyond specific word knowledge and specific reading and writing rules.

III. ARABIC MORPHOLOGY, PROCESSING AND LITERACY ACQUISITION

The difference between the MSA and the spoken vernaculars can be expressed as differences in the morphological domain as well. In general, most of the words in the Arabic language, i.e., spoken vernaculars and the literate language are morphologically derived from roots. The root presents the basic and the general semantic meanings of all words which are derived from it. In general, the roots are composed from three to four sounds that could be represented orthographically by three to four letters. The order of those sounds in specific spoken words or the root letter in the case of written word must be consistent. Changing the order of those sounds or letters produces a new and different root. Accordingly, we can find that different words which were derived from the same root share a common basic semantic relatedness. This process of derivation of words from a specific root produces different nouns or verbs depending on the pattern that instills the root letters. So, we can find verbal patterns and noun patterns. The pattern represent a functional-syntactical and lexical meaning, accordingly different words with different roots that share the same pattern could have common functional meaning. For example: the following words: مَدْرَصَة (school) مُزْرَعة (farm) مَدْرَسَة (laundry) share the same pattern مَدْرَصَة. The general lexical status of those words means a functional places were work is being done there. Although those words have different specific meanings, but because of the shared pattern we can make a deduction that there is general lexical relatedness between those words; all those words describe places of work. This lexical relatedness is a direct result of the lexical function of the shared pattern. If we keep attention to the written forms of the aforementioned written words we can notice that there are common graphemes between the different written words, which appear at the same sequences within the three written words: the first grapheme and the last one beyond the diacritical marks (the small signs above the letters). Those common graphemes represent the specific pattern’s graphemes. Different patterns have different essential graphemes that differentiate one pattern from the other. These graphemes may appear as prefixes, suffixes, infixes or all of these in the same pattern. Three or four letters within each pattern represent the places of the root’s letters in the derivation process of new words by adding the root letters to the pattern. Usually the letters ف ل represents the general pattern letters that are intended to be replaced with the specific root letters in the derivation process. The letter /ـ/ in each pattern is always replaced by the first letter of the root. The letter غ is intended to be replaced by the second letter of the root. Whereas the /ـ/ is intended to be replaced by the third letter of the root. For example, the word مكتبة which is derived from the root ك ت ب and the pattern مكتبة (figure 3 illustrates the components of the pattern مكتبة).
Some researchers tend to believe that there is a direct impact of morphological awareness among readers of Semitic languages on their reading and spelling acquisition. Taha and Saiegh-Haddad (In preparation) found that morphological awareness among native Arabic-speaking children can be reached within two different trajectories, the root awareness trajectory and the pattern awareness trajectory. According to Taha and Saiegh-Haddad, children can reach the root awareness earlier than the pattern awareness, and this can be taken as evidence that deducing the relationship between words according to their patterns is much more cognitively demanding than deducing the root relationship, and accordingly pattern awareness develops at later stages. One important finding of Taha and Saiegh-Haddad’s study was the contribution of the awareness to the morpho-orthographic features of the morphological patterns to spelling words and pseudowords inflected by those patterns. The researchers postulated that the speller relies on his/her morpho-orthographic knowledge of those patterns while spelling the words. Those essential morpho-orthographic features of different written words derived from the stem or inflected on the same pattern are consistent between the different words; accordingly this matter can lead the speller to use his/her morpho-orthographic knowledge to spell new words using the pattern analogy and his/her root knowledge. This strategy makes the spelling more accurate. It means that this morpho-orthographic knowledge among Arabic spellers can be used as a cost-effective strategy because while spelling a word, the awareness to the specific pattern that this word was inflected on and the awareness to the root letter sequences of the specific word could produce a monitoring system in spelling. It is enough to be deducing to the morpho-orthographic features of the word pattern for spelling each word that inflected by this pattern using the analogy strategy with other familiar words that have the same pattern. This finding is consistent with the Abu Rabia and Taha’s (2004) spelling model. In their model, Abu Rabia and Taha postulated that morphological knowledge is an essential component of the orthographic lexicon among native Arabic readers. For example, the two words maDRaSa and maKTaBa are derived from two different roots but share the same pattern (the pattern= MaFaAla). That means the awareness to the morphemes of the pattern can prevent spelling mistakes, while this morphological knowledge can be used as a monitoring system for controlling the spelling errors, because as it is presented in this example, the essential pattern morphemes are repeated within the different words. The awareness to this repetition may be used to control the spelling performance. So the orthographic level of knowledge in Arabic includes the morpho-orthographic knowledge (pattern morphemes and root morphemes) beyond the specific orthographic patterns of the written words. Recent findings are supporting this notion where morphological awareness and knowledge contribute to the organization of the mental lexicon among Arabic readers. For example, Abu-Rabia (2007) compared the performances of normal readers versus dyslexic readers on several measures that affect reading in Arabic. The results of his research reveal significant differences between normal and dyslexic readers on all reading and spelling skills and the linguistic and orthographic skills as well, such as morphological awareness and syntactic awareness. One important finding was the significant prediction power of morphology, which had been measured by to tasks of identification and production, to reading accuracy and reading comprehension among the both reading groups across the different grades. Abu-Rabia postulated that the knowledge of the morpho-orthographic structure of the word was the best facilitator of reading accuracy and reading comprehension. The findings of Abu-Rabia are consistent with other findings that suggest the important role of morphology in the organization of the mental lexicon of Arabic readers (Abu-Rabia & Taha, 2004; Boudelaa & Marslen-Wilson, 2001; Purnet, Béland, & Idrissi, 2000). The main source of this assumption about the role of morphology in the organization of the mental lexicon was the analysis of reading errors made by Arabic readers. Within their study, Abu-Rabia and Taha (2004) found that the morphological errors in reading were the predominate error type among normal and dyslexic readers. These results were explained by the notion that both literary Arabic and spoken Arabic are rich with morphological structures, and because when there is a visually and phonologically similarity of words, usually related to the same root, this causes morphological types of errors in reading words in Arabic.

Note that the essential letters of the pattern were written in small letters while the root letters were written in upper-case letters.
finding indicates that the reader of Arabic relies on word-recognition strategies that involve high morphological mapping. Recent data from Taha, Ibrahim and Khatib (submitted) supports this finding, as the morphological errors were found as the major type of errors made by reading-disabled readers with mild phonological deficits. Other researchers found that within deep dyslexia in Arabic we can find that morphological errors could be found as the main error type which characterizes the inaccuracy of the failure in word recognition (Bélard & Mimouni, 2001; Purnet, Bélard, & Idrissi, 2000). Also, in this case of deep dyslexia in Arabic, the authors tend to believe that roots exist as lexical units in the mental lexicon of speakers of Arabic. This postulation is supported by findings from a case study of bilingual patient with deep dyslexia who speaks Arabic and French. These errors (morphological) differ quantitatively (error rate) and qualitatively (error types) in the two languages of the same patient (Purnet, Bélard, & Idrissi, 2000).

IV. ARABIC ORTHOGRAPHY AND VISUAL WORD RECOGNITION

Along with the general domain on the research of visual word recognition, many researchers believe that the timing of those processes involved in the visual recognition are affected by the features of the written word such as the word frequency (Frost & Chambers, 1973), orthographic similarity with other written words (Davis & Lupker, 2006) and the lexicality effect (word versus non-word) (Coltheart, 2005; Sereno & Rayner, 2003). Some researchers postulate that the speed of visual recognition of words can be affected by other factors, such as the specific orthographic features of a language’s writing system. For example, some researchers postulated the gaze duration that can be measured by using oculomotor recording during the reading process differs between different languages as a direct result of the different orthographic features of the languages (Roman & Pavard, 1987). Recently, we can witness extensive data from different studies that highlight the effect of the orthographic specificity on the speed of visual word recognition (e.g., Velan & Frost, 2011). Likewise Ibrahim, Eviatar, and Ahron-Peretz (2002) found that the specific orthographic features of the Arabic orthography can affect the visual word recognition in Arabic on the very early stages of letter identification. In their research, Ibrahim and colleagues (2002) found that those specific features of the Arabic orthography compared with another Semitic orthography (Hebrew) produce a visual load within the visual word recognition and the result is slower orthographic recognition as compared with other orthographies. According to this study, the Arabic orthography has specific orthographic features that can affect the speed of visual word recognition as compared with other languages. So, what are those specific orthographic features?

Arabic is a language written in an alphabetic system of 29 letters, all consonants except three, the long vowels. Most Arabic letters have more than one written form, depending on the letter’s place in a word: beginning, middle, or end. In addition, the letters are divided into categories according to basic letter shapes, and the difference between them is the number of dots on, in or under the letter. Dots appear with 15 letters, of which 10 have one dot, three have two dots, and two have three dots. Some of the letters can be connected with former and subsequent letter within the same word, while other letter can be connected just with former letters within the same word. The result of these conditions can produce different types of written words: 1) full connected words (connected letters) 2) partially connected words 3) non-connected words. The following are examples of each of these types of words: a non-connected word=تونس (lesson), a partially connected=نور (light), and a fully connected=عسل (honey). Also, most letters have more than one shape. The shape of each letter can differ according to its place and its connectivity with former and subsequent letters. For example, the basic and non-connected shape of the grapheme سا could change according to its placement within the word: درس (lesson), رأي (idea), مثب (proof). Some authors postulate that the orthographic features of the written Arabic can produce a visual load and affect the timing and slow down the orthographic processing (Eviatar & Ibrahim, 2000). Different from this assumption, recent findings from the study of Taha, Ibrahim and Khatib (2012), showed the processing of non-connected words by proficient normal readers is more time consuming and evident from the brain activity, which was measured by electrophysiological measures and show earliest processing differences between non-connected and connected words That were measured near the N170 component. This component represents the earlier stages of visual recognition.

In addition Arabic words are a combination of consonants and vowels. Skilled and adult readers are expected to read texts without short vowels; according to Abu-Rabia (1999) this demands heavy reliance on context and other resources. Beginners and poor readers read texts with short vowels. Vowelized Arabic is considered shallow orthography, and unvowelized Arabic is considered deep orthography (Abu-Rabia & Taha, 2006a). Reading accuracy in Arabic requires vowelizing word endings according to their grammatical function in the sentence, which is an advanced phonological and syntactical ability (Abu-Rabia, 2001).

The complexity of the visual information that each written word can carry (like different shapes of different letters, dots and the vowelization marks) forces the reader to rely heavily on visual processing besides the phonological processing. This postulation was supported by Taha (2008). Within his study, Taha examined the contribution of several cognitive processing skills to the accuracy in reading isolated vowelized words among 6th grade native Arabic readers. The participants were tested with different tasks to measure their phonological processing, naming ability, and their visual processing skills. Using a stepwise regression, it was found that the visual processing skills succeeded to predict significantly the accuracy in reading isolated words ($R^2 = 16.8\%$) beyond the significant contribution of the phonological processing skills ($R^2 = 31.9\%$). According to those findings, Taha postulated that both reading and visual word recognition in Arabic depend on sufficient visual processing skills beyond the phonological skills.
V. SUMMARY AND GENERAL CONCLUSIONS

The different research findings support the notion that reading and spelling in Arabic as literacy skills are affected by the specific linguistic and orthographic features of the language. The linguistic reality and specifically the diglossic situation have a significant impact on the development of the phonological representations of words in MSA at the lexical level. It was found, that the main facilitator of these phonological representations is the orthographic exposure (Taha, in preparation). Considering the orthographic exposure, we can conclude that the development of orthographic representations is mediated by other factors such as specific morpho-orthographic representation and specific writing rules (Abu Rabia & Taha, 2004). The strong and sufficient establishment of morpho-orthographic representations could be expressed by greater accuracy in reading and spelling words (Taha & Saiegh-Haddad, in preparation). As this was illustrated previously, this morpho-orthographic knowledge among Arabic spellers can be used as a cost-effective strategy in reading and spelling. Beside the linguistic factors that affected the acquisition of reading and spelling in Arabic in general, we should not neglect the vital role of the visual and orthographic process while we try to learn about the mechanisms that are involved in reading and word recognition in Arabic (Taha, 2008). The specific orthographic features of the written Arabic words lead toward relying heavily on visual processing skills within the visual word recognition.

Considering the applied recommendations according to the abovementioned discussions, it is important to consider the different linguistic and orthographic features of Arabic while we try to develop any pedagogical or clinical tools, while taking into account the developmental differences and the impact of those differences on the efficacy of reading and spelling skills. For example, the error analysis research revealed that there are specific types of errors, those to be considered as dominant types of errors among disabled Arabic readers. This finding should be taken into consideration within clinical situations.

In sum, further research should be carried out. The duty is to bring the research to meet the applied field to help children to overcome obstacles in acquiring written language, especially among those with difficulties or risk factors for difficulties.

ACKNOWLEDGMENT

This work was supported by a post-doctoral fellowship to H. T from the Edmond J. Safra Brain Research Center, University of Haifa.

REFERENCES


© 2013 ACADEMY PUBLISHER


Haitham Y. Taha is a researcher in the field of learning disabilities. He was born in small village in the north of Israel called “Deir Hanna”. His main interest is in investigating the basic cognitive and neurocognitive process in reading among typical and disabled native Arab readers. His main research work is being done in the Safra Brain Research in the Haifa University and in the reading and cognitive lab in the Sakhnin College in Israel.

Dr. Taha published different studies in the field of reading and reading disability among native Arab readers. Currently, Dr. Taha is the chair of the special education department in the Sakhnin College.