Reading and Spelling in Arabic: Linguistic and Orthographic Complexity
Haitham Taha, Sakhnin College
Safra Brain Research for Learning Disabilities- Haifa University
December, 2013
Main Points

• How diglossia in Arabic affect phonological awareness development?
• What facilitates phonological representations in Arabic?
• The effect of the specific orthographic features of the written Arabic on word recognition (cognitive and developmental point of view)
• Arabic morphology and the contribution to reading and spelling
• Practical implications and intervention
Modern Standard Arabic

• All variations of different spoken vernaculars are different from the Modern Standard Arabic (MSA), which is considered universal literary language for Arabic speakers.

• The MSA characterized by specific grammatical rules, with specific semantic and phonological systems differentiated from all spoken variations.

• This linguistic distance between written and spoken language is a classic example of a specific linguistic situation called “Diglossia” (Ayari, 1996; Ferguson, 1959)

• Usually, the first time that Arabic-speaking children are exposed to the written standard Arabic is when they begin to read and write.
Arabic diglossia effect and literacy acquisition

• Phonemes that only exist in MSA and not within the spoken vernacular are considered as diglossic phonemes.

• Affect the phonological representation of words belong to the MSA lexicalitily status only. (Abu Rabia & Taha, 2006a; al Mannai & Everatt, 2007; Saiegh-Haddad, 2003, 2004).
# Example

<table>
<thead>
<tr>
<th></th>
<th>North</th>
<th>center</th>
<th>south</th>
</tr>
</thead>
<tbody>
<tr>
<td>/q/</td>
<td>/a/</td>
<td>/K/</td>
<td>/g/</td>
</tr>
<tr>
<td>/θ/</td>
<td>/t/</td>
<td>exist</td>
<td>exist</td>
</tr>
<tr>
<td>/d'/</td>
<td>exist</td>
<td>/z'/</td>
<td>/z'/</td>
</tr>
</tbody>
</table>
kindergarten and first grade native Palestinian Arabic-speaking children from north of Israel were tested with phoneme isolation task.

The phonemes were divided into two linguistic categories, spoken phonemes and standard phonemes.

the access to standard phonemes within the isolation process were more difficult than the spoken phonemes.

The decoding ability among the children using a pseudowords decoding task was tested.

The large numbers of decoding errors were measured for pseudowords with MSA phonemes and syllabic structures.

Saiegh-Haddad, 2003
The contribution of the orthographic exposure to the phonological representations among Arab Readers.

Taha (in preparation)

- phonological segmentation of pseudowords in Arabic was less accurate than the segmentation of real words

- Conclusion: Exposure to print and intensive reading experience facilitate the phonological processing skills in Arabic
Accuracy on phonological segmentation of words and pseudowords by means of grade

The orthographic pattern is used as a reference to control the phonological manipulation.
Emphatic phonemes and spelling challenges

- Emphatic phonemes are those that share phonological similarity with other phonemes in Arabic and share the same articulation parts of the articulatory system.

- For example: the phoneme /ض/ = /d'/ considered as emphatic to the phoneme /د/=/d/.

- One main result of the phonological similarity between one emphatic phoneme and his non-similar emphatic phoneme is the difficulties in spelling (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. ضفدع instead of ضفدع).
• Accordingly, words including emphatic phonemes can be spelled into two homophone orthographic patterns or more, one of those patterns could be the true one according to the Arabic orthographic conventions.

• spelling of emphatic words in Arabic demands proficient spelling abilities and a strong establishment of the mental orthographic lexicon.

• phonetic errors become the predominate errors along the development of orthographic knowledge, simply because there are always new emphatic words to which the speller must be exposed, while he/she does not yet have any stored orthographic patterns of those words.
Means of percentages of the phonetic errors from the total errors types (Abu Rabia & Taha, 2006)
Can our Skilled Brains detect those pseudohomomophone errors?

The importance of intensive exposure to orthographic tasks
Resolving the phonological ambiguity during visual word processing in Arabic: an event-related potential investigation

Taha, H. & Khateb A. (2013, Frontiers in Human Neuroscience)
Predictions

• We predicted that there will be a dominant orthographic and lexical level analyses and as a result the discrimination between words and their corresponding PHom would be a difficult task.

• Such a discrimination would rely on additional cognitive resources such as deep memory monitoring rather than only basic visual discrimination processes.

• Hence, differences in the ERP (between PHom and real words) were expected during the early and late stages of processing but not during time periods necessarily devoted to phonological processing.
Behavioral results

Performance: No significant difference between words and PHom (p=.55, mean= 86±10.8% and 83±19.4%).

RTs: significantly faster for words than for PHom (t=-3.0, df=17, p<.009, mean= 729±124 and 784±147ms)
ERP analysis

A Words vs Pseudo-homophones

Visual processing stages

- Left
- Center
- Right
**ERP analysis-bis**

N170: Significant main effect of condition ($F(1, 14)=5.39, p<.04$),

P2: Significant main effect of condition ($F(1, 14)=16.36, p<.002$),
P6: Significant main effect of condition (F(1, 14)=21.53, p<.0004),
Latency: t=-2.73, df=14, p<.02
Source localization

A  
P2 maps

Words

PHom.

B  
P6 maps

Words

PHom.
Summary

- At the behavioral level, words were processed more rapidly than PHom.
- At the electrophysiological level, differences were found during early components (N170 and P2) but also the late P6.
- These differences confirm our predictions that words and PHom might differ during periods related to orthographic visual processing and late memory and decision monitoring phase.
Summary-bis

- The differences during the P6 in particular were due to an earlier P6 peak latency and higher amplitude in words than in PHom.

- The source localization analysis suggested that these were explained by an earlier and higher recruitment in words of left language temporal/frontal areas.

- The component where these late phase differences were looks like the centro-parietal P600 observed after syntactic anomalies (Osterhout & Holcomb, 1992) and is thought to reflect deep rather than automatic discrimination phase.

  Result: our skilled brains can overcome this pseudohomophone challenge
Arabic morphology and the contribution to reading and spelling

- most of the words in the Arabic language, i.e., spoken vernaculars and literate language are morphologically derived from roots.

- The roots are composed from three or four sounds that could be represented orthographically by three or four letters. The root presents the basic and the general semantic meanings of all words which are derived from it.

- The order of those sounds in specific spoken words or the root letter in the case of written word must be consistent.

- Different words than those 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.
• The pattern represents a functional-syntactical and lexical meaning, accordingly different words with different roots that share the same pattern could have a common functional meaning.
Illustration of a noun pattern’s morphemes and their grapheme representations within the whole pattern.
The general lexical status of those words means a functional places were work is being doing there.
Root vs. Pattern awareness
(Taha & Saiegh-Haddad, In preparation)

Deducing the relationship between words according to their patterns is much more cognitively demanding than deducing the root relationship, and accordingly full pattern awareness develops at later stages.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Root</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd</td>
<td>88.75</td>
<td>76.66</td>
</tr>
<tr>
<td></td>
<td>15.31</td>
<td>18.83</td>
</tr>
<tr>
<td>4th</td>
<td>92.50</td>
<td>79.58</td>
</tr>
<tr>
<td></td>
<td>15.81</td>
<td>21.92</td>
</tr>
<tr>
<td>6th</td>
<td>99.46</td>
<td>99.14</td>
</tr>
<tr>
<td></td>
<td>1.87</td>
<td>2.82</td>
</tr>
</tbody>
</table>

Means of the root awareness versus the pattern awareness within the different age groups

n.s: not significant
the speller relies on his/her morpho-orthographic knowledge the spelling process.

- deducing the morpho-orthographic features of the word pattern for spelling new word that inflected by this pattern using the analogy strategy with other familiar words that have the same pattern.

- This strategy makes the spelling more accurate and cost-effective.
Means and standard deviations of the accuracy in the different spelling tasks within the different groups of age.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Spelling of transparent-morphological words</th>
<th>Spelling of opaque words</th>
<th>Spelling of transparent-morphological pseudowords</th>
<th>Spelling of opaque pseudowords</th>
</tr>
</thead>
<tbody>
<tr>
<td>2\textsuperscript{nd} (n=48)</td>
<td>50 (23.6)</td>
<td>19.58 (11.47)</td>
<td>39.16 (15.8)</td>
<td>10.3 (6.2)</td>
</tr>
<tr>
<td>4\textsuperscript{th} (n=48)</td>
<td>64.16 (21.52)</td>
<td>28.8 (16.5)</td>
<td>63.54 (23.38)</td>
<td>13 (9.23)</td>
</tr>
<tr>
<td>6\textsuperscript{th} (n=47)</td>
<td>75 (20.1)</td>
<td>41.25 (23.59)</td>
<td>68 (22.3)</td>
<td>25.41 (20.23)</td>
</tr>
</tbody>
</table>
The Arabic writing system

The Arabic writing system is considered as a consistent letter-sound alphabetical one.

The system is composed of 28 letters (+ the hamza as the 29th), of which three serve also as long vowels together with other three short vowels.

The long vowels are: أ (a), و (waw), and ي (ya).

The short vowels are represented only by additional diacritics (or not at all):

- fatha (a) ب (ba),
- damma (u) ب (bu),
- kasra (i) ب (be),
- sukon ب (eb)
Visual similarity and Complexity of the Arabic alphabet

The Arabic alphabet consists of sets of dyads or triads of letters that are highly visually similar and also follow each other in the alphabet.

Letters in the same set (dyads or triads) share a basic shape and differ only in the number of dots and their position (above or below the letters): For instance

/خ, ح, ج, د, ذ, ذ, ف, ق, س, ص, غ, ع, ش, ث, ث, ظ, ط, ر, ز/
Orthographic connectivity

Within the orthographic patterns of the written Arabic words, a majority of the letters can be connected with former and subsequent letter, while other letter can be connected only with the former letters.

As a result, different types of written words can be produced:

i) fully connected (Cw): (عَسَلٍ = Honey)
ii) partially connected (PCw): (نُورُ = Light)
iii) non-connected words (NCw): (دَرْسٌ=Lesson).
Orthographic complexity

Orthographic complexity:
Eviatar and colleagues (2004) suggested that the complexity of the Arabic orthography may cause difficulty in the process of identification and processing of letters by the reader, but also delay the creation of orthographic patterns, which may cause difficulty in reading fluency (Eviatar et al., 2004).

Of the major research issues: (bis)

Ibrahim and colleagues (2002) reported that children with Arabic as first language and Hebrew as a second language, were slower in processing Arabic than Hebrew letters (Ibrahim et al., 2002).

Hypotheses
Taha et. al., 2013 (Brain Topography)

Some authors in previous studies argued that the specific features of Arabic orthography with its unique characteristics, in particular the connectivity issue (which changes letters’ basic forms), constitute a supplementary visual load that slows down reading (referred to hereafter as the complexity assumption):

On the basis of this assumption, one could predict that NCw (in which letters are represented in their basic form) are processed more efficiently (i.e. faster, more accurately?) than Cw (where the connectivity changes the basic forms of the letter).
Hypotheses-bis

In terms of brain activity, this differential processing of NCw and Cw would be reflected in the time period devoted to visual orthographic processing.

The event-related potential (ERP) component called the N170 is the brain response hypothesized to represent the time period of orthographic analysis.

Hence, a modulation of this brain response could be predicted by the words connectivity.

*This prediction would certainly be verified if the complexity assumption is correct.*
Hypotheses-bis

However, one can also predict that since NCw constitute a much less frequent category of written words in Arabic and because skilled readers are much more used to Cw than NCw, the later will be processed slower.

For this purpose, in a first experiment we collected ERPs from native Arabic adult speakers during a lexical decision task involving the three types of words selected according to their internal connectivity.
Methods

Participants: Eighteen (15 females and 3 males) native Arabic speaking students from the University of Haifa, age from 19 into 34, mean = 23.39, SD= 3.83

Stimuli: A total of 180 words: 1) 60 NC words 2) 60 PCw 3) 60 CW, together with 360 pseudo-words (PWs)

EEG recordings and analysis: the ERPs were collected from 64 channels, waves were analyzed during the time window of the N170.
Methods-bis

Procedure: Each stimulus was presented for 150ms centrally. The response window= 1550ms. Subjects, at 90cm distance and performed a speeded lexical decision task (LDT): they were asked to decide as quickly and accurately as possible if the presented stimulus was a real word or a pseudoword.

This study was published as: Taha, H., Ibrahim, R., Khateb, A. How does Arabic orthographic connectivity modulate brain activity during visual word recognition: an ERP study, *Brain Topogr.* (2013) 26:292-302.
Results

RTs: $1 \times 3$ ANOVA, $ns$

Connectivity effect: $F(2, 34) = 1.98, p = .15$
Results

Individual SDs: 1x3 ANOVA, NCw yielded the largest response variance, Cw, the lowest
1x3 ANOVA on accuracy

Connectivity effect: $F(2, 34)=36.74, p<.00000$
ERP analysis:

We first computed point-wise t-tests on all time point and all electrodes.

Statistical analysis then focused on the posterior electrodes for the N170.
Not only there appear difference in the amplitude of the response of the N170, but also in its time latency, let’s see that.
Source localization

N170 maps

Peak inverse solutions

Cw

NCw

Cw vs NCw (P<.01)
The results indicate that:

*Cw* are processed faster by skilled readers of Arabic read more efficiently *Cw* than *NCw* and this is reflected already in the very early stages of information processing (in the *N170 amplitude and latency*)
In terms of the brain areas involved:

In both cases the maximal activity was found in left temporo-occipital areas including the inferior occipital gyrus and the fusiform gyrus (BA 19 and BA 37)

This maximal activity was found areas frequently involved in orthographic processing and reading (at Talairach's (Talairach and Tournoux 1988) xyz coordinates: -42, -65, -7 for Cw and at -42, -71, -2 for NCw).

Taken together: these data suggest that connectivity does not impact negatively reading and word recognition processes in skilled readers of Arabic.
What about children?

Indeed, one might assume that the complexity assumption is real issue and the question of connectivity impacts reading up to a certain age?

The question now is, if yes, when this effect disappear?
To address this question, we conducted the following study:

The Impact of Arabic Orthography on Lexical Decision Tasks: a Developmental Perspective (Khateb, Khateb, Taha, & Ibrahim, 2013 submitted)
Population

30 3rd graders
30 6th graders
30 9th graders

Task: lexical decision
Analysis: RT, response variance, Accuracy and detection measure
Predictions

• We predicted that in the lower grades, recognition of NCw would be faster than recognition of Cw due to the fact that NCw are visually less complex than Cw.

• In higher grades, due to the frequent exposure to connected forms, the recognition of Cw was predicted to be faster and more accurate than NCw words.
Global Results: what?

Words vs pseudowords as a function of age.
Main effect of age (F(2, 75)=11.2, p<.0001),
Main effect of word type (F (1, 75) =166.3, p<.00001).
Global Results: performance?

Words vs pseudowords as a function of age:
Main effect of age (F (2, 75) = 4.0, p < .03),
Main effect of word type (F (1, 75) = 44.8, p < .00001)
Specific Results: what?

Mean RTs per age and connectivity condition

Words as a function of connectivity and age: RT
3rd graders

• RTs: in 3rd graders, Cw yielded the slowest RTs which differed from both PCw (p<.00001) and NCw (p<.00001), with these two later conditions not differing (p<.0.14);

• Acc: in 3rd graders, PCw the highest rate of correct responses, then Cw then NCw
6th graders

- RT: in 6th graders, Cw yielded the fastest RTs, followed by PCw, and then by NCw (the slowest RTs) which differed significantly from the first two (p<.0005 and p<.0005 respectively);
9th graders

- RT: in 9th graders where no significant difference was observed between connectivity conditions.
Finally: Cw vs artificially DCw
(no difference at the lower 3rd grade)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Connectivity</th>
<th>Mean of individual RTs' medians</th>
<th>Mean of individual RTs' S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cw</td>
<td>1406 (478)</td>
<td>476 (168)</td>
</tr>
<tr>
<td></td>
<td>DCw</td>
<td>1421 (453)</td>
<td>471 (192)</td>
</tr>
<tr>
<td>3rd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cw</td>
<td>938 (156)</td>
<td>338 (107)</td>
</tr>
<tr>
<td></td>
<td>DCw</td>
<td>1285 (232)</td>
<td>429 (116)</td>
</tr>
<tr>
<td>6th</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cw</td>
<td>853 (156)</td>
<td>281 (86)</td>
</tr>
<tr>
<td></td>
<td>DCw</td>
<td>1066 (216)</td>
<td>331 (81)</td>
</tr>
<tr>
<td>9th</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What about disabled readers?

Indeed, one might assume that the complexity assumption is real issue and the question of connectivity impacts reading among disabled readers?

(Khateb, Taha, Elias-Matta, & Ibrahim, 2013)
NR = 58 (age range: 17-28.7)
DR = 20 (age range: 17-19.5)

<table>
<thead>
<tr>
<th>Connectivity</th>
<th>Normal readers</th>
<th>Disabled reader</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT-Med</td>
<td>RT-SD</td>
</tr>
<tr>
<td>NCw</td>
<td>660 (103)</td>
<td>186 (46)</td>
</tr>
<tr>
<td>PCw</td>
<td>666 (97)</td>
<td>167 (42)</td>
</tr>
<tr>
<td>Cw</td>
<td>666 (92)</td>
<td>172 (48)</td>
</tr>
</tbody>
</table>
Summary

• Our study showed that even 3rd graders, who are still considered reading beginners, were only partially affected by orthographic connectivity (since they showed higher accuracy levels for Cw than for NCw, with PCw exhibiting the higher performance).

• These findings suggest that in 3rd grade the children were apparently in an intermediary situation/a transition phase where Cw are still challenging, but in which NCw are clearly not the easiest to process.
Summary

• Our study showed that even 3rd graders, who are still considered reading beginners, were only partially affected by orthographic connectivity (since they showed higher accuracy levels for Cw than for NCw, with PCw exhibiting the higher performance).

• These findings suggest that in 3rd grade the children were apparently in an intermediary situation/a transition phase where Cw are still challenging, but in which NCw are clearly not the easiest to process.
Summary-bis

- We also showed that a transition towards a situation where $C_w$ presented the easiest condition had occurred already before the 6th grade.

- The total absence of connectivity effects among 9th graders could be explained by the very high frequency of the words used here (all words from school books, with those for 9th grade including those of the 3rd and 6th).

- After controlling the word frequency, the same connectivity effect was observed among the disabled readers also.
Orthographic vowelization

• Written Arabic words are a combination of consonants and diacritical marks represent the vowels.

• Without those marks different words could have the same orthographic patterns (homographs)
• Skilled and adult readers are expected to read texts without short vowels.

• This demands heavy reliance on context and other resources (Abu Rabia, 1999).

• Vowelized Arabic is considered shallow orthography, and unvowelized Arabic is considered deep orthography (Abu-Rabia & Taha, 2006).
Visual processing

• The complexity of the visual information that each written word can carry (like different shapes of different letters, dots and the vowelization marks) leads the reader to rely heavily on visual processing besides the phonological processing (Taha, 2013, Psychology).
Sixth Grade Normal Readers: 35 (age 12.12 (±0.39))
Sixth Grade Poor Readers: 33 (age 12.04 (SD ±0.29))

Stepwise Regression Analysis results of predictors for reading words beyond to the both groups of participants

<table>
<thead>
<tr>
<th>Predictors</th>
<th>$R^2$</th>
<th>$R^2$ Change</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological Deletion</td>
<td>.29</td>
<td>.29</td>
<td>25**</td>
</tr>
<tr>
<td>Visual perception (distinction- Beery)</td>
<td>.37</td>
<td>.12</td>
<td>19.12**</td>
</tr>
<tr>
<td>Visual search- Diamond</td>
<td>.45</td>
<td>.08</td>
<td>16.88**</td>
</tr>
<tr>
<td>Visual search- 592</td>
<td>.49</td>
<td>.04</td>
<td>14.57**</td>
</tr>
</tbody>
</table>

**$p < .001$**

Visual processing contribution to word reading accuracy in Arabic (Taha, 2013-Psychology)
Visual processing contribution to word reading accuracy in Arabic (Taha, 2013 - Psychology)
Summery and intervention implications and contents

• 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.

• Early intensive espouser to MSA during the kindergarten for establishing intact phonological representations of the MSA words. Besides, the importance of earlier exposure to print.

• 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.

• Considering the applied recommendations according to my previous discussions, it is important to consider the different linguistic and orthographic features of Arabic while we try to develop any pedagogical, intervention and clinical tools.
• **Intensive and “speeded” training** of phoneme-grapheme correspondences during the initial stages of reading acquisition (Using different tasks and specially computer assistance software- speeded responses)

• Training the visual discrimination and visual processing-perception skills beside to the phonological skills
• Using spelling tasks for training the phoneme grapheme correspondences and for contributing to the development of the orthographic lexicon (orthographic choice tasks: ضفدع ضفدع ضفدع ضفدع)

• 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).

• Intervention programs should always assess the response to intervention during different points of time throughout the intervention program.
Curriculum plan for reading and spelling acquisition for first grade readers

A
Letters (GPC rules)

Short + long syllables: CV
(Reading + Spelling)

B
Short + long syllables: CVC

Word forming tasks (R+s)

C
Reading aloud: Short texts

Reading Comprehension
Later

Morph-Orthographic Training

Reading Fluency intervention

Vocabulary

Reading comprehension strategies
Reading Acceleration Program
The Safra Brain Research for Learning Disabilities

• http://www.youtube.com/watch?v=uDq1W25UFyg
Thanks

شُكرًا