

Case Study: Dyslexic Typing Errors in Arabic (MS Word Web App vs. ATBar in their Spell Checker Tool Testing)

By Mashaal H. AlKadi

1. Methodology:

1.1. Testing Script Preparation:

1.1.1. Script Formation:

Firstly, the script has been chosen to be a general Arabic text (common language) written in an opaque writing system.

“Arabic Writing Systems:

- 1) *Transparent Writing System (Diacritized Script): each phoneme is represented by a letter or a diacritic mark such as vowel letters*
- 2) *Opaque Writing System (Non-Diacritized Script)”*

Secondly, the script contains 259 words with 100 misspelled words (colored in Red) for testing.

1.1.2. Writing Pattern of Common Dyslexic Typing Errors in Arabic:

According to studies, dyslexic errors were usually recognized by the following incorrect spelling mistakes:

- i. Swapping letters locations
- ii. Mixing similar shaped letters:
{e.g. Similar letters in shape and different in the number of dots:
 1. ف [pronounced Faa (1 upper dot),
 2. ف [pronounced Qaf (2 upper dots)] }
- iii. Incorrect replications letters
- iv. Skipping letters from words
- v. Difficulty in determining the direction
- vi. Sequential months\days remembrance

1.2. Testing Script

Testing the scripts was made by examining the spell checker’s behavior toward considering words as either correctly or wrongly spelled. In addition, testing examined the “list of suggestions” which the utility provides as a correct substitution in order study up to which extent the tool meets the requirement of a dyslexic user. In depth these categories are described in (3.1).

2. Testing Scripts:

This section includes the script used in testing in both cases: the original (correct) text (2.1) and the incorrectly spelled text (2.2).

2.1. Original Text (Non-Diacritized Script)¹ :

الدسلكسيا كلمة تأتي من أصل يوناني معناها (صعوبة مع الكلمة في اللغة) وهي نمط من عُسر القراءة المكتسبة (الكسيا) التي تصنف وفقا لمعدل الأخطاء في قراءة الكلمات والتهجئة الصحيحة. ومن الحالات المصاحبة لعسر القراءة ما يسمى بعسر الكتابة وهو اضطراب يتجلى في المقام الأول خلال الكتابة أو الطباعة، رغم أنه في بعض الحالات الخاصة، قد تؤثر أيضا على التنسيق بين اليد والعين في العمليات الموجهة بهذا الاتجاه أو التسلسل كربط العقد أو تنفيذ مهمة متكررة. ويتميز عُسر الكتابة عن خلل الأداء في أن الشخص قد يكون لديه الكلمة المراد كتابتها أو الخطوات السليمة واضحة في ذهنه، ولكنه ينفذ التسلسل بترتيب خاطئ.

من الممكن أن يظهر العسر القرائي لدى الشخص في أي جيل، غير أن أعراضه قد تختلف وتتغير من جيل إلى جيل. فلدى الرضع من الممكن أن تظهر "الديسلكسيا" على شكل صعوبات بالنطق وتعلم الكلام، بينما من الممكن أن تظهر في جيل أكثر تأخرا على شكل صعوبات بتحليل والتعرف على الأصوات، صعوبات بالقراءة والتهجئة أو حتى من خلال عدم القدرة على التعبير الشفوي السليم.

يمتاز عدد كبير ممن يعانون من الديسلكسيا بالقدرة العالية على فهم وإيصال المعلومات غير المكتوبة، بالإضافة إلى قدرتهم البارزة على التفكير بشكل منطقي ومتسلسل بسرعة أكبر من غيرهم، والحقيقة أن الكثيرين من المصابين بهذا العسر، خاصة "الحالات الصعبة" منهم، يتجاوزون الـ ١٤ نقطة في مقياس الذكاء، وهو عدد نقاط الحد الأدنى لتعريف "العبقرية".

أما الإشكالية الأساسية التي تعيق بحث ظاهرة العسر القرائي، فهي عدم قدرة الباحثين على التمييز بين الإصابة بهذا العسر وبين صعوبات القراءة العادية، بين صعوبات التعبير الكتابي والتعبير الشفوي، بالإضافة إلى عدم القدرة على التمييز بين الأنواع المختلفة ودرجات الصعوبة المتفاوتة من الإصابة بالديسلكسيا.

¹ <http://www.bokra.net/>

2.2. Misspelled Text (100 mistakes highlighted in Red colour):

الدكسلسيا كلمة تأتي من أصل يوناني معناها (صوعية مع الكلمة في اللفة) وهي نمط من عسر القراءة المتكسبية (الكسيا) التي تصنف وفقا لمعدل الأخطاء في قراءة الكلمات والتنهجة الصحيحة. ومن الحالات المصاحبة لعسر القراءة ماسيمي بعسر الكتابة وهو اضطراب يتجلى في القمام الأول خلال الكتابة أو الطاباعة، رعم أنه في بعض الحالات الخاصة، قد تؤثر أيضا على التنسيق بين اليد والعين في العمليات الموجهة بهذا الانجاه أو التسلسل كبرط القعد أو تنفيذ مهمة متكررة. ويتميز عسر الكتابة عن خلل الأداء في أن الخشخص قد يكون لديه الكلمة المرادا كتابتها أو الخطوات السليمة واضحة في ذهنه، ولكنه ينفذ التسلسل بترتيب خاطئ.

من الممكن أن يضره العسر القرائي لدى الشخص في أي جيل، غيرر أن أعراضه قد تختلف وتغير من جيل إلى جيل. فلدى الرضع من الممكن أن تدهر "الديسلسيا" على كشل صو عوبات بالنطق وتعلوم الكلام، بينما من الممكن أن تظهر في جيل أكثر تخأرا على شكل صعوبات بتلحيل والنعرف على الأصواتي، صعوبات بالقراءة والتهجئت أو حتي من خلال عدم القدرتي على التعبير الشفوي السليم.

مناز عدد كبير ممن يوعانون من الديسلسيا بالقدرت الاعلية على فهم وإبصل الماعلومات غير الموكتبة، بالإضافة إلى قدرةهم البرازة لعي التفكير بشكل منتقي ومتسلسل برسعة أكبر من غيرهم، والقحقيقة أن الكيرين من المصايين بهذا العسر، كاصة "الحالات الصعبة" منهم، يتجاوزون الـ ١٤ نقطة في ميقاس الذكاء، وهو عدد نقاط الحد الأدنى لعرتيف "العبقرية".

أما الإكشالية الأساسية التي تعيف بحت ظارهة السعر القرائي، فهي عدم قدرة الباتحن على التميز بين الإصابة بهذا العسر وبني صعوبات القراءة العادية، بين صعوبات التبغير التكامي والتعبير الشوفي، بالإضافة إلى عدم القدرت على التمييز بين الأنواع المختلفة ودرجات الصعوبة المفتاوتة من الإصابات بالديسلسيا.

3. Testing Results:

This section illustrates the testing results for both tools ATBar's spell check (as in 3.3) and the MS Word Web App's spell check (as in 3.2). In addition, it explains the results meaning through the colouring code (See Table 1 : Testing Scripts Colouring Code) classification used in the text format.

3.1. Testing Results Colour Code:

Table 1 : Testing Scripts Colouring Code

Format	Meaning
<u>Underlined</u>	Words recognized as "Names" which are English words written in Arabic letters with either <i>ignore option</i> or <i>no correction word-list supplied</i>
Bold Green	Correctly Recognized as misspelled word and the required correction was provided
Bold Blue	Correctly Recognized as misspelled word but the required correction wasn't provided
Bold Red	Misspelled words recognized wrongly as correct words
Words with Strikethrough	Correctly spelled words recognized by the tool as wrong words
Yellow Highlighted	Word Recognized as a wrong word but with no correction suggestion list

3.2. MS Word Testing Result (Text Corrected by MS Word):

الدكسلسيا كلمة تآتتي من أصل يوناتي معناها (صوعية مع الكلمة في اللفة) وهي نمط من عسر القراءة المتكسبية (الكسيا) التي تصنف وقفا لمعدل الأخطاء في قراءة الكلمات والتهجة الصحيحة. ومن الحالات المصاحبة لعسر القراءة ماسيمي بعسر الكتابة وهو اضطراب يتجلى في القمام الأول خلال الكتابة أو الطابعة، رعم أنه في بعض الحالات الخاصة، قد توثر أيضا على التنسيق بين اليد والعين في العمليات الموجهة بهذا الانجاه أو التسلسل كبرط القعد أو تنفيذ مهمة متكررة. ويتميز عسر الكتابة عن خلل الأداء في أن الخشص فد يكون لدبه الكلمة المرذا كتابتها أو الخطوات السلبيمة واضحة في ذهنه، ولكنه ينفذ التسلسل بترتيب خاطئ.

من الممكن أن يظهر العسر القرائي لدى الشخص في أي جيل، غيرر أن أعراضه قد تختلف وتغير من جيل إلى خيل. فلدی الرضضع من الممكن أن تذر "الديسلكسيا" على كشل صعوبات بالنطق وتعلوم الكلام، بينما من الممكن أن تظهر في جيل أكثر تخأرا على شكل صعوبات بتلحيل والنعرف على الأصواتي، صعوبات بالقرأة والتهجت أو حتي من خلال عدم القدرتي على التيعير الشقوي السليم.

متاز عدد كبير ممن يوعانون من الديسلكسيا بالقدرة الاعلية على فهم وإيصال الماعلومات غير الموكتبة، بالإضافة إلى قدرة هم البرازة لعي التفكير بشكل منتهي ومتسلسل برسعة أكبر من غيرهم، والحققة أن الكيرين من المصايين بهذا العسر، كاصة "الحالات الصعبة" نهم، يتجاوزون الـ ١٤ نقطة في ميقاس الذكاء، وهو عدد نققاط الحد الأدنى لعريف "العبقرية".

أما الإكشالية الأساسية التي تعيف بحت ظارها السعر القرائي، فهي عدم قدرة الباحثين على التميز بين الإصابة بهذا العسر وبنى صعوبات القراءة العادية، بين صعوبات التعبير التكاملي والتعبير الشوفي، بالإضافة إلى عدم القدرت على التمييز بين الأنواع المختلفة ودرجات الصعوبة المفتاوتة من الإصابة بالديسلكسيا

3.3. ATBar Testing Result (Text Corrected by ATBar Spell Check Utility):

الديسلكسيا كلمة تأتي من أصل يوناني معناها (صعوبة مع الكلمة في اللفة) وهي نمط من عسر القراءة المتكسبة (الكسيا) التي تصنف وقفا لمعدل الأخطاء في قراءة اللكمات والتهجة الصحيحة. ومن الحالات المصاحبة لعسر القراءة ماسيمي بعسر الكتابة وهو اضطراب يتجلى في القمام الأول خلال الكتابة أو الطاباعة، رعم أنه في بعض الحالات الخاصة، قد توءثر أيضا على التنسيق بين اليد وللعين في العمليات الموجهة بهذا الانجاه أو التسلسل كيرط القعد أو تنفيذ مهمة متكررة. ويتميز عسر الكتابة عن خلل الأداء في أن الخشخص فد يكون لدبه الكلمة المرادا كتابتها أو الخطوات السلبمة واضحة في ذهنه، ولكنه ينفذ التسلسل بترتيب خاطئ.

من الممكن أن يظهر العسر القرائي لدى الشخص في أي جيل، غيرر أن أعراضه قد تتختلف وتغير من جيل إلى خيل. فلهي الرضضع من الممكن أن تذهر "الديسلكسيا" على كشل صعوبات بالنطق وتعلوم الكلام، بينما من الممكن أن تظهر في جيل أكثر تخأرا على شكل صعوبات بتلحيل والنعرف على الأصواتي، صعوبات بالقراءة والتهجت أو حتي من خلال عدم القدرتي على التعبير الشقوي السليم.

متاز عدد كبير ممن يوعانون من الديسلكسيا بالقدرة الاعلية على فهم وإيصال الماعلومات غير الموكتبة، بالإضافة إلى قدرة هم البرازة لعي التفكير بشكل منتهي ومتسلسل برسعة أكبر من غيرهم، والحققة أن الكيرين من المصايين بهذا العسر، كاصة "الحالات الصعبة" نهم، يتجاوزون الـ ١٤ نقطة في ميقاس الذكاء، وهو عدد نققاط الحد الأدنى لعريف "العبقرية".

أما الإكشالية الأساسية التي تعيف بحت ظارها السعر القرائي، فهي عدم قدرة الباحثين على التميز بين الإصابة بهذا العسر وبنى صعوبات القراءة العادية، بين صعوبات التعبير التكاملي والتعبير الشوفي، بالإضافة إلى عدم القدرت على التمييز بين الأنواع المختلفة ودرجات الصعوبة المفتاوتة من الإصابة بالديسلكسيا.

4. Analysis

In this section, the results illustrated earlier are studied in depth. Firstly, spell checker works differently according to the punctuation marks (especially in MS Word). Secondly, (Table 2) in numbers show the difference between MS Word and ATBar spell checker in dealing with writing mistakes.

Table 2: MS Word Web App vs. ATBar in Spell Checking Accuracy through Statistics

Format	No. Of Words Recognized in each Correction Category	
	By MS Word	By ATBar
<u>Underlined</u>	4	5
Bold Green	66	61
Bold Blue	9	25
Bold Red	20	13
Words with Strikethrough	-	19

Thirdly, the table 2 gives the following facts: both tools deals with names (including English names written in Arabic letters) in the same behavior of considering them as incorrect words; whereas Word provides some substitution and ATBar provides the *ignore* option.

Fourthly, both tools give close numbers of **Bold Green** words which represent considering incorrect words as wrongly spelled and providing the required correction at the same time. Not to forget to mention, although they give close numbers but each tool has a different set of words in this category in overall. Also, this number could be higher on the ATBar's side, if the issue discussed in detail in 4.1 can be altered. That is, when a preposition is attached to a word, in some cases it's considered as an incorrect word while in true it's not.

Fifthly, in the **Bold Blue** category, the ATBar gives a higher number than the MS Word does in the number of recognized words as incorrect spelling but without the required correction. In fact, this is considered a strength point of the tool in the case of it being able to provide a higher number of detected mistakes along with the **Bold Green** category against the other tool. Also, it's considered as an advantage if it gives a higher number in the **Bold Blue** category and a lower number in the **Bold Red** category; and both cases are advantages on the ATBar's side as (Table 3: Testing Overall Result Outcome) proves.

Sixthly, the ATBar wins again in having a lower number of the undetected mistakes of the text by having only 13 incorrect words not classified as wrongly spelled and by MS Word having 20 undetected mistakes. Lastly, an issue that ATBar poses is for 19 correct words it has considered them as incorrect words and provided a list of corrections where as MS Word don't do (words with strikethrough).

In focus, the issues mentioned earlier are discussed in depth in the following part.

4.1. Notes on ATBar Spell Check Service:

1. For correct words it provides substitution as if they were incorrect (it serves as a thesaurus)
2. For names it provides (ignore) option
3. It doesn't work automatically for long scripts (on the current ATBar version [Alpha testing] longer than 248 words)
4. For incorrect words: The ATBar's spell checker behavior towards incorrect words attached to prepositions results in providing a correct substitution but without their prepositions.
5. For correct words: It recognizes words with prepositions attached as incorrect words and provides a suggested correction (the same word without preposition).
6. For certain words such as "التهجنة" in one presence it was wrongly written "والتهجة" by swapping letters' location and in another appearance it was written as a mistake "والتهجت" through spelling the last letter with similar alphabet in sound but different in shape. In fact, in the first case the correct substitution wasn't provided where in the second one it was.
7. **صعوبات** presents a case where a short vowel is written as a long vowel (letter) with an ignore option, where it was supposed to provide a correction for it.

4.2 General Notes on writing misspelling (ATBar vs. MS Words Statistics)

So, according to Table 3 ATBar's spell checker (86%) gives a higher accuracy in detecting mistakes than MS Word (75%) does. Also, it gives a lower number of undetected mistakes which provides a higher performance than MS Word do. But at the same time ATBar considers some correct words as if they were mistakes while they aren't. These numbers were calculated through the equations 1,2, and 3 based on statistics from Table 2.

Table 3: Testing Overall Result Outcome

Spell Checker Tool	No. of Mistakes Recognized Correctly ^{Eq(1)} (Out of 100)	No of Incorrect Words unrecognized (skipped) by the Spell checker ^{Eq(2)} (Out of 100)	No of Correct Words Recognized Wrongly ^{Eq(3)}
ATBar	86	13	19
MS Word (Web App)	75	20	-

Eq(1): No. of Mistakes Recognized Correctly
 = No. of **Bold Green** words + No. of **Bold Blue** words

Eq(2): No. of Incorrect Words skipped by the Spell checker
 = No. of **Bold Red**

Eq(3): No. of Incorrect Words skipped by the Spell checker
 = No. of Words with ~~Strikethrough~~

5. Conclusion

Misspelling some Arabic words either by replacing similar letters (alphabets with dots) or by switching their places results in creating new correct which the grammar may not consider as an incorrect word! Thus, grammar must be used to study the context of the sentence; as it is not just a spell checker. Even though mistakes were done through simulating a dyslexic user behavior these kind of errors must be studied in depth to collect accurate facts about the boundaries of such writing problem (e.g. for to how many columns a dyslexic user switch or replace letters). That is, gathering such knowledge will help the spell checker to compare the user's writing with a calculated set of dictionary words based on that knowledge.

In conclusion, ATBar fortunately has achieved a higher accuracy in spell checking results through getting a higher number of detected mistakes and a lower number of skipped mistakes than MS Word do. Also, not to forget the two issues of ATBar must to be checked which are: considering correct words as mistakes and considering correct words attached to prepositions as incorrect words while they aren't.