

# المعجم العربي في ضوء اللسانيات الحاسوبية

عمر مهديوي

كلية الآداب والعلوم الإنسانية، عين الشق - الدار البيضاء

**الكلمات المفتاحية:** المعجم العربي- اللسانيات الحاسوبية أو المنظور اللساني الحاسوبي- المعجم الورقي- المعجم الآلي- الأدوات الحاسوبية- الأدوات اللسانية- الحوسبة المعجمية- لسان العرب.

١٥٧

**ملخص:** سنتطرق في هذا العرض إلى المعجم العربي في ضوء المنظور اللساني الحاسوبي الحديث، مع الوقوف عند معمارية المعجم الورقي التقليدي ومقوماته وثغراته التي تفرض على المعجم العربي في الوقت الراهن، أكثر من أي وقت مضى، إعادة النظر في أسس الصناعة المعجمية القديمة والحديثة ومرتكزاتها في أفق بناء معاجم إلكترونية تنتشر بالأدوات والإجراءات التي توفرها الآلة مما يتصل بالسرعة في إنجاز العمليات، والدقة المتناهية في جمع المعطيات وفرزها وتخزينها واسترجاعها ومعالجتها بأقل ما يمكن من الكلفة والجهد والوقت. كما سنتناول الوسائل والأدوات الحديثة التي يمكن اعتمادها في تحديث المعجم العربي بنوعيه القديم والحديث، والتي من شأنها أن تساعد على بناء قواعد معطيات معجمية للمفردات العربية بنوعها البسيط والمركب، ومن ثم مساءلة الكثير من التطبيقات الحاسوبية من قبيل الترجمة الحاسوبية، والمدقق الإملائي والنحوي، والمشكل الآلي، والمولد والمحلل النحوي وهلم جرا. وهكذا، فقد اتخذنا من "لسان العرب" لابن منظور منطلقاً لاختبار تجربة حوسبة المعجم التقليدي، من خلال تطبيق المفاهيم والإليات التي توفرها اللسانيات الحاسوبية عموماً، والمعجمية الحديثة خصوصاً.

## مدخل

أصبح من المؤكد اليوم أن المعلومات تعتبر علماً دقيقاً أكثر من أي علم آخر (الطاقة مثلاً)، وعليها يتوقف التقدم والتطور السوسيوثقافي للأمم والشعوب؛ لهذا أضحت تطوير هذا المجال من الأولويات الأساسية؛ بالنظر إلى أهميته في تطوير العلوم والتقنيات وتعليم اللغات...

على هذا الأساس، لم يعد اليوم الاقتصار على المعاجم الورقية الموضوعية على طريقة المناهج التقليدية والمسماة في الأدبيات المعجمية بالصناعة المعجمية أو المعجمية، كافياً؛ لأن تقنيات التخزين ومعالجة المعلومات التي توفرها الآلة تمكن من بناء معاجم آلية أكثر استيعاباً لأكبر قدر من المعلومات، وفي أسرع وقت ممكن، وبأقل ما يمكن من الكلفة وفق ضوابط لسانية وحاسوبية صارمة.

## الحاسوب والمعجمية العربية

يقوم الحاسوب بدور رئيس في المعجمية الحديثة، خصوصاً فيما يتعلق بحفظ مراحل التطور التاريخي لمعاني المفردات، فاستخدام الحاسوب يمكننا من حفظ معلومات عن كل كلمة، بل مقاطع من النصوص التي كانت تستخدم فيها في حقب زمنية مختلفة. وإذا ما تمكنا من استخدام هذه التقنية في اللغة العربية؛ فإنه سيكون بإمكاننا تتبع مراحل تغيير معاني المفردات العربية عبر مراحل تاريخية مختلفة، ومعرفة مدى تردد كل مفردة، وتغيير الدلالات في النطق. وسيوفر لنا كل هذا معلومات دقيقة عن التغيرات التي تحدث على كل جذر واشتقاقاته، وما إلى ذلك. وبعبارة أخرى يمكننا من حفظ سجل شامل للغة.

وقبل أن نتطرق إلى كيفية بناء معاجم حديثة باستعمال تقنية الحاسوب، يجدر بنا إلقاء نظرة سريعة على الخصائص البنوية والعامية للمعجم.

تعتبر المعاجم، مثلها مثل كتب الآداب، أداة للتعبير بعمق عن الثقافة والحضارة والعلم في أي أمة من الأمم. كما أنها ترتبط أشد الارتباط بتاريخ الآداب، ويتجلى دورها الرئيس في وصف اللغة أو الترجمة من لغة إلى أخرى. إنها تكشف عن التاريخ والجغرافيا والفنون والعلوم والتقنيات... إن المعاجم بهذا المعنى تمثل منطقة الإحالة القوية المتمثلة أساساً في التراث اللغوي والثقافي والعلمي العام.

وتكمن أهمية المعجم في كونه صنف لغوي وكتاب يمد القارئ/ المستعمل بما يلي:

- معلومات حول العالم، وهذا ينطبق على الموسوعات كالموسوعة الإسلامية المعاصرة والمعاجم ثنائية اللغة أو متعددة اللغة، والتي تكمن وظيفتها في تحقيق التواصل مع جماعات لغوية أخرى في العالم.
- معلومات حول اللغة الخاصة للمتكلم اللغوي، وهذا ينطبق على المعاجم الأحادية اللغة، التي تسمح للمستعمل بضبط نظامه اللغوي العام.



والتواصل مع أفراد جماعته اللغوية.

وهكذا « لا نكون في حاجة إلى التأكيد على أهمية المعلومات في وصف أنظمة اللغات الطبيعية، فذلك أصبح من بديهيات الأمور في عصرنا الراهن، كما أن المتنفس الطبيعي لكل النظريات والمناهج اللسانية أصبح ما تحققه المعلومات من تقدم في صياغة البرامج القادرة على تقييس دماغ الإنسان، وبذلك أصبحت قادرة على صياغة قوانين صورية تقوم بدور مزدوج، من جهة وصف النظام اللغوي في سائر مستوياته باستعمال لغة عقلانية، ومن جهة أخرى فإن هذه اللغة تصبح قادرة على توليد سائر بنيات اللغة وفق قوانين الاستعمال العادي لها» .

إن المعجم إذن، هو الذاكرة الحية التي تجيب عن أسئلة القارئ، وتقوم بدور الحفاظ على تراث الأمة من الضياع، كما أنه مؤسسة اجتماعية تصون تكامل المعارف المقدمة إلى أفراد الجماعة اللغوية. ويعتبر أيضا نموذجا للكفاية اللغوية لدى المتكلم (في دماغه البشري)، أي منظومة من الوظائف المسجلة في الكفاية المعجمية على شكل متواليات لغوية يصطلح عليها في علم المعجم بالمدخل المعجمية.

١٥٨

وإذا كان المعجم باعتباره مصدرا للمعلومات-كيفما كانت سعة الكتاب أو غايته- فإنه يتضمن معلومات غير مستمرة، مرتبة ترتيبا محكما بطريقة ألفبائية أو اعتبارية، وكيفما كان الحال، فإن هذه المعلومات كثيرة العدد، ومكثفة الدلالة بشكل مثير للانتباه، فالمعجم الفرنسي (Le Petit Robert) مثلا يحتوي على ٤٧٠٠٠ مدخلا، بينما يتوافر (Le Petit Larousse) على ٧٠٥٠٠ مدخلا معجميا.

المعلومات في المعجم:

إن المعلومات في المعجم تكون ذات طبيعة لسانية في الغالب الأعم، ما عدا الصور والرسوم التي تمثل المعلومات غير اللسانية، وهي معقدة من حيث التركيب والبنية. أما مادة المعجم فليست نصا حرا، ولكنها برنامج بالمعنى التقني للكلمة، إنها متتالية من المعلومات المنظمة والمصاغة بشكل شمولي وكلي، وبطريقة محكمة، على شكل بنية هرمية مزدوجة، فهي:

(١) بنية رئيسية على شكل قائمة بالمفردات أو المدخل، مرتبة ترتيبا خاصا، تمثل الهيكل الرئيس للمعجم.

(٢) بنية ثانوية، وهي المعلومات المقدمة عن المدخل، ويمكن لمحتواها أن يتغير بحسب طريقة العرض التي يمكن أن تقترب بصورة موازية، تقوم بدور الشرح، وتختزل مئات الكلمات .

إن بناء معاجم آلية يستلزم إذن استحضار ثلاثة أمور رئيسية وهي:

أ - الكمية الكبيرة والمعقدة من المعلومات.

ب - التنظيم النسقي وبنية المعلومات.

ج - الفحص التوثيقي للنص المعجمي ومراجعته.

إن جمع المعلومات والمواد وإحصاءها في ضوء المناهج التقليدية، يتطلب عدة أفراد لهم كفاءة علمية عالية، أو بالأحرى مؤسسات ومعاهد تتمتع على شخصيات مرموقة ومهمة. في مقابل ذلك، إن أهم ميزة في المعجم الحديث تتجلى في طابعه الجماعي، على سبيل المثال: المعجم الموسوعي الفرنسي، الذي ظهر سنة ١٩٦٠، يضم حوالي ١٦٣٢٧٠ مادة معجمية، من بينها أكثر من ٤٠٠٠٠ مفهوما اصطلاحيا، واستغرق تحضيره سنوات عدة، وشارك في إعداده فريق علمي وثقافي رفيع المستوى يقارب ٢٠٠٠ فردا.

الأدوات الحاسوبية التي توفرها الآلة للمعجم:

لقد ساعد التطور السريع الذي عرفته تكنولوجيا الحواسيب، في الآونة الأخيرة، على معالجة المعلومات غير العددية بأقل تكلفة وبأسرع ما يمكن من الوقت، مما يتصل بالتخزين والبنية وإنشاء الإرساليات/ الرسائل اللسانية انطلاقا من قواعد مضبوطة وصارمة. السؤال الذي يطرح إذن هو: كيف نبني قاعدة بيانات معجمية للمفردات اللغوية بنوعها البسيط والمركب موجهة لخدمة المعاجم الإلكترونية للغة العربية؟ ثم ما هي الإمكانيات التي توفرها الآلة للمعجم؟

يمكن للمعجمي أن يستخدم معالجات ذكية سواء في البيت أو المكتب، تساعده على التواصل مع حاسوب آخر أكثر ضخامة، والاستفادة من المعطيات من خلال برنامج معلوماتي معين، وهذا الحاسوب الذكي قادر على القيام بعمليات حسابية سريعة تسرع من عمل المعجمي وتنفذه بأقل تكلفة، وبدقة عالية ومهارة غاية في الإحكام.

يستطيع المعجمي من خلال المعالج الآلي الحصول على المعطيات والمعلومات بطريقة أشبه ما تكون بالباشرة، ويستمد هذه المعلومات من بنك المعطيات المخزنة في الحاسب على الشبكة، وفي الوقت نفسه يمكنه معالجتها.

إن استخدام الحاسوب بالنسبة إلى المعجمي لا ينحصر فقط في رصد الخصائص المميزة لهذه التقنية، بقدر ما يتعداه إلى الإمكانيات

التي يتيحها الحاسوب على مستوى التطبيقات الآلية، إذ يساعد على تنظيم بنوك المعطيات الضخمة، وهو ما لا يتيسر له القيام به اعتماداً على الذاكرة البشرية المحدودة.

ومن بين الإمكانيات التي تتيحها الآلة للمعجمي نذكر ما يلي:

- تحرير النص على الآلة.
- استرجاعه بكل يسر وفي أسرع وقت ممكن.
- ضعف كلفة تخزين المعلومات وتكثيفها.
- استخدام المنهجية الصورية (الخوارزميات) في معالجة النصوص.
- الاستفادة من تقنية المدقق الإملائي والنحوي في مراجعة النص المعالج.
- توظيف تقنية القارئ الضوئي.
- التمكن من تجاوز المراحل التقليدية التي كان يعتقد خطأ أنها إجبارية في معالجة نص من النصوص.
- تقديم طيف من المعلومات التي تسهل العمل المعجمي.
- يساعد الحاسوب على التسريع في إنجاز العمليات والقيام بالتعديلات الممكنة إما بالحذف أو الزيادة في محتوى الكتاب، مما يسمح بتطوير العمل المعجمي وبلورته في أية لحظة.

تهدف المعلومات إذن، إلى جمع المعلومات وحفظها في بيانات ومعطيات مبنية واسترجاعها، وفي هذا الإطار طبقت تقنيات تخزين المعلومات ومعالجة البيانات على بنوك توثيق المعلومات وحفظها، ويمكن الإشارة في هذا السياق إلى بنوك المعلومات المصطلحية المنتشرة في مناطق مختلفة من العالم العربي (باسم) بالمملكة العربية السعودية و(مكازن) الشركة العالمية للإلكترونيات بالقاهرة، (المعربي) Lexar، والذي تم بلورته وتطويره فيما بعد بمعهد الدراسات والأبحاث للتعريب بالرياض.

تتوجه المعلومة المعجمية المبنية المراجعة لخدمة تقنيات التخزين؛ لهذا فالحاسوب يساعد على تخزين الوحدات غير المستمرة وبرمجة الذاكرة على المعلومة النحوية الأكثر أو الأقل تعقيداً مقترنة بوحداتها اللسانية، ويتعرف عليها في معجم اللغة وفق برنامج المؤشرات النحوية وطريقة النطق أو الكلام، وأصول الكلمات (الإيمولوجيا)، التاريخ والتعريف والأمثلة. إن المعجم الثنائية اللغة تمدنا بالمقابلات الضرورية للمصطلحات العلمية والتقنية، في حين ينحصر دور المعجم الأحادية اللغة في تزويد المستعمل بالمصطلحات المعاصرة المرتبطة بكفائته اللغوية.

أصناف المعجم العربية:

تعتبر المعجم، مثلها مثل كتب الآداب، أداة للتعبير بعمق عن الثقافة والحضارة والعلم في أي أمة من الأمم، إنها ترتبط أشد الارتباط بتاريخ الآداب، ويتجلى دورها الرئيس في وصف اللغة أو الترجمة من اللغة المصدر إلى اللغة الهدف. إنها تكشف عن التاريخ والجغرافيا والفنون والعلوم والتقنيات. إن المعجم، بهذا المعنى، تمثل منطقة الإحالة القوية المتمثلة أساساً في التراث اللغوي والثقافي والعلمي الخاص والعام.

المعجم بالمعنى العام هو مصنف لغوي، وكتاب يمد القارئ أو المستعمل بما يلي :

- معلومات حول العالم، وينطبق هذا على الموسوعات والمعجم الثنائية اللغة أو متعددة اللغة. وتتجلى وظيفة هذا النوع المعجمي في تحقيق التواصل مع أفراد الجماعات اللغوية في العالم.
- معلومات حول اللغة الخاصة للمتكلم اللغوي، هذا ينطبق على المعجم الأحادية اللغة، والتي تساعد متكلم اللغة على ضبط نظامه اللغوي، والتواصل مع الآخرين.

وتنقسم المعجم الأحادية اللغة إلى صنفين كبيرين وهما:

- أ- المعجم القديمة، وأهم وظيفة تضطلع بها هي القيام بدور الإحالة والميعار اللساني كلسان العرب (ق ١٤) وتاج العروس على سبيل المثال لا الحصر.
- ب- المعجم الحديثة، مثل متن اللغة لأحمد رضا والمعجم الوسيط، ويتميز هذا الصنف من المعجم بطابعه الأدبي، ويوصف الاستعمال اللغوي العربي من القرن الرابع إلى القرن العشرين، ويتوظيف واستعمال التقنية المعاصرة، لكنه يفتقر إلى وسائل وصف وتفسير الوقائع اللسانية في العالم العربي.

وإلى جانب هذا الصنف الورقي الحديث يمكن أن نذكر المعجم الإلكترونية التي تبنى أساساً اعتماداً على تطبيقات التقانة الحديثة مما يتصل بالمعالجة الآلية للغات الطبيعية على مستوى المعجم، وهي في الأساس عبارة عن قاعدة بيانات لغوية مشفرة، تشمل جميع مستويات التحليل اللساني مما يتصل بالنحو والأصوات والدلالة، ويشترط فيه أن يكون شاملاً وعماماً؛ لأن البرنامج اللساني المعد للمعالجة الآلية لا ينبغي أن يفشل في العثور على أية معلومة كلما كان نوعها، وكيفما اتفق؛ لأن أي خطأ في المعلومات المدخلة من شأنه أن يتسرب إلى باقي مفردات الجملة أو قل النص برمته، ومن ثمة يعرقل عملية اشتغال البرنامج الحاسوبي.

في البداية، لا بد من الوقوف على الأدوات التي تطوع العربية للمعالجة الآلية من قبيل: الشفرة العربية الرومانية والتي تمت صياغتها انطلاقاً من نصوص مختلفة: القرآن الكريم، النثر الكلاسيكي، النصوص المعجمية، والنصوص المعاصرة (الكفاية اللغوية المعاصرة). وللإشارة، فقد استوعبت هذه النصوص/ الشفرة جميع العمليات بشكل كاف، وإن كنا لا نتوقع أن تكون مطلقة؛ لأن التجربة الطويلة هي التي ستدفع إلى تطوير وبلورة الجزئيات.

ولقد وقع اختيارنا التطبيقي على معجم "لسان العرب" لابن منظور الإفريقي؛ لأنه نص معجمي شامل، ويقوم على نظام معجمي واضح، وقابل للدراسة والمعالجة بواسطة الآلة الحاسوبية.

تشفير المفردات العربية:

تعتبر الشفرة العربية الرومانية نتاج تجربة مركز الأبحاث والتطبيقات اللسانية العربية. فقديماً حُصرت الأبجدية العربية في ٢٨ علامة، بدون احتساب لام ألف (لا). وتمثل هذه العلامات سواء في بداية الكلمة أو وسطها أو آخرها بأشكال مختلفة. على أن هذا النوع من التوظيف يرجع إلى طبيعة الكتابة العربية العادية السريعة، والمتضمنة للأبجدية العربية بعد ٩٤، يضاف إلى هذا الرقم علامات أخرى: التاءان (ت-ة)، الألفان (ا-ى)، الهمزة ومختلف أشكال رسمها: ئ، و، ؤ، أ، ء. بل إن القراءة تقتضي ذكر الحركات القصيرة (فتحة- ضمة- كسرة) وتغيير حركة السكون، في مقابل الحركات القصيرة والسكون، التي لم يمثل لها في الأبجدية العربية؛ لأن القراءة تفرض التمييز بين الحركات الطويلة والصامتتين (و/ى).

فهل من الضروري إذن، التذكير برسوم الحروف من قبيل: الشدة، المدة، الوصلة والتتوين... التي يتطلبها النص العربي المعاصر؟ ليس من الطبيعي فرض هذا العدد بمختلف علاماته على الكتابة العربية؛ لأن هذه الشفرة ذات خصائص لاتينية. وبناء على مبدأ الاقتصاد، فإن ما يهم النص العربي هو المعلومات الدالة، بل لا ينبغي الأخذ بعين الحسبان الاختلاف الرمزي أو الهندسي للعلامات، بل على العكس من ذلك، لا بد أن نميز التاء المربوطة من التاء الطويلة، وأشكال رسم كتابة الهمزة عن الجزئيات الدالة.

الصوامت

فيما يخص الأبجدية العربية الكلاسيكية، هناك عدد من الصوامت والجهيرات (= Sonantes) لها مقابلاتها في اللغات الرومانية، ولهذا فهي لا تطرح مشاكل على مستوى التشفير، مما جعلنا ننقلها بحسب استعمالها اليوم، يبدو أن الكل متفق عليه:

N	ن	S	س	B	ب
H	ه	F	ف	T	ت
W	و	Q	ق	G	ج
Y	ي	K	ك	D	د
		L	ل	R	ر
		M	م	Z	ز

على أن أغلب الصوامت معروف في الأبجدية العامة للغات الرومانية، ولكي ننقلها إلى العربية لا بد من صناعة شفرة للمفردات.

إن نصيب الاعتبارية في هذه المواضع كبير جداً؛ لهذا سنعمل على انتقاء الرموز الأكثر شيوعاً تقادياً للرموز المبهمة، التي من المتوقع أن يلفظ بها وعي المستعرب اللساني، وفيما يخص هذا الاختيار لا بد من استحضار الاعتبارات الرمزية والمعطيات الصوتية (الفونتيكية) وشروط الشكل (الرسم)، والأعراف القديمة الخ. فقد رمزنا للصوامت المفخمة بتكرير الصامت البسيط على النحو الآتي:

إلى جانب ت T توجد ط  
إلى جانب د D توجد ض  
إلى جانب س S توجد ص

وانطلاقاً من التشابه الصوتي ('Analogie phonétique') ولفظنا Z Z بالنسبة إلى "ظ" وموازاة مع هذه الصوامت (vélarisés) (اللهوية) رمزنا للصوت الصفييري الحنكي الاحتكاكي "غ" ب RR، وللصوت الصفييري الحلقي المهموس "ح" ب HH.

وفيما يتعلق بالصفيريات (= Spirantes) بين الأسنان، فقد تم استعمالها مثلها مثل الصوت الشجري "ش"، أما الصوامت

المضغفة، فقد أوردناها كما كانت تستعمل في القديم : SH-DH-TH .

لقد دفعنا بعض الأسباب المتعلقة بالمائلة الرمزية إلى ضبط الهمزة -بوصفها صامتا مجهورا وشديدا في العربية الكلاسيكية- بالعلامة C، وحسب رسوم كتابتها أوردناها على النحو الآتي :

CY	ئ
%C	ء
CW	ؤ

وبناء على هذا النقل، قبلنا بترميز الصوت الحلقي المجهور (sonore) "ع" بـ "cc"، إذ اخترنا للتاء المربوطة رمزا خاصا بها، يمكن من دراسة T الصغيرة منعزلة عن T ؛ لاعتبارات صوتية صرفية.

في الواقع نعرف أن التاء المربوطة يمكن أن تختزل باعتبارها صوتا تنفسيا (Aspiratoire) وضعيفا، وهي على العموم لاحقة للتأنيث، وأخيرا لا بد من التفكير في العناصر والأدوات الضرورية الممكن استعمالها في لوحة المفاتيح اللاتينية. كتبنا الصوت الرخوي اللهوي (Spirante vélaire) "خ" بـ J الإسبانية، وإن كانت الـ x الإغريقية قادرة على الترميز لهذه الظاهرة، لكن فضلنا استعمال لها علامة D .

الحركات:

تنقسم الحركات في العربية إلى قصيرة وطويلة، يمكن التمثيل لها على النحو التالي:

A %	ا
U W	و
I Y	ي
A	--
U	-
I	--

رمزنا إلى الألف المذكورة، والأكثر ترددا في القرآن بـ: %P (ألف صغيرة) وعند ما يتعلق الأمر بالألف المقصورة (القصيرة) فالأمر يبدو معقدا للغاية، الشيء الذي دفعنا إلى تمثيلها بالحركة E، ومثل هذا الحل لا يخلو من عمق صوتي (فونتيكي) بالكاد. أما السكون (الحركة الصفر) أو عدم الحركة، فقد مثلنا له بمربع صغير □ في لوحة مفاتيح الآلة المستعملة. وأما التوين، فيمكن أن يدرج ضمن الصوامت، وبما أنه يلحق آخر الصوامت ضمن شروط معينة، أدرنا تصنيفه في إطار الحركات، ثم مثلنا له بـ "N" مسبوقة بحركة:

-IN
-UN
-AN

### علامات إضافية (تكميلية)

يظهر أن الصوامت أو المصوتات كافية للقراءة السليمة في العربية، في حين نجد بعض العلامات تبدو ضرورية بالنسبة إلى القراءة السهلة والسريعة مثل الشدة وهي تضعيف لصامت، يمثل لها بالنجمة \* . ولإنتاجها من جديد في النص لا ينبغي تمييز التضعيف الجذري كما في سد، من التضعيف الناتج عن المماثلة التأخيرية أو الرجعية الحاصلة في الصوامت الشمسية المسبوقة بالتعريف مثل.

### أسهل

وغني عن البيان ما بين هاتين الحالتين المتباعدتين من فرق، وهناك حالات أخرى للتضعيف مستعصية الحل مثل: اضطرب-عدت، حيث التضعيف صوتي (فونتيكي) بالكاد وليس شكلي، على الرغم من كوننا نفضل معالجة مجموع المشاكل في إطار بحث موجه . وانطلاقاً من قانون اقتصاد الرموز، وضعنا للوصلة نفس العلامة المعطاة للهمزة الأولى C% ، في حين وضعنا الهمزة مخالفة للسكون، أما المدة فهي رسم يرد في الوقائع اللسانية المختلفة، يمثل لها بالرمز آ (الهمزة ممددة شكل الألف)، وللحركة الطويلة (الفتحة الطويلة) علامة هي : â ومن خلال الشفرة مثلنا لها ب A ، %C% . وللإشارة فالرسم القرآني للمدة يتطابق مع هذا التحليل، والدليل على ذلك أن ءامنوا، حيث الرمز ~ يلعب دوراً آخر في القرآن، إنه للاختصار والإيجاز : يس، ص، المص، ربي، بني إسرائيل. وفي الحالتين معا فقد مثلنا له بخط صغير - .

وللإشارة، فهذا مجرد جزء من الرسم القرآني وباعتبار النص القرآني كتاباً مقدساً، فهو يتطلب مجموعة من الرموز والعلامات التي تستوعب خصائصه الشكلية: معايير النقط، والتجويد، والمؤشرات الفونتيكية، وإلى جانب هذا، هناك نصوص تحتفظ بقيود التطريز، وتعرض بعض العلامات المنسجمة مع الإيقاع الشعري.

الشفرة العربية الرومانية			
SH	ش	%	ا
SS	ص	E	ى
DD	ض	C	ء
TT	ط	C %	أ
Z Z	ظ	CW	ؤ
GG	ع	CE	ئ
RR	غ	B	ب
F	ف	T	ت
Q	ق	PT	ة
K	ك	G	ج
L	ل	HH	ح
M	م	J	خ
N	ن	D	د
H	ه	DH	ذ
W	و	R	ر
Y	ي	Z	ز
		S	س

العلامات التكميلية	
A	-
U	-
I	-
□	-
N	-
*	-
c%	أ
P%	ء (هذا)

الحركات الطويلة	
A %	ا
UW	و
LY	ي

### حوسبة المعجم العربي التقليدي

تتحدد بنية المعجم العربي في عائلات وأسر مورفولوجية تنحدر من أصل اشتقاقي يعبر عنه عادة بالجذر، ويتحقق الاشتقاق نظريا بشكل مباشر انطلاقا من الجذر، وبشكل غير مباشر انطلاقا من المشتق. وهكذا نصنف المشتقات ونرتبها من خلال علاقتها بالجذر-الأصل (الأم). وبموجب ذلك نحصل على المشتقات الأولية (= primaires) المرتبطة مباشرة بالجذر، والمشتقات الثنائية (= Secondaires) المنبثقة عن المشتقات الأولية، والمشتقات الثلاثية (= tertiaires) المصاغة انطلاقا من الشائيات الخ. ينقسم المعجم العربي على المستوى النحوي، إلى ثلاثة أقسام: فعل، واسم، وحرف أو أداة. وانطلاقا من نظام الاشتقاق (التوليد)

والمقولة النحوية للكلمة، سنعالج المعجم العربي في مستويات ثلاثة:

- ١ - مستوى الفعل المجرد: الفعل الثلاثي، الأداة، الفعل الرباعي، الفعل المزيد.
- ٢ - مستوى المشتقات الفعلية.
- ٣ - مستوى المشتقات الاسمية.

١٦٣

في مرحلة أولى سنتوقف على الأفعال المجردة، باعتبارها أصولا نظرية لعائلات مورفولوجية. ويبدو هذا النوع من الاختيار صائبا؛ إذ يتيح إمكانية تشغيل مدونة الجذور العربية. إن التحليل الكافي لهذه المدونة والتوارد الميكانيكي طرحا جملة من العراقيل في وجه الباحثين مما يتصل بالترتيب المعجمي أو الترتيب الصوتي (الفونتيكي)، والتي تمثل في سائر الأحوال مجال اشتغال اللسانيات العربية والسامية، على أن أي جهد في مجال الصوتيات العربية هو في حد ذاته يمس صميم اللسانيات الصوتية السامية والعامية.

### أسباب اختيار «لسان العرب

يعتبر لسان العرب لابن منظور أحد النصوص المعجمية العربية التي تستحق الدراسة؛ بالنظر إلى غناه وشموليته، فهو أضخم معجم في اللغة، كما أن أحد مميزاته التطبيقية تصنيفه الألفبائي الذي لا يتطابق مع الترتيب الألفبائي المعاصر، حيث ينطلق من الصامت الختامي؛ مما يسمح بدراسته على نحو أمثل، بخلاف المعاجم الأخرى كجمهرة اللغة العربية لابن دريد أو التهذيب للأزهري.

### المعلومات والشفرة

بعدما حددنا أسباب الاختيار، سنقوم بتشفير المعلومات الخاصة بكل وحدة معجمية، بعضها أولي كرقم الترتيب، والصنف النحوي، وعدد الحروف، والبعض الآخر معقد (مركب). وعلى هذا الأساس، أنشأنا شفرة مطابقة لهذه المعطيات يمكن استعمالها في لوحة المفاتيح.

وفيما يلي نورد الترميز المطبق في الشفرة على شكل جدول:

تحليل رموز الشفرة:

الخانة: ١-٥: تشير إلى المرجع المدرس، والمختصر للمعلومات المقدمة.

مثل: ل: لسان العرب، ت: تاج اللغة، الخ.

الخانة: ٦-١٠: رقم الترتيب الألفبائي المعاصر للمواد المعجمية الأفعال: المجردة والمزيدة، الكلمات المسكوكة، وخارج الاستعمال، الأدوات الخ.

الخانة ١١-١٤: رقم المشتقات مقرونة ماديا برؤوس المواد (مصدر الصيغ الفعلية، الأسماء، الجموع السالبة والمكسرة الخ).

الخانة ١٥: رقم الحروف (الصوامت والمصوتات الطويلة).

٠ = أداة أو اسم أكثر من صامتين.

١ = أحادي.

٢ = ثنائي.

٣ = ثلاثي.

٤ = رباعي.

الخانة ١٦-١٧: رقم اصطلاحي للصيغ الفعلية:

٠٠ = صيغة غير فعلية.

٠١ = صيغة أولي.

٠٨ = صيغة ثامنة.

.. = صيغة (ن).

الخانة: ١٨-٢٠: صوائت الأفعال الثلاثية:

٣ = فتحة (a)

٦ = ضمة (u)

٩ = كسرة (i)

١٨ : الصائت الثاني.

١٩ : الصائت الثاني.

٢٠ : الصائت الثالث.

الخانة ٢١-٢٩ : صوائت الثلاثي في المضارع:

٢ = فتحة (a)

٦ = ضمة (u)

٩ = كسرة (i)

١٦٤

الخانة ٢١-٢٢ : صوامت المضارع المطابقة للصائت الثاني من الخانة ١٨ .

الخانة ٢٤-٢٦ : صوائت المضارع المطابقة للصائت الثاني من الخانة ١٩ .

الخانة ٢٧-٢٩ : صوائت المضارع المطابقة للصائت الثاني من الخانة ٢٠ .

الخانة ٣٠ : بدائل (متغيرات):

٠ = ليست هناك بديلة.

١ = بدائل صامتية.

١- = بديلة صائتية للشخص المتكلم والمخاطب مثل

تر يتر تررت.

٢- = بديلة الواو والياء للشخص المتكلم والمخاطب مثل:

تلا يتلو وتليت.

ملحوظة: يرافق رقم البديلة كتابة الكلمة.

الخانة ٣١ : الكتابة الصوتية بحسب الشفرة العربية الرومانية.

الكتابة المزدوجة الموضع: لحرف واحد خانتان، وللضراع نقطة.

مشاكل وحلول

تبين لنا من خلال الوقفات الأولى عند مواد اللسان، أنه يطرح مجموعة من الصعوبات تتعلق أساساً بالترتيب المطبق عليه، اقترحنا لها بعض الحلول نوردها على النحو الآتي:

- إن ترتيب اللسان ترتيباً ألفبائياً بحسب الأواخر مهم وأساسى بالنسبة إلى الشعراء في نظم القوافي، وهو يتعارض مع الترتيب الألفبائي المعاصر (بحسب الأوائل). ولعل الهدف من هذا التعديل هو تسهيل عملية تكوين الملف والأعمال الأخرى الخاصة بالقوائم.

- حددنا المعلومات الواردة في اللسان، وعندما يتعلق الأمر بخطأ ما في حركة المضارع، فإننا نعود توالاً إلى "الصحاح" و"القاموس المحيط" و"تاج اللغة"، قصد تصويب الخطأ الخ.

- صنفتنا الألف في الأفعال الجوفاء أو المهموزة بحسب أصلها في "اللسان"، وعندما نصادف خطأ في المعلومات الواردة، فإننا نستند إلى الترتيب المشار إليه أدناه، ولا سيما عندما تختلط الهمزة بالألف مثال:  
دعا ← دعو

أها تظل أها.

- إن قلب الألف في بعض الأفعال إلى واو أو باء، قد أكثر من عدد مداخل الملف، مقارنة مع مداخل اللسان؛ لأننا عزلنا الصيغة الواردة في الواو، عن صيغة الياء.

جاءت التاء المربوطة على شكل الهاء مثل:

إرة

- الحجج الواردة في اللسان هي جزء منه.

- أسماء الأعلام والدخيل:

لقد فكرنا في بعض الأحيان وضع أسماء الأعلام والدخيل حيناً، ثم فصلها عن الشرح حيناً آخر، ورغم إيجابيات هذا الحل، فإن تحقيقه في الوقت الراهن يبقى صعب المنال بالنسبة إلى الأبحاث الإيمولوجية العربية. ومن جهة أخرى، لم نطرح كمبدأ للبحث فصاحة الأصل المعجمي، بقدر ما اقترحنا صياغة للمعجم العربي من جديد في ضوء الأدوات المعلوماتية، وعلاوة على ذلك، يبدو أن السلوك اللغوي تجاه العناصر الأجنبية وظواهر الاقتراض يمكن أن تقدم معلومات حول توظيف واشتغال اللغة.

- الكلمات الثنائية مثل أب، أخ، يد، فقد تم دمجها في المواد المعجمية، وذلك بذكرها مع الثلاثي تارة، ومع الرباعي تارة أخرى بطريقة أشبه ما تكون نسقية.

### خاتمة

حاولنا في هذه الورقة معالجة وضعية المعاجم العربية القديمة والحديثة، والإمكانيات التقنية والإجرائية التي توفرها التقانة الحديثة للسانيات عموماً، وللمعالجة الآلية للغات الطبيعية خصوصاً، مما يسمح ببناء معاجم عصرية على أسس ومرتكزات لسانية وحاسوبية صارمة في الآن نفسه.

إن الوضع الحالي للغة الضاد، يفرض أكثر من أي وقت مضى التفكير جدياً في بناء معاجم إلكترونية للغة العربية على غرار ما هو موجود في اللغات الأجنبية، قصد تقليص هوة الفجوة الرقمية بين المعاجم العربية والمعاجم الأجنبية.

إن حوسبة المعجم العربي اليوم لم تعد مطلباً من قبيل الرفاهية، وإنما أصبح ضرورة ملحة؛ ولهذا فالعرب أفراداً ومؤسسات مدعوون لبذل الجهود الملائمة والتنسيق بين الخبراء والفنيين والمهتمين واللغويين والمهندسين قصد بناء معجم عربي عصري يستمد أسسه ومرتكزاته النظرية والتطبيقية من الصناعة المعجمية الحديثة.

### المراجع

التعريف المعجمي بين المعجم الورقي والمعجم الآلي: عمر مهديوي، اللقاء العلمي الدولي الرابع للقاموسية بتونس حول 'التعريف وقضاياها في القاموس' أيام ٢٢ و ٢٣ و ٢٤ جوان، ٢٠٠٦.

اللغة العربية والحاسوب: قراءة سريعة في الهندسة اللسانية: د. محمد الحناش، مجلة التواصل اللساني، مج ٩، ٢٠٠٣.

توليد الأسماء من الجذور الثلاثية المعتلة، مقارنة لسانية حاسوبية: عمر مهديوي دكتوراه السلك الثالث، كلية الآداب والعلوم الإنسانية، جامعة سيدي محمد بن عبد الله، فاس، المغرب، ١٩٩٩.

تحليل اللغة العربية بواسطة الحاسوب سعد بن هادي القحطاني، مركز اللغة الإنجليزية معهد الإدارة الرياض، مجلة مجمع اللغة العربية الأردني ص ٢٤٦

[http://www.majma.org.jo/G05/05/04/068\\_\\_files/%D8%A7%D9%84%D9%81%D9%87%D8%B1%D8%B3.htm](http://www.majma.org.jo/G05/05/04/068__files/%D8%A7%D9%84%D9%81%D9%87%D8%B1%D8%B3.htm)

تأهيل اللغة العربية وتثبيتها حاسوبياً، عمر مهديوي، الندوة الدولية حول 'اللغة العربية ومتطلبات التهييء اللغوي'، معهد التعريب بالرباط، أيام ٢٧/٢٨ أبريل ٢٠٠٦.

توليد الأسماء من الجذور الثلاثية الصحيحة، مقارنة لسانية حاسوبية: ع. مهديوي، دكتوراه، كلية الآداب والعلوم الإنسانية، عين الشق، الدار البيضاء، جامعة الحسن الثاني، المغرب، قيد الإعداد.

حوسبة المعجم العربي، عمر مهديوي، الأيام اللسانية الوطنية التاسعة عشر سنة من البحث اللساني المقارن 'جمعية اللسانيات بالمغرب بتعاون مع كلية الآداب والعلوم الإنسانية، جامعة محمد الخامس أكادال، أيام ٢٨-٢٠ يونيو ٢٠٠٦.

خوارزميات صرفية للغة العربية، عمر مهديوي، المؤتمر السنوي الخامس لمجمع اللغة العربية السوري حول 'اللغة العربية في عصر المعلوماتية دمشق'، أيام ٢٠-٢٢، نوفمبر ٢٠٠٦.

مشروع نظرية حاسوب لسانية في بناء معاجم آلية للغة العربية: د. الحناش محمد، مجلة التواصل اللساني ع ٢، س ١٩٩٠.

مدخل إلى المعالجة الآلية للمعجم العربي، عمر مهديوي صحيفة الحوار المتعدن، ع ١٥١٨، ١٢/٤/٢٠٠٦.

- La binarité en Arabe Classique. Les verbes quadrillitères. W.Attalah et Y.Ayach. C.R.A.L..Nancy 2; 1961.
- Bases de données informatisées et dictionnaires Bernard Quemada Lexique2.PRESSES UNIVERSITAIRES DE LILLE1 ;981 ,p105.
- Les dictionnaires électroniques ; Michael Zock John Carroll ; .Revue de l'association pour le traitement automatique des langues.Revue de l'association pour le traitement automatique des langues ; ATALA .vol44 no2/2003.p7.
- Dictionnaires électroniques et Analyse Automatique de textes ; Max Silberztein. Masson ; 1993.
- Introduction Aux Recherches Linguistiques Arabes Sur Machine ; W.Attalah et Y.Ayach ; C.R.A.L.no 23 ; 1973.
- UNGEGN Working Group on Romanization Systems. REPORT ON THE CURRENT STATUS OF UNITED NATIONS ROMANIZATION SYSTEMS FOR GEOGRAPHICAL NAMES. Arabic Version 2.2. January 2003