في توليد المفردات بالحاسوب

أ.د. موفق دعبول

أ. مروان البواب

لا يخفى أن للغة العربية خصائص ومزايا تجعلها قابلة للمعالجة الحاسوبية، ففيها خاصية الاشتقاق الصرفي، واعتماد المعجم على الجذور، والصلة الوثيقة بين المبنى والمعنى، وإطراد القياس في كثير من الحالات الصرفية والنحوية والصوتية.

وتكمن أهمية معالجة اللغة العربية بالحاسوب في التطبيقات الهامة والمتعددة المرجوة منها: كالترجمة الآلية من العربية وإليها، وتعرف الكلام وتركيبه، والقراءة الآلية للنصوص المكتوبة، والكتابة الآلية للنصوص المقروءة، واكتشاف الأخطاء اللغوية في النصوص وتصحيحها، والتحاور مع الآلة باللغة الطبيعية، وشَكل النصوص غير المشكولة.

وللمعالجة الحاسوبية للغة العربية وجوه متعددة، فهناك المعالجة الصرفية، والنحوية، والدلالية، والصوتية...، وغيرها، وتنقسم المعالجة الصرفية إلى قسمين أساسين هما: الاشتقاق، والتحليل.

يتناول هذا المقال بعض النقاط المتعلقة بالقسم الأول من المعالجة الصرفية وهو الاشتقاق (أو توليد المفردات بالحاسوب).

1 – الجذور في العربية:

مما تمتاز به العربية أن مفرداتها ترد في الأعم الأغلب إلى جذور (أو مواد) ثلاثية، وأن عددا قليلا نسبيا من المفردات ترد إلى جذور رباعية أو خماسية أي إنه يمكننا أن ننظر إلى هذه المفردات على أنها مجموعة كلية تتألف من عدد من المجموعات الجزئية، تحتوي كل منها على عدد من المفردات ينظمها انتماؤها إلى جذر واحد، هو عنوان هذه المجموعة الجزئية.

فالمجموعة الجزئية للجذر (ك ت ب) مثلا تحتوي على عدد من المفردات نحو: [كَتَبَ، كاتَبَ، تكاتب، استكتب، كتابة، مكاتبة، استكتاب، كاتِب، مكتوب، كِتاب، وكتبية...الخ]، وجميع هذه المفردات تحوي في تركيبها حروف الجذر (ك ت ب).

أما عدد المفردات في كل مجموعة جزئية فيتناسب طردا مع خصوبة الجذر، من بضع مفردات إلى عشرات المفردات.

هذا وقد بلغ عدد الجذور المُحْصبة – أي التي يتولد منها أفعال – (7420) جذرا، منها (5590) جذرا ثلاثيا، و(1830) جذرا رباعيا.

منها: أننا بمعالجتنا للمفردات عن طريق الجذور سنوفر الكثير من الجهد والزمن، فبدلا من أن نعالج مئات الألوف من المفردات، نعالج بضعة آلاف من الجذور، دون التفريط بشيء من صحة هذه المعالجة ودقتها.

ومنها: أن عدد حروف الجذر غالبا ما يكون أقل من عدد حروف الكلمة؛ فالجذور هي في الأعم الأغلب ثلاثية، أما الكلمات فقد يصل طول بعضها إلى عشرة حروف، أو أكثر، ومن ثم فإن معالجة الكلمات عن طريق جذورها فيه تيسير واختصار.

ومنها: أن الجذر يتألف من الحروف فقط، ولا يحوي الحركات، وهذا يعني أن معالجة المفردات عن طريق جذورها سيجعلها أكثر سهولة وسرعة، على أن الحاسوب سيعالج هذه الحركات عن طريق أوزان المفردات كما سنرى قريبا.

وهناك طرائق عديدة لتخزين الجذور في الحاسوب، منها مثلا: تخزين الجذور الثلاثية في مصفوفة ثلاثية الأبعاد(28، 28،28)، [28 هو عدد حروف الهجاء]، يختص البعد الأول بالحرف الأول لجذر، والبعد الثاني للحرف الثاني، والثالث للثالث، فالجذر (ن ظ ر) – مثلا- نجده في الخانة ذات الإحداثيات (25، 17، 10)، والجذر (ن ح ر) في الخانة (25، 6، 10)، وهكذا...أي إن إحداثيات الخانة في المصفوفة تدل على الجذر المحدد، أما الخانة نفسها فنخزن فيها المعلومات اللغوية الخاصة بالجذر.

وهكذا نجد أن مزية الجذور تيسر المعالجة الحاسوبية للعربية، إذ يستطيع الحاسوب أن يولد جميع المفردات – التي تبلغ مئات الألوف- اعتمادا على عدد قليل نسبيا من الجذور (لا يتجاوز بضعة آلاف)، وعلى عدد قليل أيضا من قواعد التوليد (لا يتجاوز العشرات)، وبذلك نضمن السرعة والدقة، ونقلل من حجم مكان التخزين في آن معا.

2 – أبواب تصريف الأفعال الثلاثية:

الفعل الثلاثي له – باعتبار الماضي مع المضارع – ستة أبواب ؛ هي: (فعَل يفْعل)، و(فعّل يفعِل)، و(فعّل يفعّل)، و(فعِل يفعَل)، و(فعُل يفعُل)، و(فعِل يفعِل)، وقد جمعوها منظومة بقولهم:

فَتْحُ ضَمٍّ فَتْحُ كَسْرٍ، فَتْحَتَان كَسْرُ فَتْحٍ، ضَمُ ضَمٍّ، كَسْرَتَان

بلغ عدد هذه الأفعال الثلاثية (7508) فعلٍ، توزعت على الأبواب الستة المذكورة كما يلي:

الباب	مثاله	عدد الجذور	النسبة المئوية
الأول	نصَر ينصُر	2110	28.17%
الثاني	جلَس يجلِس	1838	24.48%
الثالث	منَع يمنَع	1152	15.34%
الرابع	علِم يعلَم	1894	25.23%
الخامس	كرُم يكرُم	484	6.45%
السادس	حسِب يحسِب	25	0.33%
		7508	100%

وقد درجت بعض المعجمات القديمة (كالقاموس المحيط) على إيراد أمثلة أبواب تصريف الفعل الثلاثي لدلالة عليها، نحو: [فَخرَ: كفَرِحَ ومنع: تَكَبَّر]، أي إن الفعل (فخر) يتصرف من الباب الرابع: (فخِر يفخَر) ومن الباب الثالث: (فخَر يفخَر).

أما المعجمات الحديثة (كالمعجم الوسيط)، فتورد الفعل الماضي وإلى جانبه رمز عين مضارعه للدلالة على باب الفعل، نحو: [قتَر فلان ـُـ قترا: ضاق عيشه]، أي إن الفعل (قتر) يتصرف من الباب الأول: (قتَر يقتُر).

وأما في المعالجة الحاسوبية فيجري التعبير عن هذه الأبواب بالأرقام من 1 إلى 6، وتخزَّن هذه الأرقام مع جذور الأفعال.

ويقوم الحاسوب بتوليد الفعل الماضي والمضارع والأمر وفق الباب المذكور، اعتمادا على رقم الباب وعلى حروف جذر الفعل.

فهو يولّد الفعل الماضي، اعتمادا على الصيغة التالية:

c1+v1+c2+v2+c3+v3

حيث c1 و c2 و c3 حروف الجذر.

و (ـَـ) = v1 لأن حركة فاء الفعل الفتحة دائما.

و (ـَـ) = v2 إذا كان الرقم المقترن بالجذر هو (1) أو (2) أو (3)، و (ـُـ) إذا كان الرقم المقترن بالجذر هو (5)، و (ـِـ) إذا كان الرقم المقترن بالجذر هو (4) أو (6).

وأما v3 فلها قواعد مُحَوْسَبَة بحسب اتصال الفعل بالضمائر.

ويولّد الفعل المضارع، اعتمادا على الصيغة التالية:

c0+v0+c1+v1+c2+v2+c3+v3

حيث c0 أحد حروف المضارعة (أنيتت)، و c1، و c2، وc3 حروف الجذر.

و (ـــَـــ) = v0 لأن حركة حرف المضارعة الفتحة دائما.

و (ـــْـــ) = v1 لأن حركة فاء الفعل السكون دائما.

و (ـــَـــ) = v2 إذا كان الرقم المقترن بالجذر هو (3) أو (4)، و (ـــُـــ) إذا كان الرقم المقترن بالجذر هو (1) أو (5)، و(ـــِـــ) إذا كان الرقم المقترن بالجذر هو (2) أو (6).

وأما v3 فلها قواعد محوسبة بحسب اتصال الفعل بالضمائر، وكون الفعل المضارع مرفوعا أو منصوبا أو مجزوما.

ويولّد فعل الأمر، على نحو مماثل.

وينبغي التنبيه هنا على مسألتين: (الأولى) هي أن الحاسوب يولّد هذه الأفعال مسندة إلى جميع الضمائر (المتكلم، والمخاطب، والغائب)، وفي جميع حالاتها الإعرابية، و (الثانية) هي أن هذه القواعد خاصة بالأفعال السالمة (التي تُكوّن قرابة 60 % من الأفعال كما يتضح من هذا الجدول).

نوع الفعل	النسبة المئوية لعدد الأفعال
السالم	57.04%
المضعف	7.6%
المهموز	7.7%
المثال	4.94%
الأجوف	11.51%
الناقص	10.32%
اللفيف	0.87%

أما الأفعال غير السالمة (وهي: المضعّف، والمهموز، والأجوف، والمثال، والناقص، واللفيف) فتحتاج إلى قواعد إضافية، تتعلق بالإدغام، والإعلال... كما سنرى لاحقا.

3 – أبنية الأسماء والأفعال:

مما تمتاز به العربية أيضا، أن مفرداتها تنتظمها أوزان (أبنية) محددة، بعضها يختص بالأسماء وبعضها الآخر يختص بالأفعال، فمن أمثلة أوزان الأسماء: فَعْل (بَيْت)، وفَعَل (جَبَل)، وفاعل (عالم)، وفواعِل (قَوارِب)، وفعالل (دراهم)...، ومن أمثلة أوزان الأفعال: فَعِلَ (شَرِبَ) وفَعْلَلَ (دحرج)، واسفعل (استغفر)...، ومع أن عدد هذه الأوزان يبلغ المئات، فإنها تقل كلما زاد عدد حروف الجذر، فلقد أكثر العرب التصرف في التصرف في الثلاثي لخفته في الاستعمال، فاستخدموا منه أكثر الأوزان التي يحملها، وأهملوا أقلها، أما الرباعي فقد ضيقوا نطاق استعماله، لثقله على اللسان، فاستخدموا منه القليل، وأهملوا الكثير.

وما يعنينا من هذه الأوزان في المعالجة الحاسوبية، هو أنها تعطينا قوانين التوليد الأساسية للأفعال والأسماء على حد سواء.

فإذا استعرضنا أوزان الأفعال الثلاثية المزيدة مثلا وجدناها تنقسم إلى ثلاثة أقسام:

قسم زِيد فيه حرف واحد، ويأتي على ثلاثة أوزان: (الأول) أَفْعَلَ كأكرَمَ، (الثاني) فَاعَلَ كقاتل، (الثالث) فعَّلَ كفرَّحَ.

وقسم زيد فيه حرفان، ويأتي على خمسة أوزان: (الأول) اِنْفَعَلَ كانكسر، (الثاني) اِفْتَعَلَ كاجتمع، (الثالث) اِفْعَلَّ كاحمرّ، (الرابع) تَفَاعَلَ كتباعد، (الخامس) تَفَعَّل كتعلم.

وقسم زيد فيه ثلاثة أحرف، ويأتي على أربعة أوزان، (الأول) اِسْتَفْعَلَ كاستخرج، (الثاني) اِفْعَوْعَلَ كاعشوشب، (الثالث) اِفْعَوَّلَ كاجلوّذ، (الرابع) افْعَالّ كاحمارّ.

وقد بلغ عدد هذه الأفعال الثلاثية المزيدة (13837) فعل، توزعت على هذه الأوزان على النحو الآتي:

	وزن المزيد	عدد الأفعال	النسبة المئوية
مزيد بحرف واحد	أفعل	2917	% 21.08
	فعّل	2719	% 19.65
	فاعل	1266	% 9.15
مزيد بحرفين	انفعل	652	% 4.81
	افتعل	1710	% 12.36
	افعّل	96	% 0.69
	تفاعل	1023	% 7.39
	تفعّل	2329	% 16.83
مزيد بثلاثة أحرف	استفعل	953	% 6.89
	افعوعل	62	% 0.45
	افعوّل	13	% 0.09
	افعالّ	97	% 0.70
	المجموع	13837	% 100

وكذلك فإن للفعل الرباعي المجرد وزن واحد هو فَعْلَلَ كدحرج، وأما الرباعي المزيد فيأتي على ثلاثة أوزان هي: تَفَعْلَلَ كتدحرج، وافْعَنْلَلَ كانحرنجم، وافْعَلَلّ كاقشعرّ.

وقد بلغ عدد هذه الأفعال الرباعية المجردة والمزيدة (2310) فعل.

وبذلك يكون مجموع أوزان الأفعال الثلاثية المزيدة والأفعال الرباعية (16) وزنا، فإذا رمزنا هذه الأوزان برموز مناسبة، وخزّنا هذه الرموز مع الجذور الخاصة بها، استطاع الحاسوب أن يولّد جميع هذه الأفعال (13837 + 2310= 16147 فعل) على حسب أوزانها المحددة على النحو الذي مر معنا آنفا، اعتمادا على هذه الرموز الستة عشر وعلى جذور الأفعال.

أما الأسماء (كالأسماء الجامدة، والأسماء المشتقة، والمصادر...)، فلها أيضا أوزانها الخاصة بها فاسم الفاعل مثلا يصاغ للثلاثي المجرد على وزن واحد هو (فاعِل) نحو: واضع، وبائع...، ويصاغ لغير الثلاثي المجرد على وزن الفعل المضارع المبني للمعلوم، مع إبدال حرف المضارعة ميما مضمومة، وكسر ما قبل آخره، نحو مُخْرِج، ومُتَوَاضِع، ومُسْتَغْفِر...

يولّد الحاسوب جميع هذه الأسماء اعتمادا على أوزانها وجذورها بالطريقة التي أشرنا إليها آنفا فهو يولد أسماء الآلة على وزن (مِفْعَلَة) مثلا وفق الصيغة التالية:

«م» + « ـِـــ» + ف + « ـْـــ» + ع + «ـَـــ» + ل + «ـَـــ» + « ة»

(حيث: ف = فاء الفعل، ع = عين الفعل، ل = لام الفعل)

فمن الجذر (ل ع ق) يولّد (مِلعقة)، ومن الجذر (ن ش ف) يولّد (مِنشفة)، ومن (ك ن س) يولّد (مِكنسة)، ومن (س ط ر) يولّد (مِسطرة)، وهكذا...

وينبغي التنبيه هنا على مسألتين: (الأولى) أن بعض الأسماء لها أوزان قياسية مطردة، فهذه يولّدها الحاسوب دون أن تكون رموز أوزانها مخزنة مع الجذور، وبعضها الآخر لها أوزان سماعية لا يقاس عليها، فهذه لا بد لتوليدها من أن تكون أوزانها مخزنة مع الجذور، (الثانية) أن توليد الحاسوب للأسماء يتناول تصريفها أيضا، أي توليد الأسماء في حالة التأنيث أو التثنية أو الجمع أو التعريف، أو الإضافة، أو التصغير...

4 – قواعد التصريف المشترك:

تشترك في هذه القواعد كل من الأفعال والأسماء، وتشتمل على قواعد الإعلال والإدغام، والإبدال ورسم الهمزة، وسنكتفي بإيراد أمثلة عنها طلبا للاختصار:

الإدغام:

مثال: يولّد الحاسوب الفعل الماضي المسند إلى ضمير المتكلم من الجذر (ع ل م) وفق ما يلي:

ف + « ـَـــ» + ع + « ـِـــ» + ل + « ـْـــ»+: تُ» ←

« ع» + «ـَـــ»+ «ل» + «ـِـــ»م»+ «ـْـــ» + «تُ»←عَلِمْتُ.

فإذا كان لام الجذر حرفَ التاء -(سكت) مثلاً- أعطى الحاسوب:

«س» + «ـَـــــ» + «ك» + «ـَـــــ» + «ت» +

«ـْـــــ» + «تُ» ⟵ «سَكَتْتُ»

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة الإدغام الصغير:

«ت» + «ـْـــ» + «ت» + «ـَـــ» ⟵ «ت» + «ـُّـــ» ⟵ «تُ»

وبتطبيق هذه القاعدة نجد: نَحَتْتُ ⟵ نَحَتُّ

مثال آخر: يولد الحاسوبُ الفعلَ المضارعَ المُسْنَدَ إلى ضمير الغائب من (ن ص ر) وفق ما يلي:

«يَـ» + ف + «ـْـــــ» + ع + «ـُـــــ» + ل +

«ـُـــــ» ⟵

«يَـ» + «ن» + «ـْـــــ» + «ص» + «ـُـــــ» + «ر» +

«ـُـــــ» ⟵ «يَنْصُرُ»

فإذا كان الفعل مضعّفًا – (مدّ) مثلا- أعطى الحاسوب:

«يَـ» + «م» + «ـْـــــ» + «د» + «ـُـــــ» + «د» +

«ـُـــــ» ⟵ «يَمْدُدُ»

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة الإدغام الكبير:

«م» + ـْـــــ + «د» + ـُـــــ + «د» +

ـُـــــ ⟵

«م» + ـُـ + «د» + ـْـ + «د» + ـُـ «م» + ـُـ + «د» + ـُـ

وبتطبيق هذه القاعدة نجد: يَمْدُدُ ← يَمُدُّ

الإبدال:

مثال: يولد الحاسوب الفعل الماضي المزيد المُسْنَدَ إلى ضمير الغائب من الجذر (ج م ع) على وزن (افْتَعَلَ) وفق ما يلي:

«ا» + ف + «ـْـ» + «ت» + «ـَـ» + ع + «ـَـ» + ل + «ـَـ» ←

«ا» + «ج» + «ـْـ» + «ت» + «ـَـ» + «م» + «ـَـ» + «ع» + «ـَـ» ← «اجْتَمَعَ»

فإذا كان فاء الجذر حرفَ الصاد –(صبر)مثلا– أعطى الحاسوب:

«ا» + «ص» + «ـْـ» + «ت» + «ـَـ» + «ب» + «ـَـ» + «ر» + «ـَـ» ← «اصْتَبَرَ»

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة الإبدال:

«ص» + ـْـ + «ت» + ـَـ ← «ص» + ـْـ + «ط» + «ـَـ» (في صيغة افتعل)

وبتطبيق هذه القاعدة نجد: اصْتَبَرَ ← اصْطَبَرَ

الإعلال:

مثال: يولّد الحاسوبُ الفعلَ المضارع المُسْنَدَ إلى ضمير جماعة المتكلمين

من الجذر (ج ل س) وفق ما يلي:

«نَـــ» + ف + «ـــْـــ» + ع + «ـــِـــ» + ل + «ـــُـــ» ⟵

«نَـــ» + «ج» + «ـــْـــ» + «ل» + «ـــِـــ» + «س» + «ـــُـــ» ⟵ «نَجْلِسُ»

فإذا كان الفعل أجوف – (سار) مثلا– أعطى الحاسوب:

«نَـــ» + «س» + «ـــْـــ» + «ي» + «ـــِـــ» + «ر» + «ـــُـــ» ⟵ «نَسْيِرُ»

[ملاحظة: ورد حرف الياء هنا، لأن الألف في (سار) أصلها الياء].

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة الإعلال بالتسكين:

«س» + ـــْـــ + «ي» + ـــِـــ ⟵ «س» + ـــِـــ + «ي»

وبتطبيق هذه القاعدة نجد: نَسْيِرُ ⟵ نَسِيرُ

مثال آخر: يولّد الحاسوب اسم الفاعل للمفرد المذكر من الجذر (ض ر ب) وفق ما يلي:

ف + «ا» + ع + «ـــــــ» + ل + «ـــــــٌـــ» ⟵

«ض» + «ا» + «ر» + «ـــــــ» + «ب» + «ـــــــٌـــ»

فإذا كان الفعل ناقصًا – (غزا) مثلا– أعطى الحاسوب:

«غ» + «ا» + «ز» + «ـــ» + «و» + «ـــٌـــ» ⟵ «غازِوٌ»

[ملاحظة: ورد حرف الواو هنا، لأن الألف في (غزا) أصلها الواو].

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة الإعلال بالحذف:

«ز» + ـِـــــ + «و» + ـٌـــــ ← «ز» + ـٍـــــ

وبتطبيق هذه القاعدة نجد: غازِوٌ ← غازٍ

مثال ثالث: يولد الحاسوب اسم الفاعل للمفرد من الجذر (ن ص ر) وفق ما يلي:

ف + «ا» + ع + «ـِـــــ» + ل + «ـٌـــــ» ←

«ن» + «ا» + «ص» + «ـِـــــ» + «ر» + «ـٌـــــ»

← «ناصِرٌ»

فإذا كان الفعل أجوف – (باع) مثلا- أعطى الحاسوب:

«ب» + «ا» + «ي» + «ـِـــــ» + «ع» + «ـٌـــــ»

← «بَايِعٌ»

[ملاحظة: ورد حرف الياء هنا، لأن الألف في (باع) أصلها الياء].

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة الإعلال بالقلب:

«ا» + «ي» + ـِـــــ ← «ا» + «ئـ» + ـِـــــ

وبتطبيق هذه القاعدة نجد: بايِعٌ ← بائِعٌ

رسم الهمزة:

مثال: يولّد الحاسوب الفعل المضارع المسند إلى ضمير جماعة الغائبين

من الجذر (م ن ع) وفق ما يلي:

« يَـ» + ف + « ـْـ»+ ع + «ـَـ»+ ل + « ـُـ»+ « ونَ»

« يَـ» + م + « ـْـ»+ ن + «ـَـ»+ ع+ « ـُـ»+ « ونَ»

يَمْنَعُونَ.

فإذا كان الفعل مهموزا – (قرأ) مثلا- أعطى الحاسوب:

« يَـ» + ق + « ـْـ»+ ر + «ـَـ»+ ء+ « ـُـ»+ « ونَ»

← يَقْرَءُون.

هنا ينبغي على الحاسوب أن يعالج هذه الحالة، وذلك بأن يطبق قاعدة رسم الهمزة:

ـَـ+ « ء»+ ـُـ ← ـَـ+ « ؤ» + ـُـ

وبتطبيق هذه القاعدة نجد: يَقْرَءُون ← يَقْرَؤُون.

المراجع:

– **إحصاء الأفعال العربية في المعجم الحاسوبي**، البواب، مرياتي، مير علم، الطيان، مكتبة لبنان، ط1، 1996

– **أسلوب معالجة اللغة العربية في المعلوماتية**، في كتاب « استخدام اللغة العربية في المعلوماتية»، إصدار المنظمة العربية للتربية والثقافة والعلوم، تونس، 1996

– **تصريف الأفعال والأسماء**، فخر الدين قباوة، مكتبة المعارف، 1988

- **جامع الدروس العربية**، مصطفى الغلاييني، المكتبة العصرية، ط 26، 1992
- **شذا العرف في فن الصرف**، أحمد الحملاوي
- **اللغة العربية والحاسوب**، نبيل علي، دار تعريب، الكويت، 1988
- **المعالجة الصرفية للعربية بالحاسوب**، المجلة العربية للعلوم، المنظمة العربية للتربية والثقافة والعلوم، العدد 32، 1998
- **النحو الوافي**، عباس حسن، دار المعارف بمصر، ط4، 1966
- **نظام الاشتقاق والتصريف في العربية**، مجلة معلوماتي، الحاسوب والتقنيات، العدد 86، كانون الأول 1999