

العنوان:	التحليل الصرفي للأسماء العربية
المصدر:	مجلة أبحاث لسانية - معهد الدراسات والأبحاث للتعريب بالرباط - المغرب
المؤلف الرئيسي:	حمداني، عبدالفتاح
مؤلفين آخرين:	اليزيدي، توفيق(م. مشارك)
المجلد/العدد:	ع23,24
محكمة:	نعم
التاريخ الميلادي:	2007
الشهر:	ديسمبر
الصفحات:	58 - 31
رقم MD:	596077
نوع المحتوى:	بحوث ومقالات
قواعد المعلومات:	AraBase
مواضيع:	حوسبة اللغة العربية، المعالجة الآلية للغة العربية، التحليل الصرفي
رابط:	http://search.mandumah.com/Record/596077

التحليل الصرفي للأسماء العربية*

مقدمة

تزايد الحاجة يوما بعد يوم للمعالجة الآلية للغة العربية ليتزايد معها حجم الدراسات والأبحاث والمقاربات التي أنجزت وتتجز في هذا الإطار. ويعد التحليل الصرفي جزءا من هذه المعالجة، حيث قام العديد من الباحثين العرب وغير العرب، نذكر منهم كلا من باكوالتر (Buckwalter) ولاركي (Larkey) وشين وجي (Chen & Gey) وخوجة ودرويش ودياب والسعداني وحشيش وغيرهم كثير، بتطويره وإغنائه باقتراحهم لمجموعة من الأنظمة التي تتباين فيما بينها من حيث الشكل أو المضمون أو هما معا. إلا أن تحليل الأسماء لم يحض بعناية كبيرة في مجال التحليل الصرفي مقارنة بالعناية الفائقة التي حضيت بها الأفعال. لذلك نود في هذه الورقة أن نعرض تصورا لسانيا يعنى بوضع قاعدة معطيات لغوية يدوية للأسماء العربية مزودة بزمرة من المعلومات الصرفية والمعجمية والتركيبية الضرورية، وتقديم تصنيف جديد لها مقرون بعدد لا يستهان به من القضايا العالقة. وتتوزع عناصرها الأساسية على المحاور الآتية. نقدم في البداية بعض النماذج من المقاربات المعروفة لإنجاز المجذعات (stemmers) والمحللات الصرفية، التي نبدي بعض الملاحظات العامة والموجزة على المنهجية المتبعة فيها لبلوغ الغاية المرجوة منها في الفقرة الأخيرة من هذه الورقة. ثم نعرض المنهجية التي نروم اعتمادها لإنجاز محلل صرفي عربي للأسماء يراعي عددا غير قليل من الخصوصيات التي

* قدم جزء كبير من هذه الورقة في الندوة الدولية التي نظمها معهد الدراسات والأبحاث للتعريب (2007). ونستغل هذه الفرصة لنتقدم بالشكر الجزيل لأعضاء فريق المعالجة الآلية للغة العربية الذين ساهموا بقسط وافر في تطوير التصور المتبني من خلال ملاحظاتهم القيمة ونقاشاتهم المفصلة والعميقة.

تتفرد بها اللغة العربية عن غيرها من اللغات الطبيعية الأخرى، ويأخذ بعين الاعتبار عمليات الحذف والإبدال التي تنطبق على أسماء دون غيرها، على أن نتناول بعد ذلك، بتفصيل، تصورنا لقالب الأسماء، بدءا بتصنيفها، وانتهاءا بالسمات التي ترتبط بها.

1. مقاربات جوهرية ومحللات صرفية للغة العربية

لقد تم في غضون السنوات القليلة الأخيرة اقتراح عدد لا بأس به من المحللات الصرفية من قبل باحثين، وخصوصا من قبل شركات ومؤسسات تعنى بوضع جملة من المحللات والقواميس والمترجمات الآلية رهن إشارة المتعلم والباحث العربي تسهيلا للتحصيل وتيسيرا للبحث اللغوي وغير اللغوي. وإذا تفحصنا هذه المحللات الصرفية من وجهة نظر لسانية محضة أو من وجهة نظر لسانية حاسوبية، تعترضنا زمرة من الصعوبات الناجمة عن بعض مكامن الضعف الوثيقة الصلة بعدم الاستفادة من نتائج البحثين اللساني واللساني الحاسوبي عبر تاريخهما القصير ومحطاتهما الحاسمة القليلة.

ويعد مفهوم التجذيع (stemming) في مجال اللسانيات الحاسوبية مفهوما جوهريا قامت على أساسه العديد من المجذعات والمحللات الصرفية التي تكثر الحاجة إليها في جل التطبيقات اللسانية الحاسوبية، وبالأخص فيما يعرف في هذه الأدبيات بأنظمة استرجاع المعلومات. والمقصود بالتجذيع، بصفة عامة، عملية حذف أو إزالة اللواصق من الكلمات السطحية/النصية للحصول على الجذع أو الجذر. وقد نالت الكلمات العربية قسطا لا بأس به من الاهتمام عكسته الأعمال الكثيرة التي أنجزت في هذا الإطار، والنتائج القيمة التي تمكنت من تحقيقها في فترة وجيزة للغاية.

ويكاد يجمع جل الباحثين في هذا الحقل المعرفي الفتحي على أن صرف اللغة العربية يتطلب إجراءات إضافية لكونه صُرفياً واشتقاقياً في الآن نفسه. فمشكل الأواسط طرح بحدّة خصوصاً في الأعمال التي كانت تروم بلوغ الجذر بدل الاكتفاء بالجدع، كما أن ظاهرة جمع التكسير (غير القياسي) شكلت عقبة صعبة حالت دون التوصل إلى نتائج وحلول دقيقة، هذا علاوة على قضية تخلي الكتابة العربية عن الحركات القصيرة وما يخلفه من لبس على مستوى الكلمة في حد ذاتها، مما يعيق عمل بعض الأنظمة التي تأخذ بعين الاعتبار جوانب معينة من المعنى (لا سيما تلك التي تقدم مقابلات إنجليزية للكلمات العربية).

وإذا ألقينا نظرة شاملة على جل هذه الأعمال يمكن تلخيص مضامينها باعتماد

المقابلات أو الثنائيات التالية:

- أعمال تستهدف الجذع وأخرى تستهدف الجذر،
- أعمال تعتمد على قواميس وأخرى لا تعتمد عليها،
- أعمال تستعين بالأوزان وأخرى لا تستعين بها،
- أعمال تلجأ إلى الترجمة الآلية (المقابلات الإنجليزية للكلمات العربية) وأخرى لا تلجأ إليها،
- أعمال قائمة على لائحة تتضمن عدداً أكبر من عدد اللواحق الموجودة في اللغة العربية وأخرى قائمة على لائحة تتضمن البعض منها فقط،
- أعمال تدرج معلومات صرف-تركيبية عن عناصر أو مكونات الكلمة النووية وأخرى لا تقدم عنها أية معلومة على الإطلاق.

جل هذه الأعمال تتدرج بكيفية مباشرة أو غير مباشرة في إحدى المقاربات الثلاث الموالية التي تدور في فلكها معظم الأبحاث الحديثة التي تسهم بشكل فعال في تطوير هذا الميدان: المقاربة الإحصائية، ومقاربة حذف اللاصقة، ومقاربة التحليل الصرفي، وهي التي سنتطرق إليها في الفقرات الفرعية الموالية لكونها أغنت حقل المعالجة الآلية للغة العربية.

1.1. المقاربة الإحصائية

جوهر مقاربة التجذيع الإحصائي هو حساب ترددات توارد الجذوع واللواسق، وتعتمد على ما يعرف بأنظمة n-grams لاستخراج الجذوع من الكلمات السطحية. يتم تكوين طبقات من الكلمات التي تبدأ بالحرف الأول نفسه (n-grams)، أو الجزء الأول منه، واستعمال تقنيات التصنيف لتثذيب هذه الطبقات. إنها مقاربة تعتمد على التحليل الصرفي الآلي وعلى تحليل التوارد الذي يقيم طبقات من الجذوع على أساس قياس التوارد. وقد طبقت هذه المقاربة على اللغة العربية من قبل كسو وكروفت (Xu & Croft) وميفيلد وآخرين (Mayfield & al.) ودو روك (De Roeck).

ينبني مجذع شين، مثلاً، على الترجمة الآلية (إنجليزي/عربي)، إذ يستخلص الجذع من الكلمات العربية بناءً على مقابلاتها الإنجليزية. فالمقابل الإنجليزي لكلمة عربية من قبيل *أطفال* هي *children*، وبما أن المجذع الإنجليزي له القدرة على تحويل الأسماء المجموعة إلى مفرد عن طريق إزالة لاصقة الجمع (نظراً للطبيعة الصرفية لهذه اللغة)، فإنه سيعثر على كلمة *child* التي تقابلها في اللغة العربية كلمة *طفل*. بهذه الطريقة يتم التوصل إلى جذوع الكلمات العربية بما في ذلك مفرد جمع التكسير. لقد قام شين بتطوير مجذع خفيف يقوم بحذف وإزالة السوابق واللواسق

المترجمة إلى اللغة الإنجليزية، واضعاً ست لوائح تتضمن الثلاث الأولى منها الحرف الأول والحرفين الأولين والحروف الثلاثة الأولى، بينما تشتمل اللوائح الثلاث الأخرى على الحرف الأخير والحرفين الأخيرين والحروف الثلاثة الأخيرة. الخطوات المتبعة في هذا المجذع هي:¹

- (أ) حذف الحروف الثلاثة الأولى إذا كانت موجودة في لائحة السوابق الخاصة بها عندما يكون طول الكلمة خمسة أحرف على الأقل،
- (ب) حذف الحرفين الأولين إذا كانا موجودين في لائحة السوابق الخاصة بهما عندما يكون طول الكلمة أربعة أحرف على الأقل،
- (ج) حذف الحرف الأول "و" إذا كان طول الكلمة أربعة أحرف على الأقل وتبدأ ب "و"،
- (د) حذف الحرف الأول "ب" أو "ل" إذا كان طول الكلمة أربعة أحرف على الأقل وتبدأ ب "ب" أو "ل"، وإذا كانت موجودة في مجموعة الوثائق العربية المعتمدة،
- (هـ) الإزالة المتكررة للاحتقتين المواليين في اللائحة المناسبة إذا كان طول الكلمة أربعة أحرف على الأقل قبل حذف هذه اللائحة،
- (و) الإزالة المتكررة للائحة الأولى الموالية في اللائحة المناسبة إذا كان طول الكلمة ثلاثة أحرف على الأقل قبل حذف هذه اللائحة.

¹ شين، ص (4-5). ويعتمد شين على المقاربة الإحصائية لحساب ترددات الحرف الأول والثاني والثالث من بداية الكلمة العربية وعدد الكلمات العربية التي تبدأ بسابقة خاصة، وترددات الحرف الأول والثاني والثالث من اللوائح العربية. ويقدم اللوائح الآتية:

بالنسبة للحرف الأول: و- ا- ب- ل- م- ت- س- ف- ك- ي- ن- د- ر- ش- غ- ه- ج- أ- ع- ح- بالنسبة للحرفين الأولين: ال- وا- با- لل- وم- وت- لا- سي- وس- وي- ول- كا- فا- او- ما- لي- بو- كو- ان-.

بالنسبة للحروف الثلاثة الأولى: وال- الم- بال- الا- الت- الب- است- وسي- السن- فال- الع- للم- الك- الف- كال- للا- الن- واس- الح- ولل- مال- ال- سال- لال-.

بالنسبة للحرف الأخير: ا- ن- ي- ة- ه- ت- م- ر- و- ل- س- د- ح- ف- ك- ب- ز- ش- ع- بالنسبة للحرفين الأخيرين: -ها- ين- ان- ات- ون- ية- هم- نا- سا- وا- يا- ني- ته- ري- لى- ار- ير- يه- تي- يل-.

بالنسبة للحروف الثلاثة الأخيرة: -تها- هما- تين- لتين- تهم- يان- يتش- تان- اتى- يون- رات- سمان- اته- تتا- رين- نية- رها- نها- ينا- لها-.

2.1. مقارنة حذف اللاصقة

تهدف مقارنة حذف اللاصقة، وهي المعروفة أيضا في الأدبيات باسم التجذيع الخفيف (light stemming)، إلى حذف السوابق واللواحق (دون الأواسط) من الكلمات السطحية/النصية العربية بغية استخلاص جذوعها من دون أن تهتم بجذورها الممكنة أو أوزانها المحتملة. من بين الباحثين الذين اشتغلوا في إطار هذه المقاربة وحاولوا تطبيقها على اللغة العربية، نذكر على سبيل المثال لا الحصر كلا من لاركي وآخرين.

التجذيع الخفيف حسب لاركي لا يتقيد بوجود قاموس للكلمات العربية، كما أنه لا يتقيد بالمقياس القائل بإمكان حذف اللاصقة عندما تكون البقية كلمة موجودة في اللغة العربية. ويقوم مجذع لاركي على ما يعرف بخفيف10 (light10) الذي لا يتوفر على جميع اللواحق الموجودة في اللغة العربية، بل يتضمن المستعمل منها بكثرة بحسب قياس ترددها في المتن المعتمد، ويعتمد على الخطوات الآتية:²

أ) حذف "و" من خفيف2 وخفيف3 وخفيف8 إذا كان ما تبقى من الكلمة هو ثلاثة أحرف أو أكثر،³

ب) حذف أداة التعريف إذا كان ما تبقى من الكلمة هو حرفان،

² لاركي، ص (13).

³ يتكون المجذع الخفيف للاركي مما يلي:

- خفيف1 يتضمن السوابق الآتية: ال- وال- بال- كال- فال-، ولا يتضمن لواحق.
- خفيف2 يتضمن السوابق الآتية: ال- وال- بال- كال- فال- و-، ولا يتضمن لواحق.
- خفيف3 يتضمن السوابق الآتية: ال- وال- بال- كال- فال- و-، واللاحقتين -ة -هـ.
- خفيف8 يتضمن السوابق الآتية: ال- وال- بال- كال- فال- و-، واللواحق -ها -ان -ات -ون -ين -يه -ية -هـ -ي.
- خفيف10 يتضمن السوابق الآتية: ال- وال- بال- كال- فال- لل- و-، واللواحق -ها -ان -ات -ون -ين -يه -ية -هـ -ي.

ج) العودة إلى لائحة اللواحق وحذف أية لاحقة موجودة في آخر الكلمة إذا كانت هذه العملية تبقي حرفين أو أكثر.

3.1. مقاربة التحليل الصرفي

تقوم هذه المقاربة بالبحث عن الجذع أو الجذر المناسب للكلمة السطحية التي يتم التعامل معها بعد إزالة كل اللواحق العالقة بها، مستعملة لهذه الغاية تقنيات مختلفة وأنظمة متباينة. ولعل ما يميزها عن سابقتها كونها حاولت فهم عدد كبير من الظواهر المعتادة في الصرف العربي مستغلة بعض النتائج القيمة التي تم التوصل إليها في هذا المضمار، مما أضفى عليها نوعاً من المعقولية في التعامل مع الكلمة العربية ولواحقها المختلفة. فبرزت للوجود مجموعة من المحللات الصرفية التي نخص بالذكر منها المحلل الصرفي لدرويش المعروف باسم "سيبويه"، والمحلل الصرفي لخوجة، والمحلل الصرفي لبكوالتر، والمحلل الصرفي لدياب، والمحلل الصرفي للسعداني وحشيش وغيرها. وبما أنه لا يسعنا في مقام كهذا التطرق بتفصيل لجميع هذه المحللات، نفضل أن نتناول المحلل المقترح من قبل بكوالتر وخوجة ودرويش على اعتبار أن بعضها يعتمد قاموساً يدوياً ويستهدف الجذع، وبعضها يعتمد الوزن ويبحث عن الجذر، في حين أن بعضها الآخر يقوم على قواعد آلية لاستخلاص الجذر.

1.3.1. المحلل الصرفي لخوجة

يرتكز هذا المحلل على أربعة لوائح متفاعلة فيما بينها، هي: لائحة السوابق، ولائحة اللواحق، ولائحة الجذور، ولائحة الأوزان. ويشغل هذا المحلل بالكيفية المالية: فيبعد أن يقوم بإزالة اللواحق على أساس العودة إلى القاموسين الخاصين

بهما، يتم التحقق مما تبقى على أساس العودة إلى قاموس الأوزان للبحث فيه عن وزن له نفس الطول. ففي حال عثوره على هذا الوزن يتوصل إلى الجذر، ويقوم بالتحقق من وجوده في قاموس الجذور. وهو بهذا محلل يزيل اللواحق ويتأكد من أن ما تبقى هو جذر موجود له وزن موجود. وفي حال عدم عثوره على الجذر يعيد الكلمة الدخلى في صورتها الأصلية.

2.3.1. المحلل الصرفي لبكوالتر

يبحث هذا المحلل عن الجذع اعتماداً على ثلاث قواميس (يدوية)، هي: قاموس السوابق، وقاموس اللواحق، وقاموس الجذوع، وكذا على ما يسميه بكوالتر بقبوء التوافق، وهي قبوء تكبح مجموعة من التأليفات التي لا تستسيغها الكلمات السطحية في اللغة العربية، وتنصب على ثلاثة تأليفات أساسية، هي: التأليف بين السوابق، والتأليف بين اللواحق، والتأليف بين السوابق واللواحق. ويقدم هذا المحلل نتائج مقرونة بعدد لا بأس به من المعلومات الصرف-تركيبية المتعلقة بمكونات الكلمة النووية: السابقة واللاحقة والجذع. وتجدر الإشارة إلى أن محلل بكوالتر يتعامل مع الكلمات السطحية برمتها، سواء أكانت مصحوبة بالحركات القصيرة أو مجردة منها.

3.3.1. المحلل الصرفي لدرويش (سيبويه)

يقترح درويش منهجية لإنشاء محلل صرفي (في يوم واحد) مؤسس على إحصاءات وقواعد آلية بمقدوره توليد الجذور الممكنة لأية كلمة عربية، كما أنه يقدم طريقة معينة لاشتقاق الجذر من الجذع. محور هذه المنهجية هو القيام بتجميع الإحصاءات المتعلقة بالزوج جذر-كلمة، وتتلخص في الخطوات التالية:

أ) وضع لائحة صرفية لاستخراج الجذور من الكلمات،

ب) تكوين لائحة السوابق واللواحق،

ج) تقدير احتمال استعمال القواعد واحتمال ظهور السابقة واللاحقة.

هذا المحلل الذي لا يتطلب لوائح يدوية خاصة باللواحق والقواعد، يقتضي قائلين: يقوم الأول باستعمال الأزواج جذر-كلمة لوضع لائحة السوابق واللواحق ولتكوين الجذوع وأيضا لحساب الحالات التي تظهر فيها السابقة واللاحقة والجذع الصيغي. في حين يتعامل القالب الثاني مع خرج الكلمات العربية وذلك بتكوين التأليفات الممكنة بين السابقة واللاحقة والجذع الصيغي، ويربط هذا الخروج بالجذور الممكنة. فلائحة الأزواج جذر-كلمة يتم تكوينها يدويا ووضعها في قاموس، مما يخول لهذا القالب التعامل مع هذا الزوج كخرج. وبمقارنة الكلمة بالجذر يحدد النظام كلا من السابقة واللاحقة والجذع الصيغي. فإذا قدمنا لهذا النظام الزوج (وكتابهم، ك ت ب)، سيولد "و" كسابقة و"هم" كلاحقة، وفعال كجذع صيغي، وسينمي عدد تواردات السابقة "و" واللاحقة "هم" والجذع⁴.

⁴ هناك عدد كبير من المحللات التي لا يمكن أن نتناولها كلها في هذه الورقة، ولكن نريد أن نشير إلى أن محلل السعداني وحشيش يندرج في إطار ما يعرف في الأدبيات بالنظام الصرفي العربي القائم على خطوتين، ذلك أنه يتكون من محلل صرفي عربي يرتكز على قائلين أساسيين، هما: قالب التحليل، وقالب التوليد/الشكل. الغاية من القالب الأول هي استخراج الجذر والربط بين السمات الصرفية وخرج الكلمة غير المشكولة. ويبدأ التحليل بحذف السابقة واللاحقة وفصل اللواحق عن بعضهما البعض، ثم يتم الربط بين سمات الكلمة وسمات اللاصقة للتوصل إلى الجذع، ومن ثمة الجذر. ويتم في الأخير فحص مجموعة الجذور الممكنة وسماتها بالنظر إلى القاموس الحاسوبي العربي. وبذلك، فخرج قالب التحليل عبارة عن مجموعة الجذور العربية المشروعة وسماتها الصرفية. أما قالب التوليد/الشكل، فهو يتضمن مجموعة الجذور الممكنة مقرونة بسماتها الصرفية، يعتمد عليها المولد الصرفي العربي ليقوم بتوليد الكلمات (المصرفية) المشكولة. إضافة إلى ذلك تطبق بعض القيود والعمليات قبل الحصول على مجموعة التحاليل الممكنة، إذ ينبغي، مثلا، أن يكون هناك توافق بين بداية الكلمة ولائحة السوابق، وبين سمات الكلمة وسمات السابقة.

2. الأسماء العربية: ملاحظات عامة وقضايا مطروحة

للأسماء العربية خصوصيات كثيرة يتعلق البعض منها بخصائص اللغة العربية في حد ذاتها، بينما يتعلق البعض الآخر بالطبيعة المركبة للمقولات الاسمية. فهذه اللغة، كما هو معلوم، تكتب دون حركات قصيرة أثناء الكتابة، مما يصعب معالجتها آلياً ويجعل الاعتماد على قاعدة معطيات غير تامة وغير دقيقة أمراً صعباً للغاية، ولا يسعنا في تفادي الوقوع في الخطأ والإلمام بجميع النتائج المحتملة. فكلما نحو كتاب ملتبسة في وضعها هذا بين الكلمات المشكولة الآتية: *كتاب* و*كتاب* و*كتاب*⁵. الأمر نفسه يصدق كذلك على كلمة معلمة التي يمكن أن تكون *معلّمة* أو *معلّمة* وهلم جرا. هذه التفاصيل ذات أهمية قصوى لكونها تبرز الإمكانات الحركية المتاحة والمعاني المختلفة التي يمكن أن تنتج عن هذه الثقاليب الحركية التي لا تعكسها الكتابة. زد على هذا أن بعض الحروف الأصلية في الكلمة تكون مماثلة للواصق في حالات من قبيل *ليمون* أو *مكان* أو *ولد* أو *بلد الخ.* وهناك حالات أخرى يبلغ فيها الالتباس أشده، ومثال ذلك كلمة من قبيل *برد* التي تحتمل أن تكون *بَرْد* أو *بِرْد* باعتبار أن الباء سابقة (حرف جر) و*رَدّ* كلمة مستقلة (بمعنى جواب).

وتتخذ بعض الأسماء العربية الأخرى صوراً (سطحية) مختلفة متأثرة في ذلك إما بطبيعة الحركة الإعرابية الواردة، أو بطبيعة العناصر الملحقة بها. فالحرف الأخير من الاسم المهموز يكتب بأشكال متعددة بحسب محله من الإعراب: ماء أو مائه (في حالة الجر) أو ماؤه (في حالة الرفع)، كما أن نفس الحرف في الاسم المقصور يتحول إلى ألف (ا) أو ياء نحو: مرمى التي تصير *مرماه* أو *مرمياه* عندما يتم إقحام اللواحق، أو عندما تكون مصرفة في المثني، في حين يحذف

⁵ كلمة *كتاب* يمكن أن تكون جمعاً ل*كتاب*، ويمكن أن تعني مكان تعليم الصغار وجمعها *كتاتيب*.

الحرف الأخير من الاسم المنقوص نحو: محامي محام محامون. ومن ناحية أخرى تصير التاء المربوطة مبسوطة عندما تتلوها لاحقة، وتحذف في حالة جمع التكسير مثل: ورقة ورقته ورقات، كما أن اللاحقة *لل* لا تعبر عن أصلها الذي هو *ل* في الحالات العادية، إذ عندما تكون اللام الثانية أصلية، كما في كلمة من قبيل للغة، نقرأ بكيفيتين: للغة أو للغة. اللام الأولى في الكلمة الأولى هي في الأصل *ل*، وفي الكلمة الثانية تبقى على حالها.

الملاحظ أن بعض اللواحق الاسمية في اللغة العربية تتشابه من حيث الشكل وتختلف من حيث الوظيفة. فالواو يمكن أن يكون للعطف أو للابتداء أو للحال أو للمعية أو للقسم، كما أن الياء يمكن أن تكون للنسبة أو ضمير ملكية. أما الألف (ا) في آخر الكلمة فهو إما يمثل لاحقة العدد المثني في حالة الإضافة (وبالتالي يمكن أن يكون متلوا بلاحة)، أو يمثل الألف الحامل للتثنية الذي يجب أن يكون في نهاية الكلمة. التاء المربوطة بدورها يمكن أن تكون صنيفة أو علامة دالة على التأنيث أو غير دالة عليه، ولاحقة العدد المثني وجمع المذكر السالم تفقد النون في حالة الإضافة.

3. التصنيف الاسمي المعتمد: الموارد والأسس

تتطلب الأسماء معالجة خاصة من شأنها المساهمة في بلورة تصور جديد يضبطها ويحول حوسبتها وترجمة خصائصها ألياً، نقادياً للفوضى ودرءاً للأحكام الأولية التي قد يكون جزء كبير منها خاطئاً. فعلى المستوى التصنيفي للأسماء إلى طبقات فرعية، نعتقد أن أي تصنيف كيفما كان نوعه لا بد أن يكون مؤسسا، وإن بكيفية غير مباشرة، على معلومات لسانية دقيقة تأخذ بعين الاعتبار دور الحاسوبي

في التمثيل للعمليات المعقدة التي يمكن أن تقوم بها الآلة عندما يتعلق الأمر بموضوع متشعب اسمه اللغة. أول شيء، إذن، ينبغي الانطلاق منه هو عملية تصنيف الأسماء إلى أصناف تتقاطع في خصائصها العامة والبارزة (أي الشكلية)، وتتعارض في عدد وطبيعة اللواحق المسموح لها بالانضمام إليها، لما لهذه العملية من أهمية من الناحية العملية، خصوصا في المرحلة التي يحاول فيها الحاسوبي أن يجد ترجمة آلية للنظام المفترض تمكنه، بدقة وبساطة واقتصاد، من تحقيق الأهداف المرجوة. أما على المستوى المنهجي للتعامل مع النمط المقولي الاسمي، نشدد على ضرورة الانتباه لأمرين أساسيين، أحدهما هو أن النصوص العربية، التي من المفروض أن تشكل مجال عمل المحلل الصرفي، هي نصوص غير مشكولة، يلتبس فيها الاسم بالفعل وتتداخل فيها الأصناف الاسمية بعضها ببعض. والأمر الآخر هو أن عددا لا يستهان به من العناصر المدرجة ضمن لائحتي السوابق واللواحق تتحقق كحروف أصلية في بعض الأنماط المقولية الأخرى، وهذا الشبه الخطي/الشكلي غير الوظيفي يشرعن بعض العمليات التي تؤدي إلى نتائج لا قبل لها بما تزخر به الذاكرة الصرفية للغة العربية.

هناك عدة تصنيفات لغوية وغير لغوية للأسماء العربية مستمدة من التصنيفات التقليدية أو الحديثة على اختلاف أنواعها ومصادرها وخلفياتها النظرية، لكن لا ينبغي أن ننسى أننا بصدد المعالجة الآلية للغة العربية، وأن أي تصنيف عليه أن يراعي هذه الخصوصية. أصناف الأسماء التي اعتمدها في القاعدة محصورة فيما يلي: اسم الجنس، واسم العلم، واسم الآلة، واسم المكان، واسم الزمان، والتصغير، واسم الحدث، واسم الفاعل، واسم المفعول، وصيغة المبالغة.

الأصناف الستة الأولى هي أسماء ذوات باستثناء اسم الزمان،⁶ وهي تنقسم الخصائص الآتية: (أ) تقبل لاحقتي المثني وجمع المؤنث السالم، (ب) تجمع جمع تكسير فقط، (ج) التاء المربوطة ليست لاحقة، ولا تنفصل عن الكلمة، ولا تعبر عن التأنيث بالضرورة، (د) تختفي التاء المربوطة من الكلمة عند ما تتلوها لاحقة أو عندما تكون مجموعة جمع مؤنث سالم، (هـ) تقبل النسبة. وفي مقابل ذلك، يبدي اسم العلم سلوكا مخالفا من حيث أنه غالبا ما يرد في صورة صرفية واحدة قلما تتغير، لكنه لا يخرج عن إطار الخصائص السابقة الذكر شأنه في ذلك شأن اسم الحدث الذي لا يدل على الذوات، والذي لا يقبل أحيانا أن يثنى أو يجمع. وتختلف الأصناف المذكورة عن الثلاثة الأخيرة في كونها تملك الخصائص ذاتها: (أ) التاء المربوطة لاحقة للتأنيث دائما، (ب) النسبة غير ممكنة، (ج) جمع المذكر السالم ممكن.

بناء على ما سبق، قمنا بتجميع بضعة أصناف اسمية في طبقات كلما تشابهت في قبولها لنفس العدد من اللواحق، فكان أساس هذا التقسيم مستمدا من كون الأصناف الاسمية الموجودة في اللغة العربية تنتمي بالضرورة لإحدى الطبقات الثلاث الرئيسية الآتية:

• الطبقة الأولى، وهي التي لا تقبل لاحقة التأنيث *ة* ولاحقة الجمع المذكر السالم *ون/ين*، تتضمن الأصناف الاسمية التالية (وهي تشكل نسبة كبيرة من قاموس الأسماء): اسم الذات/الحدث/العلم/جمع التكسير.

• الطبقة الثانية، وهي التي لا تقبل لاحقة الجمع المذكر السالم *ون/ين*، ولاحقة المثني *ان/ين*، تتضمن اسم النوع فقط، وهو إما اسم ذات أو حدث.

⁶ لقد أدرجنا اسم الزمان تجاوزا ضمن الصنف الخاص باسم الذات، على أن نقوم بفصله عنه مستقبلا.

• الطبقة الثالثة، وهي الطبقة التي لا تقبل لاحقة النسبة، تتضمن الأصناف الاسمية التالية: اسم الفاعل/المفعول/صيغة المبالغة.

لقد أدرجنا جمع التكسير إلى جانب اسم الذات والحدث والعلم في الطبقة الأولى لكونه لا يخضع لقاعدة مطردة (في حالة الثلاثي)، ولكونه يتقيد بخصائص الطبقة التي ينتمي إليها. هناك بعض الأصناف الاسمية التي وردت في طبقتين، ونخص بالذكر هنا اسمي الذات والحدث الواردين في الطبقتين الأولى والثانية، ويرجع ذلك إلى أن بعض الأسماء المدرجة ضمن هذين الصنفين تقبل التاء المربوطة التي تكون صنيفة الوحدة، وهي لاحقة كما في قولنا: شجر-شجرة، رقص-رقصة. هذه الأسماء ينبغي أن تجمع في طبقة خاصة لا تتضمن اسم الوحدة.

4. الموارد اللسانية لقاعدة المعطيات اليدوية

1.1.4. القواميس

هناك ثلاثة قواميس يتضمن كل واحد منها عنصرا من العناصر الثلاثة المكونة للكلمة السطحية (أو الكلمة الدخل): السوابق واللواحق والجذوع. فقاموسا السوابق واللواحق يحتويان على كل اللواحق الاسمية البسيطة والمركبة الموجودة في اللغة العربية، وهي مشكولة وغير مشكولة، مفككة وغير مفككة ومصحوبة بمصروفة من المعلومات المتعلقة بها. ويقوم قاموس الأسماء على تقسيم ثلاثي الطبقات تدرج ضمنه كل المركبات الاسمية التي يمكن العثور عليها في اللغة العربية، وهو مرتب ألفبائيا على أساس الجذر (الصامت) الذي يحتضن قائمة من الكلمات المصنفة أسفله حسب الترتيب الألفبائي أيضا. هذه الكلمات مشكولة وغير مشكولة ومقرونة هي الأخرى بالمعلومات المتعلقة بالطبقة والصنف والصنف الفرعي، فضلا عن سمتي الجنس والعدد.

وضمن الأصناف الاسمية المعتمدة نجد أسماء ذات طبيعة خاصة لكونها تظهر في أشكال مختلفة، كما أشرنا إلى ذلك من قبل، يتعلق الأمر هنا بالمهموز والمنقوص والمقصور والأسماء الخمسة التي تشكل ما نسميه بالأصناف الفرعية.⁷

2.4. الجنس والعدد

يوسم الجنس المذكر على مستوى القاعدة بالنسبة لجميع الطبقات باستثناء الحالات التي تكون فيها الكلمة السطحية مصرفة في المثني المذكر أو جمع المذكر السالم، في حين أن الجنس المؤنث يوسم على مستوى القاعدة بالنسبة للطبقة الأولى وعلى مستوى اللاحقة بالنسبة للطبقتين الثانية والثالثة في المفرد والمثنى المؤنث وجمع المؤنث السالم.⁸ وفي ما يخص العدد نشير إلى أن جميع الكلمات موسومة بالعدد المفرد على مستوى القاعدة، ما عدا تلك المدرجة في الصنف الرابع، يعني جموع التكسير التي تكون موسومة على مستوى القاعدة دائما على أنها جمع. وعندما تكون الكلمة السطحية مصرفة في المثني والجمع السالم بنوعيهما، تستخلص سمة العدد المناسبة من نفس السمة الموجودة في اللاحقة (وهي سمة متضمنة في قاموس اللواحق).⁹

لدينا، إذن، ثلاث طبقات اسمية، وضمن كل طبقة هناك أصناف وأصناف فرعية موزعة على النحو الآتي: بالنسبة للصنف: 1= اسم الذات، 2= اسم الحدث،

⁷ نظرا لقلّة الأسماء الخمسة في اللغة العربية سنزود قاموس الأسماء بكل تمظهراتها الممكنة.

⁸ فكلمة من قبيل مدرسة أو دار موسومة في القاعدة على أنها مؤنث، بيد أن سمة الجنس المؤنث في كلمة

نحو معلمة أو شجرة أو معلمتان أو شجرات مستخلصة من نفس السمة في اللاحقة الواردة.

⁹ الهدف هو تقديم أكبر عدد ممكن من المعلومات حول عناصر الكلمة الدخلى المفككة، والاستغلال

المستقبلي لهذه المعلومات في إنجاز محلات تركيبية أو دلالية أو في وضع مترجمات آلية للخ.

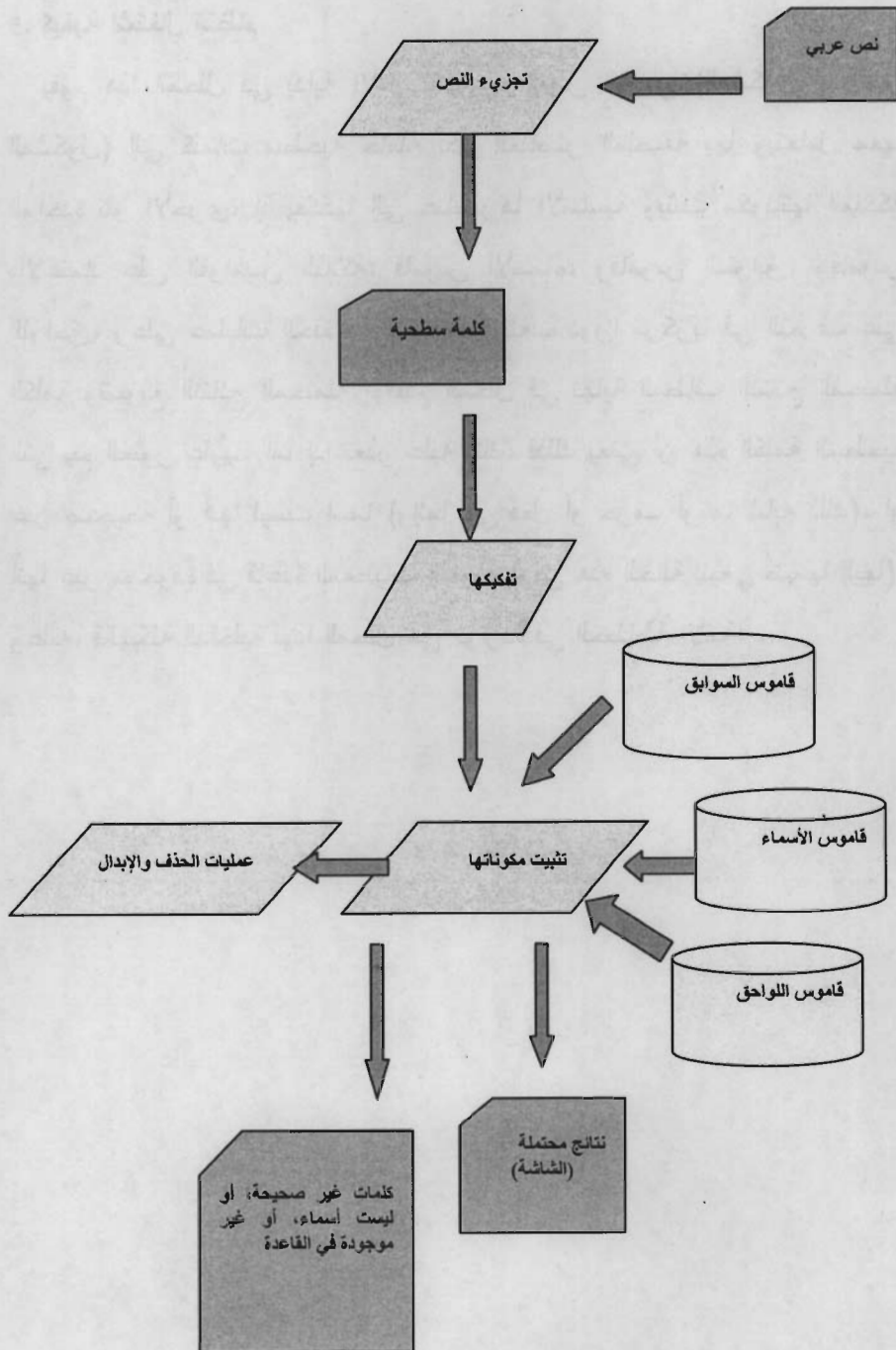
3= اسم العلم، 4= جمع التذكير، 5= اسم الفاعل، 6= اسم المفعول، 7= صيغة المبالغة. بالنسبة للصفة الفرعي: 1= المقصور، 2= المنقوص، 3= المهموز. بالنسبة للجنس: 1= المذكر 2= المؤنث. بالنسبة للعدد: 1= المفرد، 2= جمع التذكير.

نموذج من قاموس الكلمات المعتمد

الجنس	العدد	الصفة	الطبقة	الكلمة المشكولة	الكلمة	الجذر
						ك ت ب
1	1	2	1	كَتَبَ	كتب	
2	2	4	1	كُتِبَ	كتب	
1	1	1	1	كِتَاب	كتاب	
1	2	4	1	كُتِّبَ	كتاب	
1	1	1	1	كُتِّبَ	كتاب	
2	1	2	1	كِتَابَةٌ	كتابة	
1	1	1	1	كُتِّبَ	كتيب	
1	1	1	1	مَكْتَب	مكتب	
2	1	1	1	مَكْتَبَةٌ	مكتبة	
1	1	2	1	اِكْتَاب	اكتتاب	
2	2	4	1	مَكَاتِب	مكاتب	
1	1	5	3	كَاتِب	كاتب	
1	1	6	3	مَكْتُوب	مكتوب	
2	1	2	1	مُكَاتِبَةٌ	مكاتبة	
2	1	2	1	كُتِبَتْ	كتبة	
						ك ت ح
1	1	2	1	كُنِح	كنح	

5. كيفية اشتغال النظام

يقوم هذا المحلل في بداية الأمر بتجزئء النص العربي (المشكول أو الغير المشكول) إلى كلمات سطحية حاملة لكل العناصر الملتصقة بها ويتعامل معها الواحدة تلو الأخرى، إذ يفككها إلى عناصرها الأساسية ويثبت مكوناتها المفككة بالاعتماد على القواميس الثلاثة: قاموس الأسماء، وقاموس السوابق، وقاموس اللواحق، وعلى عمليات الحذف والقلب التي تلعب دورا مركزيا في التعرف على الكلمة وتسويغ النتائج المحتملة. ويقدم المحلل في نهاية المطاف النتائج المحتملة التي يتم العثور عليها. أما إذا تعذر عليه ذلك، فذلك يعني أن هذه الكلمة السطحية غير صحيحة أو أنها ليست اسما (وإنما هي فعل أو حرف أو ما شابه ذلك)، أو أنها غير موجودة في قاعدة المعطيات اللغوية (وفي هذه الحالة ينبغي ضمها إليها). وعليه، فالهيكلية الداخلية لهذا المحلل هي الواردة في الخطاطة الآتية:



يتم البحث عن كل كلمة سطحية في قاموس الأسماء أولاً قبل أن تفكك بالنظر إلى أحد القواميس المعتمدة، ثم تفكك على أساس قاموس السوابق (لنحصل على: سابقة/كلمة محتملة)، وتتكرر هذه العملية حتى تشمل جميع السوابق الواردة، وفي كل مرة يتم البحث من جديد عن ما تبقى من الكلمة في قاموس الأسماء. بعد ذلك تنطبق قاعدة التفكيك مرة أخرى، ولكن على أساس قاموس اللواحق هذه المرة (لنحصل على: كلمة محتملة/لاحقة) فيتم اللجوء إلى قاموس الأسماء للمرة الثالثة، وبشكل متكرر، للبحث عن ما تبقى بعد إزالة اللواحق بكيفية تدريجية على الشاكلة المشار إليها قبل قليل. وعند القيام بعملية التفكيك الثالثة والأخيرة سيتجرد الجذع من كل اللواحق العالقة به باعتماد القواميس الثلاثة (لنحصل على: سابقة/كلمة محتملة/لاحقة). في هذه المرحلة يتم احتساب كل التاليفات الممكنة بين العناصر الثلاثة المكونة للكلمة السطحية بالعودة المتكررة لقاموس الأسماء للتحقق من وجودها فيه.

6. عمليات القلب والحذف

لمعالجة جملة من التغيرات التي تطرأ على الكلمات السطحية نقترح هذه العمليات التي من شأنها أن تسعفنا في الإلمام بكل السيناريوهات الممكنة:

1) في حالة وجود ا (الألف) بعدها ثلاثة حروف على الأكثر نقلب ي (الألف المقصورة) شريطة أن يكون ما بعدها موجوداً في قاموس اللواحق، مما يعني أن البحث سينصب على الكلمة مرمى بدلاً من مرما الواردة في كلمة سطحية من قبيل مرماهم. أما في حالة وجود ياء بعدها ألف بمفرده أو متلو بحروف أخرى (أقصاها ثلاثة)، يحذف ما بعد الياء وتقلب ألفا مقصورة، وينصب البحث عن هذه الكلمة في قاموس الكلمات (وتحديداً في الصنف الفرعي 1 المتعلق بالمقصور). فإذا

ما وجدت، تكون عملية القلب صحيحة ويتم التحقق من أن ما بعد الياء هو بالفعل لاحقة استنادا إلى قاموس اللواحق، وإذا لم يكن موجودا، تلغى عملية القلب مباشرة.

(2) بالنسبة للمنقوص (محمم ومحامون)، يتم البحث للوهلة الأولى، وكالعادة، في قاموس الأسماء الذي لا يتوفر على ما يبدو على كلمات من هذا القبيل. عندئذ تتم إضافة ي إلى هذه الكلمة بعد إزالة اللاحقة (إذا كانت موجودة) وقبل البحث عنه في الصنف الفرعي 2 الخاص بالمنقوص. وإذا ما تم العثور على الكلمة في قاموس الأسماء منذ الوهلة الأولى، فلن تضاف الياء إلى الجذع مطلقا.

(3) تبديل الشكل و والشكل نـ بالشكل الأصلي ء والبحث عن الكلمة في الصنف الفرعي 3 المتعلق بالمهموز إذا كانا مثلوين بحروف موجودة في قاموس اللواحق، وإلا فهي أصلية ولا ينبغي تبديلها.

(4) في حالة الكلمات التي تبدأ بـ ال، تحذف اللام الأولى ويتم البحث عن البقية في قاموس الأسماء، فإذا وجد فهذا يعني أن هذه اللام يمكن أن تكون ل بمفردها أو لل (التي تفكك استنادا إلى قاموس السوابق على النحو الآتي: ل/ال)، وأن اللام الثانية هي أصلية. أما إذا تعذر وجودها، تحذف لل ويتم التحقق من وجود الباقي في قاموس الأسماء.

(5) كل كلمة تنتهي بالتاء المبسوطة ت بعد تجريدتها من جميع اللواحق يتم البحث عنها في بداية الأمر في قاموس الأسماء، وسواء وجدت أو لم توجد تحولت إلى ة، مما يتطلب القيام ببحث جديد عن هذه الكلمة الجديدة في قاموس الأسماء. وفي حالة جمع المؤنث السالم الذي ينصب على الأسماء التي تنتمي للطبقتين الأولى والثانية، يتم حذف لاحقة العدد المؤنث من الكلمة وإضافة تاء مربوطة لما تبقى منها قبل البحث عنها في قاموس الأسماء.

6) تكون عملية التفكير مشروعة فقط إذا كان ما تبقى من الكلمة بعد إزالة اللواصق لا يقل عن حرفين.

7) أية كلمة تنتهي بالياء ينطبق عليها ما يلي: إذا كانت موجودة في الصنف الفرعي 2، فإن هذه الياء أصلية ولا يمكن أن تكون لاحقة.

7. مناقشة وتقييم

يندرج العمل الذي قدمناه في هذه الورقة ضمن مقارنة التحليل الصرفي، ويستمد بعض أسسه من العمل الذي قدمه بكوالتر. وقد يبدو للوهلة الأولى أن المحلل الصرفي الذي نقترحه مشابه لذلك المقترح من قبل بكوالتر، بيد أن التفاصيل تثبت أنه مخالف له شكلا ومضمونا. وقبل خوضنا في تفاصيل الفرق الشاسع الموجود بين المحللين، يجدر بنا أن نناقش الأعمال التي تطرقنا إليها في الفقرة الأولى بتقديم بعض الملاحظات حولها، وخاصة ما يتعلق بالأسس اللسانية التي تقوم عليها. ولعل الملاحظة العامة الأساسية الأولى التي يمكن ان نبديها بخصوص هذه الأعمال إجمالا، هي عدم مسايرة الجانب اللساني للحوسبة، إذ يقع معظمها في العديد من الأخطاء التي تنعكس سلبا على دقة النتائج المنتظرة. ولا بأس أن نقدم بعض مظاهر إهمال العمل اللساني في هكذا محاولات.

المجذعات الخفيفة في تعاملها مع اللواصق على ضربين: أحدهما يعتمد لوئح تتضمن مجموعة محدودة من اللواصق الموجودة في اللغة العربية، وهو بهذا يمثل المجذعات الضعيفة التي لا تحيط بكل الإمكانيات التي تتيحها اللغة المراد تحليلها. أما الضرب الثاني فهو يتضمن من اللواصق أكثر من اللازم وأكثر مما يوجد في اللغة العربية، وهو بهذا يمثل المجذعات القوية التي تنتج أكثر مما تخوله اللغة

المراد تحليلها. وعليه، يفشل الضرب الأول من المجذعات الخفيفة في بلوغ الجذع عندما لا تكون اللاصقة واردة في اللائحة، كما تعاني، شأنها في ذلك شأن الضرب الثاني، من مشكل مشابهة بعض اللواصق للحروف الأصلية للكلمة العربية.

عندما نتأمل لائحتي السوابق واللواحق اللتين يقوم عليهما مجذع شين، يمكن أن نتصور عدد النتائج الخاطئة التي يمكن أن يقدمها. فهاتين اللائحتين تتضمنان عناصر لا صلة لها بما يوجد في اللغة العربية لا من قريب ولا من بعيد. نفس اللوائح تقريبا يقوم عليها مجذع TREC 2001 ومجذع TREC 2002 (وهما صيغتان معدلتان لمجذع لاركي).¹⁰ ولما تعامل شين مع جمع التكسير اعتبر ان (أ) و(ا) في كلمة نحو *أطفال* هي لواصق: (أ) سابقة، و(ا) واسطة. واستادا للعمليات التي يعتقد أنها ضرورية يمكن أن يتم حذف حروف أصلية من كلمة مثل *بالكم*، لأن الشرط في حذفها هو كون طول الكلمة هو خمسة أحرف مع وجود السابقة المركبة في اللائحة المناسبة. وبخصوص حذف الواو من كلمة تتضمن أربعة أحرف لكونه يعد في مثل هذه الحالات حرفا أصليا، نشير إلى وجود كلمات عربية من قبيل *ولد-وليمة*.

وفي مقابل ذلك، ينهج المحلل الصرفي لخوجة نهجا معقولا لبلوغ الجذر، يتجلى في اعتماده على الوزن. وعندما نتحدث عن الأوزان تتبادر إلى الذهن مسألة العلة التي تتمظهر بكيفيات مختلفة بحسب طبيعتها وطبيعة النمط المقولي للكلمة والطبقة التي تنتمي إليها. لذا لا بد من اعتماد القواعد الصرفية الضابطة لهذه المسألة في اللغة العربية سواء تعلق الأمر بالأفعال أو الأسماء. ويبقى السؤال المطروح هو

¹⁰ انظر الهامش (1).

الآتي: ما الغاية يا ترى من بلوغ الجذر بدل الجذع إذا لم نهتم بالكلمة الأصلية؟ فإذا قدمنا لهذا المحلل كلمة سطحية/نصية من قبيل *يقاتلك*، سوف يزيل السابقة (ي) واللاحقة (ك) ليحصل على *قاتل* التي سيقيسها فيما بعد بالوزن الذي يملك نفس الطول والذي هو *فاعل*، وبالتالي يسهل عليه بلوغ الجذر الذي هو *ق ت ل*.

غير أننا لا نعرف شيئا عن الكلمة التي تم الحصول عليها بعد إزالة اللواصق: هل هي فعل أو اسم، وهنا بالتحديد يطرح مشكل تداخل الأوزان الذي يجعل مثل هذا التمييز مستحيلا. وعندما لا يجد الجذر في قاموس الجذور يرجع الكلمة الدخل دون أن يتحقق من أنها كلمة موجودة في اللغة العربية أو غير موجودة فيها (لغياب قاموس الجذوع). أما النظام المتبع في إزالة اللواصق فهو بدوره غير فعال لاعتمادا على اللوائح فقط، مما يؤدي إلى إزالة بعض الحروف الأصلية في الكلمة والتعامل معها على أنها لواصق، وذلك لعدم وجود استراتيجية واضحة قادرة على إقامة تمييز دقيق من هذا النوع.

ونظرا لكوننا نشغل في إطار مقارنة التحليل الصرفي، ونظرا كذلك للتشابه السابق الذكر بين المحلل الذي نقترحه والمحلل الذي اقترحه بكوالتر، نود أن نوضح أمرا هاما يمكن إجماله في القول إن محللنا يختلف جملة وتفصيلا عن محلل بكوالتر، إذ العلاقة الوحيدة القائمة بين المحللين لا تتعدى اعتمادهما على ثلاثة قواميس يدوية تمثل العناصر الثلاثة الأساسية للكلمات السطحية. وما عدا ذلك، تبدأ مظاهر الاختلاف الكثيرة التي نستهلها بالإشارة إلى أن محلل بكوالتر يكتفي بالبحث عن الجذع، بينما يقوم محللنا بالبحث عن الجذع والجذر ويقدمهما معا في صفحة النتائج.

إن تصنيف الكلمة الوارد في هذا المحلل هو: الاسم، والفعل، وما يسميه بكوالتر بالكلمة الوظيفية، ويقصد بها الأدوات والضمائر والكلمات الأخرى التي ليست فعلا ولا اسما. فكل جذع تسند له مقولة صرفية معينة، وهذه المقولات كثيرة ومتنوعة نذكر منها، على سبيل المثال لا الحصر، الاسم، والاسم المثنى، والاسم المثنى المعرف، واسم العلم.. الخ، والفعل الماضي، والفعل المضارع وغيرهما. وهناك قائمة من السوابق واللواحق التي يمكن أن تلتصق بهذا الجذع أو ذلك. وتوخيا للدقة نقتصر على مناقشة تصنيف بكوالتر للأسماء، بحكم أن هذه الأخيرة تمثل الجانب الصعب والغامض في أية معالجة آلية كيفما كانت، وبحكم كذلك أنها لم تتلحقها من الاهتمام إلا مؤخرا. ونجد في هذا التصنيف حوالي عشرين صنفا اسميا أو مقولة صرفية، على حد تعبيره، تتحدد بموجب خصائصها الإلصاقية الصرفية للجذع الاسمي الذي يتعلق به الأمر.

ملاحظتنا على هذا التصنيف متعددة الجوانب والأحياز، نستلها بالعدد الكبير للأنماط الاسمية التي يمكن إدماج بعضها في بعض للحصول على طبقات كبرى. فعندما يقول بكوالتر إن هناك اسما مثنى كصنف من الأسماء، فهو يعني بذلك أن الجذع الاسمي يأخذ لاحقة المثنى، أو عندما يقول إن هناك صنفا آخر هو الاسم الذي ينتهي ب (ت)، فإنه يعني بذلك أنه يأخذ لاحقة الجمع المؤنث. وعليه، من الجلي أن الأساس الذي يقوم عليه هذا التصنيف غير متين مادام مستخلصا مما هو سطحي ولا يعكس بالضرورة الخصائص الحقيقية والعميقة للأسماء، ومن الطبيعي أن تكون إحدى نتائجه اقتراح هذا الكم الهائل من الأصناف.

من بين الملاحظات الأخرى التي ينبغي أن نشير إليها نذكر بعض الصعوبات التي واجهت بكوالتر في وضعه لجملة من المقولات في غير موضعها، كوضعه،

مثلا، لاسم من قبيل "أسرة" في صنف الأسماء التي تنتهي بالتاء المربوطة والتي تجمع بإضافة لاحقة الجمع المؤنث (ت)، وهذا بالطبع أمر غير صحيح. كذلك هو الشأن بالنسبة لإدراجه لاسم ك "كتابة" في صنف الأسماء التي تنتهي بالتاء المربوطة والتي تكون في الأصل مفردا مؤنثا وغير قابلة لأن تجمع. هذا علاوة على أنه لا يقيم حدا فاصلا بين الأسماء والصفات والأعداد، إذ يدرجها جميعها ضمن الأصناف الاسمية التي اقترحها في نموذج.

يلاحظ كذلك أن صاحب هذا النموذج لم يكن موقفا كثيرا في التعامل مع السوابق التي أغفل ذكر البعض منها، كما هو الحال بالنسبة لألف الاستفهام، وهي سابقة حسب منطق بكوالتز (لعدم وجود فراغ يفصلها عن الاسم الذي ترد معه في مثل قولنا: "محمد نائم؟"). وهو كذلك لم يفصل بأي شكل من الأشكال في مسألة بعض السوابق التي لا تختلف من حيث شكلها وإنما تختلف من حيث وظيفتها، كالواو الذي يعد في تصور بكوالتز أداة عطف، والصواب أنه يمكن أن يكون للعطف أو للمعية أو للقسم أو للحال، والياء التي يقول إنها للنسبة، يمكن أن تكون صنيفة.¹¹

¹¹ للواو في اللغة العربية وظائف عدة أوردناها أعلاه، ونمثل لها، تباعا، بالجمل الآتية:

(أ) ودع الطفل المدرسة والمدرسين

(ب) تجولت والرجل

(ج) والله

(د) عاد الطفل وهو يحمل رصاصة

الياء أيضا تتوزع وظيفتها في هذه اللغة بين الترميز للنسبة والترميز لما يعرف في الأدبيات اللسانية بالصنيفة، وهي بكل بساطة الياء التي تدخل على اسم نحو "مصر" فيصير "مصري" (علاوة على أنه تكون دالة على الملكية في قولنا "كتابي"). وينتج، بطبيعة الحال، عن الخلط بين هذه الوظائف عدم تقديم أوصاف دقيقة للعناصر المفككة في الخرج.

وبالنسبة للتاء المربوطة فهو يعتبرها، كباقي الأعمال التي أشرنا إليها، لاحقة تحذف من جميع الكلمات باستثناء الأسماء التي تكون فيها هذه التاء غير دالة على التانيث (كما في *طلحة*)، والتي يضعها في خانة خاصة حتى يتعرف عليها النظام لكي لا يقوم بنفكيكها لأنها ليست لاحقة. التاء المربوطة في التحليل الذي قدمناه في هذه الورقة على ضروب عدة يتعامل معها النظام وفق الطبقة التي الطبقة الاسمية التي تنتمي إليها وليس وفق الصنف، ودون الحاجة إلى عزل أسماء عن غيرها لمجرد أنها تمثل حالة خاصة.

ويستعين محلل بكوالتر بعدد من قيود التوافق التي تلغي جملة من التأليفات غير الممكنة في اللغة العربية بين عناصر الكلمة النووية، وهي قيود لا أثر لها في تحليلنا لأن التوافق الحقيقي في نظرنا يتم بالعودة المتكررة إلى قاموس الجذوع في كل مرة طبقت فيها عملية التفكيك المرتكزة على القواميس الثلاثة الأساسية. فوجود الجذع في هذا القاموس هو بمثابة عملية فحص أولى يتم بموجبها التأكد من أن عملية التفكيك المقامة صحيحة، ثم تأتي بعد ذلك عملية فحص ثانية تتم بالنظر إلى قاموسي السوابق واللواحق للتحقق من أن اللواحق المزالة موجودة في اللغة العربية أو غير موجودة فيها، وأنها مسوغة في الطبقة الاسمية التي تنتمي إليها الكلمة أو غير مسوغة فيها، حتى تكون عملية التفكيك مشروعة أو تلغى فلا يقدم خرجها كنتيجة محتملة.

وعموماً، فكل النواقص التي تمت الإشارة إليها أعلاه شائعة في الأدبيات التي تعاملت مع الكلمة العربية، وهي ناتجة بالدرجة الأولى عن الأسباب الآتية:

- غياب التدقيق اللساني المحض للمعطيات والمصادر اللغوية المعتمدة،

- التركيز على الصياغات الصورية (الخورزمات) التي تطبع مجال الحوسبة،
- الزعم المسبق أن صرف اللغة العربية في غاية الصعوبة ويطرح مشاكل عويصة لا يمكن ضبطها،
- عدم مسايرة القفزات النوعية التي يعرفها الحقل اللغوي الحديث، وعدم الاطلاع على نتائجه والنقاشات العميقة الدائرة حوله.

خاتمة

أثناء المعالجة الآلية للغة العربية لا بد من مراعاة الدقة وتوخي الحذر لكثرة تمظهراتها الناتجة بالدرجة الأولى عن كونها غير مرفقة بالحركات القصيرة. وقد قدمنا في هذه الورقة ما نعتقد أنه أساس لساني متين مكنا من التوصل إلى تنظيم معقول للعناصر المدرجة في قاعدة المعطيات اليدوية لبلوغ الكفاية والشمولية والإحاطة بكل الاحتمالات الممكنة. فكان النظام المقترح أدنويا من ناحيتين: أ) اعتماد قاموس مصنف للأسماء لا يتضمن الكلمات السطحية التي يمكن اشتقاقها بإضافة لواصق إليها، ب) اعتماد النظام على إجراءين تكراريين مرتبطين بعدد محدد من عمليات الحذف والإبدال، هما: التفكيك والعودة للقواميس (قاموس الأسماء على وجه الخصوص).

المراجع

- Buckwalter, T: Qamus. Arabic lexicography. <http://www.qamus.org/>
- Chen, A. and Gey, F.: 2002, Building an Arabic stemmer for information retrieval. In TREC 2002. Gaithersburg: NIST, pp 631-639,.
- Darwish, K.: 2002, Building a shallow morphological analyser in one day. ACL Workshop on computational Approaches to Semitic languages.
- Diab, M.: 2004, Arabic SVM Tools.
http://www.stanford.edu/~mdiab/software/arabic_SVMTools.tar.gz.
- El-sadany.T.A&Hashish.M.A.: 1999, An Arabic Morphological System. IBM Systems Journal. Vol 28-NO 4.
- Khoja, S.and Garside, R.: 1999, Stemming Arabic test. Computing Department, Lancaster University.
<http://www.comp.lans.ac.uk/computing/users/khoja/stemmer.ps>
- Larkey, Leah, S., Ballesteros, Lisa, and Connell, Margaret: 2002, Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis In Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2002), Tampere, Finland, pp. 275-282.
<http://www.ldc.upenn.edu/Catalog/>.
- Larkey, Leah, S., Ballesteros, Lisa, and Connell, Margaret. Light Stemming for Arabic Information Retrieval.
www.mtholyoke.edu/~lballest/Pubs/arab_stem05.pdf.
- LDC, Linguistic Data Consortium. Buckwalter Morphological Analyser. Version 1.0, LDC2002L49, 2002.