مجلة المعجمية - تونس

ع 9-10

1994

المدقّق الإملائيّ العربيّ :
إنجاز قاعدة معلومات إملائيّة عربيّة لتوليد معجم مدقّق إملائيّ للنّصوص على الحاسوب

بقلم : عبد الرّزاق بنّور

مـــدخـــــل :

0 - 1. يتنزّل هذا العمل في نطاق اللسانيّات الإعلاميّة وبصفة أدقّ في الاختصاص المصطلح عليه بـ" الصّناعات اللّغويّة" (les industries de la langue)، في ميدان المعالجة الآليّة للّغة المكتوبة.

0 - 1 - 1. ليس من الغريب أن يكون هذا النّوع من البرامج نادرا في العربيّة إلى حدّ يقارب العدم مع أنّه متوفّر توفّرا ظاهرا في جلّ اللّغات الأوروبيّة، فإنّ لغة الإعلاميّة ليست اللّغة العربيّة، التي تعسر معالجتها الصّوريّة (traitement formel) عسرا شديدا لكثرة طرق اشتقاق الكلمات فيها. وبالفعل فإنّ التعقيد الّذي تتّسم به اللغة العربيّة في الصّياغة الإملائيّة الكتابيّة (orthographique) والتّصريفيّة (flexionnelle) يفسّر بعض الشّيء صعوبة بناء معجم إملائيّ بالمفهوم العامّ (ككتاب على الورق) أو بالمعنى الخاصّ (كمدقّق إملائيّ، أي كبرنامج الكتروني).

0 - 1 - 2. فعلى مستوى اللّغة العربيّة يتطلّب البحث عن كلمة ما في المعجم حدّا أدنى من المعرفة اللّغويّة وحتّى اللّسانيّة في بعض الأحيان (مثل قواعد الاشتقاق، ونوع التكوين، وما إذا كانت الكلمة التّي يتمّ البحث عنها ذات جذر ثلاثيّ أو جذر رباعيّ ...)، فنحن لا نجد في أيّ من المعاجم العربيّة المرتّبة بحسب الجذور كلمة "مرتجف" مثلا إلاّ بالرّجوع إلى أصل تكوينها واشتقاقها، أي إلى جذرها "ر ج ف"، وهذا ليس

بالأمر الهيّن بالنسبة إلى المستعملين المبتدئين من الأطفال ومن غير النّاطقين باللّغة العربيّة الّذين هم أكثر النّاس حاجة إلى المراجع والمعاجم والتّثبّت الإملائيّ، وهنا تبدو لنا أهمّيّة المدقّق الإملائيّ في اللّغة العربيّة إذ يمكن لنا مثلا بفضل هذه الوسيلة التّثبّت آنيًا من صحّة كتابة كلمة ما دون الرّجوع إلى أصلها وجذرها الخ..

0 - 1 - 3. وأمّا عمليًا فإنّ جدوى هذا المعجم الإملائيّ - بقطع النظر عمّا له من فائدة في حدّ ذاته باعتباره معينا على الكتابة (على الأقل كرافد نفسانيّ، لأنّنا عندما نتأكّد من صحّة ما نكتب نقدّم ما كتبنا ليقرأ مطمئنّين) - تكمن في تغيير نظرة المستعمل إلى الحاسوب الّذي لا يرى فيه في الغالب أكثر من آلة كاتبة متطوّرة. ويمكن أن ينطبق ما سبق على المستعملين من ذوي اللّغات التي لا تفتقر - افْتِقَارَ العربيّة - إلى معاجم إملائيّة، فحريّ بلغة مثل لغتنا أن يُوَفَّرَ لها الحاسوب ما لم تتمكن منه بالطرق التقليديّة. وبالإضافة إلى ما أسلفنا من الأدلّة فإنّ المدقّق الإملائيّ يعتبر لبنة أخرى تضاف إلى مشروع منظومة المعالجة الآليّة الشّاملة للّغة العربيّة.

0 - 2. لقد كانت مهمّتنا تنحصر في إنجاز قاعدة معلومات معجميّة يعتمدها المدقّق الإملائيّ المندمج(correcteur orthographique intégré) في "وينتكست"(Wintext) (برنامج منسّق نصوص من انتاج وينسوفت (Winsoft))، وكانت مهمّتنا هي التّبويب وما بعد التّبويب والتّصنيف والتّنسيق مع كلّ الألغوريتمات (algorithmes) التي تتمّ إعادة كتابتها في لغة البرمجة المعدّة للغرض من قبل مؤسّسة نويتيك سرفيس (الأمريكيّة Circle Noetic services) والّتي تعتمد لتوليد القاعدة. وكانت مهمّتنا أيضا تتمثّل في امتحان مدى صلاحيّة المدقّق الإملائيّ المنجز بالاعتماد على القاعدة المولّدة، مع فحص تطابق الاقتراحات الّتي يقدّمها المدقّق باعتماد التقارب الصّوتيّ وتلاؤمها.

أ - التّقييدات المرتبطة بالمدقّق الإملائيّ بصفة عامّة :

إنّ بعض المعطيات التي تبدو بديهيّة قَيْقع تجاهلها في إنجاز المعاجم القياسيّة(dictionnaire analogique) وحتّى الإملائيّة إن وجدت، لا يمكن بأيّ حال تجاهلها عند إنجاز مدقّق إلكتروني. فمثلا نجد أنّ الرّصيد الإملائيّ للغة الفرنسيّة لاروس(1) يكتفي

(1) Le vocabulaire orthographique, Larousse, 1988.

بالتّنصيص على الجمع بالنّسبة إلى الكلمات الّتي يمكن أن تكون محلّ شكّ ولا يذكر أيّ شيء عن الصّيغ أو الأسماء التي يعتبرها خاضعة للقاعدة أو ليست محلّ شكّ، وكذلك الشّأن بالنّسبة إلى الأفعال فيكتفي بذكر المنوال (le modèle de conjugaison) الذي يندرج الفعل تحته في التّصريف ويبقى على المستعمل استخراج الصّيغ الملائمة. ولكن خاصيّة المعجم المدقّق، خلافا لما ذكر، هي أنّه يحتوي بصفة صريحة على الصّيغ والأشكال وحتّى الصّيغ الأكثر شيوعا واستعمالا وبداهة وإن أدّى ذلك إلى التّكرار المملّ، ولهذه الخاصيّة تأثير مباشر في بعض اللّغات الّتي تمثّل لها هذه الضّرورة العمليّة مشكل تضخّم مَرَضي(2) .

وبنفس الحدّة فإنّ بناء قاعدة معجميّة لمدقّق إملائيّ يقابله سؤال فيتغنشاين (Wittgenstein) (3) "إلى أيّ حدّ؟"، وتواجهه - خلافا للمعاجم العامّة العادية - مشكلة إدخال الألفاظ المبتذلة أو الجهويّة أو الإقليميّة في قاعدة المعلومات مع الكلمات التي ترفضها المذهبيّات أو يمنعها الحياء والقيود الاجتماعيّة العقائديّة مهما يكن نوعها.

ويتلخّص المأزق عمليّا فيما يلي : كيف يمكننا تكوين قاعدة ذات حجم متناه وغير محظور مع وجوب تكوين قاعدة مستوفاة وأكثر شموليّة مع مراعاة النّزعة التكاثريّة التّي يميل إليها نظام الاشتقاق العربيّ ؟

وقد أجبرنا بذلك على التخلّص من شرط الاستقصاء لأسباب فنيّة مرتبطة بطبيعة اللّغة العربيّة وتعقيداتها التصريفيّة(4). ولم يكن أمامنا اختيار آخر غير التّخلّي عن هذا الشّرط في الإصدار النّموذجي وتمكين المستعمل من تكوين معاجم خاصّة (بالمستعملين) بقدر ما تسمح به المساحة المتوفّرة على قرص التّخزين بالحاسوب.

أ - 1. لم يكن في نيّتنا تمثيل العربيّة الفصحى القديمة بل تمثيل لغة المستعمل العادي في العالم العربيّ الحديث / المعاصر. ولكي لا نعيد المشي في طريق قد سلك اعتمدنا الصّيغ الواردة في الأوساط المدرسيّة على اختلاف أشكالها. وكلّ صيغة لا يحتمل أن

(2) انظر أسفله وما ينتج من تضخّم في حجم المعجم الالكتروني.

(3) "Où s'arrêter?", v. L.Wittgenstein, Investigations philosophiques, Idées/ Gallimard .

(4) انظر أسفله مثالا على بعض هذه التعقيدات.

يعترضها متكلّم باللّغة العربيّة مثقّف في الكتب المدرسيّة كالمصطلحات الفنيّة أو التّقنيّة المتخصّصة والعبارات المنبوذة أو المحرّمة دينيّا أو اجتماعيّا تكون مسبّقا مغربلة وتكون قاعدتنا بفعل ذلك مضبوطة ومقنّنة. ولكنّنا لا نعتبر ذلك من الميزات أو الحسنات لأننا لا نجهل أن أحسن قاعدة معجميّة (على الأقل القاعدة المكوّنة لهذا الغرض) هي الّتي تقترح أكبر عدد ممكن من الأشكال الواردة سواء أكانت من المستعمل أم من الممكن المهمل، وأنّ الاستيفاء هو معيار النّجاح لمثل هذا العمل..

ومن هذا المنظور انطلقنا من الرّصيد الاساسي(5) الذي كوّنه فريق بحث مغاربيّ من تونس والجزائر والمغرب انطلاقا من دراسات وتحقيقات وتجارب ميدانيّة. وبما أنّ هذا الرّصيد لا يتعدّى 4000 كلمة بما فيها الأفعال فإنّ جرد الكتب المدرسيّة التّونسيّة منها والعربيّة مكّننا من أن نقارب 50 ألف لفظة مدوّنة. ولو لم نسقط المهمل (أو الغريب المستعمل في يومنا هذا) من العربيّة الفصحى لتحصّلنا على عدد كبير من الكلمات. وهذا العدد الكبير من الكلمات بمختلف تصاريفها (نظرا إلى الخاصيّات التصريفيّة العربيّة التي ذكرنا) سينتج حتْما قاعدة معجميّة لها ما لها من الحجم بحيث تستعصي معالجتها إذا أدخلنا فيها كلّ ما نعرف من اللّغة. ويكفي في هذا السّياق أن نذكر صيغة اشتقاق "النّسبة" : فإنّها في العربية ذات قاعدة مفتوحة ولذلك فإنّها تُعطي دون قيد : "قرطبة ← قرطبي و "تونس ← تونسي"، وبدون أي قيد أو تقييد "الجاحظ ← جاحظيّ" و "أرسطو ← أرسطي" و "فَ ← فائي"، الخ... وليست هناك إذن حدود لتوليد النّسب انطلاقا من الأسماء والأعلام وحتّى من الأدوات.

ويستفحل التّكاثر برسم الكلمات المعرّبة الّتي ليست مقنّنة بعد فنلاحظ وجود عدّة رسوم واردة لكلمة مثل (electronic/ électronique) فترسم ألكتروني أو الكتروني أو

(5) الرّصيد اللّغوي الوظيفي (1976)، قائمة في الكلمات أعدتها اللّجنة الدّائمة للعربيّة الوظيفيّة وهي تضمّ معهد الدراسات والبحوث للتعريب، جامعة محمد الخامس ومعهد اللّسانيّات والصّوتيّات بالجزائر ومركز الدّراسات والأبحاث الاقتصاديّة والاجتماعيّة بتونس (قسم اللّسانيّات). وتحتوي هذه القائمة على قرابة الـ4000 كلمة.

الكتروني أو ألكتروني (6).

ومن البديهي - نظرا إلى كلّ هذه الحالات - أن يقترن مثل هذا العمل وجوبًا بموقف لسانيّ فيما يتعلّق بحدود القاعدة المعجميّة وباختيار الكلمات (وأقلّها مشكل الألفاظ الإقليميّة أو المحليّة (7) الّتي يجب دراستها حالة حالة).

ب - معطيات نظريّة تتعلّق باللّغة العربيّة (الخاصيّات التصريفيّة للغة العربيّة)

إنّ اللغة العربيّة معقّدة أكثر من أيّ لغة أخرى إلى درجة أنّها تكاد تستعصي معها أيّ معالجة آليّة. فزيادة على ما فيها من تعقيدات نحويّة لها خاصيّات صرفيّة (morphologique) وإملائيّة (orthographique) تجعلها صعبة الدّراسة من النّاحية الإعلاميّة الصّوريّة (formelle) بصفة خاصّة.

ب - 1 . تصنيف نوعي :

تقع اللّغة العربيّة في التّصنيف النّوعي التكويني (typologie généalogique/ genealogical typology) المبنيّة بصفة عامّة على نظام التّصريف(8) ضمن اللّغات التّأليفيّة (synthétique) (مقابلة باللّغات التّحليليّة (analytique)، ومع ذلك فإنّها من بعض النّواحي

(6) يطرح مشكل الكلمات المقترضة نفسه بالفعل لأنّ الدول النّاطقة بالعربيّة تنهل من مصادر لغويّة مختلفة، فالبعض فرنكوفونيّ والآخر انجلوفونيّ وبعضهم له احتكاك بالإسبانيّة والآخر بالايطاليّة. . . وبذلك فإنّ نفس الكلمة المقترضة تنطبع باللّهجة الأصليّة فتنطق وتكتب بحسب ذلك. فقد يقترض بلد عربيّ فرنكفُونيّ كلمة بينما يقترح لها بلد فرنكوفونيّ آخر ترجمة فمثلا نجد "إطارات" تتنافس مع "كوادر" وهو مثال نموذجيّ للمعاملة غير المقنّنة أو المنسّقة للتّوليد المعجميّ.

(7) إنّ المشكل المطروح في العربيّة اليوم هو أنّ اللّغة "الحديثة" تحتوي على كلمات إقليميّة يجب أخذها بعين الاعتبار. ولكن ما الّذي يبرّر الاحتفاظ بكلمة مثل "طابور" ورفض كلمة مثل "شنطة" والاثنتان مصريّتان.

(8)مع أنّ ذلك لا ينطبق على بعض التصنيفات (مثل Tesnière, Éléments de syntaxe structurale (1959), Klincksieck. الذي يتبعُ Schmidt (1926) وينطلق من "المعلم النّمطي" (trait de structure/structure feature). انظر مثلا لزيادة التفاصيل P.Ramat, Typologie linguistique. PUF. Paris (1985)

كأنّها لغة تحليليّة. فمثلا تحدّد بعض العلاقات النّحويّة بواسطة كلمات وظيفيّة مستقلّة كـ"في" الظّرفيّة و"ثمّ" الزّمانيّة و"كـ" التشبيه. على أنّنا نلاحظ أنّه لا تنتمي إلاّ قلّة من اللّغات إلى صنف واحد دون غيره. ولذلك فإنّنا نَرى في كلّ لغة هيمنة مظهر من المظاهر مع وجود هيئات أخرى تجذبها إلى الأصناف المقابلة. ويمكننا القول بالنّسبة إلى العربيّة إنّها تظهر هيمنة تأليفيّة(9). وتكون اللّغة التّأليفيّة إمّا لغة متصرّفة وإمّا لغة لاصقة، ويكون الإلصاق أحد المظهرين اللّذين يميّزان اللّغات التّأليفيّة التي تبدو في اعتماد اللّغة (في بناء كلماتها وعلاقاتها النّحويّة) على إضافة زوائد إلى الجذور.

وتكون الكلمات في هذا النّوع من اللّغات عبارة عن مجموعة من الوحدات لها معان واضحة ومستقلّة وذات وظائف محدّدة (ومثال اللّغات اللاصقة : اليابانيّة والتّركيّة)، وتكون بعض اللّغات أكثر إلصاقا من غيرها ممّا يجعل من الممكن تدرّج اللّغات حسب مدى إلصاقها. أمّا العربيّة فهي لغة تصريفيّة (10)، وهي تختلف عن اللّغات الإلصاقيّة بكون مورفيمات التّصريف (الزّوائد) يمكنها أن تدلّ على أكثر من وظيفة نحويّة في كلمة واحدة خلافا للّغات اللاّصقة حيث يوجد تطابق كامل بين الشّكل والوظيفة.

ب - 1 - 1. تنقسم اللّغات التّصريفيّة بدورها إلى لغات جذريّة ولغات جذعيّة (thématique/radicale) ولغات مختلطة (جذريّة وجذعيّة). وتكون معالجة اللّغات الجذريّة أسهل من النّاحية الصرفيّة لأنّ مورفيمات التّصريف تضاف إلى الجذر فيسهل بذلك عزله، وأمّا في اللّغات الجذعيّة التي تحتوي على تصريفات داخليّة (تناوب الصّوائت vocal gradation / ablaut أو الصّوامت Alternance / alternation / umlaut) فإنّ التّغييرات تصيب الجذر فتصعب بذلك المعالجة لأنّ تنوّع أشكال التّغيير تفرض تعدّد الصّور والقواعد اللاّزمة لتوليد هذه الأشكال المختلفة.

إنّ العربيّة لغة مختلطة، يبدو مظهرها الجذريّ خاصّة في جدول تصريف الأفعال الصّحيحة وكذلك في تكوين المؤنّث (والمثنّى) والجمع السّالم (المؤنّث والمذكّر) فمثلا

---

(9) اللّغة التّأليفيّة هي اللّغة التي تحدد العلاقات النحويّة بالتصريف أو بدمج زوائد وإلصاقها بالجذور.

(10) اللّغة التصريفيّة هي لغة تحدد العلاقات النحويّة بالتصريف وبالحاق الزوائد بالجذوروبالكلمات التي تحتوي على أكثر من مورفيم.

الفعل 'علّم' ← 'علّمت' ← 'علّما' والاسم 'معلّم' ← 'معلّمة' و'معلّمان' (مثنّى مذكّر) ومعلّمتان (مثنّى مؤنّث) و'معلّمون' (جمع مذكّر سالم) و'معلّمات' (جمع مؤنّث سالم)، وصيغة الملكيّة، مثلا 'مُعلّم' ← 'مُعلّمي'.

ويبدو مظهرها الجذعيّ في تكوين المؤنّث انطلاقا من المذكّر وكذلك جموع التكسير (مؤنّثة ومذكّرة) واشتقاق الاسماء والصّفات انطلاقا من الصّيغ الفعليّة... فمثلا انطلاقا من فعل 'ضرب' نحصل على 'ضَرْبٌ' و 'ضارب' و 'مضروب' وانطلاقا من 'قاض' ← 'قضاة' و'فارس' ← 'فرسان' و'رجل' ← 'رجال' و'فؤاد' ← 'أفئدة' و'قرد' ← 'قردة' وانطلاقا من 'أسْود' ← 'سود' (المذكّر والمؤنّث) وسوداوات (المؤنّث فقط)... وتعدّ الأفعال المهموزة واللفيفة الأكثر إشكالا وصعوبة في المعالجة.

ب- 1- 1- 1. وإذ إنّ اللّغة العربيّة لغة تصريفيّة مختلطة فهي بطبيعة الأمور لغة جذريّة، إلاّ أنّنا إذا اعتبرنا هذه الخاصيّة فإنّه يجب علينا أن نميّز في مستوى اللّغات الجذريّة بين ما هو لاحقيّ (suffixante/suffixing) وبين ما هو سابقيّ (préfixante/prefixing) ومن هذا المنطلق فإنّ اللّغة العربيّة لاحقيّة\سابقيّة في الآن نفسه. فهي لغة لاحقيّة مثلا في الدّلالة على علاقة الملكيّة النّحويّة (مثلا 'كلب'← 'كلبه') وهي سابقيّة في الدّلالة على علاقات نحويّة مثل المكان والتّتابع والمرافقة (مثلا 'كلب' ← 'بكلب').

ب-1-2. تكمن الفائدة من معرفة انتماء لغة معيّنة إلى تصنيف معيّن خاصّة بالنّسبة إلى بناء مدقّق إملائي في أنّ هذا المدقّق في اللّغة التّأليفيّة- خلافاللمدقّق الذي يبنى في اللّغات التّحليليّة- لا يمكن أن يكون إملائيّا رسميّا صرفا بل يكون بالضّرورة نحويّا. ففي اللّغة الفرنسيّة مثلا (وهي لغة تحليليّة) يجب اللّجوء الى مدقّق نحويّ لإثبات أنّ السّلسلة "Le mon chien" أو "le chien mon" ليست مقبولة ولا يمكن مع ذلك لأيّ مدقّق إملائيّ أن يرى فيها خللا ولكنّ الأمور غير ذلك بالنّسبة إلى اللغات التّأليفيّة مثل العربيّة حيث نجد أنّ الشكل المعادل ليس مقبولا في مدقّق إملائيّ (مثلا 'الكَلْبُهُ').

ب-2. دراسة بنية الكلمة :

إذا استثنينا العلامات المميّزة كحركة التّضعيف (الشدّة) والهمزة فإنّ النّظام التّصريفيّ الرّسميّ العربيّ يعتمد بصفة تكاد تكون مطلقة على الصوامت (28 صامتا). ولا تكون

الصّوائت (7 صوائت) موسومة إلاّ إذا كانت تحمل نبر الدّيمومة (= طويلة). ويمكن لكلّ صائت أن يضعّف ويحمل الشّدّة باستثناء الحروف في أوّل الكلمة. وأمّا رسم الهمزة ونوعها وطبيعتها فأمور مقعّدة.

إنّ العربيّة لغة دمجيّة (Holophrastique/holophrastic) : ولمّا كان مفهوم الكلمة البسيطة المجرّدة لا يوجد في العربيّة إلاّ نظريًا فإنّ مفهوم الكلمة المركّبة (كلمة تحتوي على بياض\ فراغات(11) بين أجزائها) مثل الفرنسيّة "pomme de terre" ، أو على فاصلة عليا مثلا "l'ami" ، أو على خطّ وصل مثلا "vis-à-vis" لا يعني شيئا. فالبياض الفاصل، حدّ فيصل بدون منازع في جلّ اللّغات الرّومنيّة، ليست له نفس الوظيفة في اللّغة العربيّة فيمكن لبياضين أن يكونا حدّي جملة كاملة إن هي كتبت بلغة تحليليّة مثل الفرنسيّة (مثلا : "فَضَرَبْتُهُ" يناسبها "alors, je l'ai frappé"). فالتّجزئة في العربيّة لا تهمّ الكلمة البسيطة والكلمة المركّبة بل الكلمات بالعربيّة مركّبة بالقوّة أو قابلة نظريًا للتّركيب ويحلّ هنا مفهوم الكلمة-الجملة مكان المفاهيم الأساسيّة في اللّغات الرّومنيّة. (12).

ب- 2-1 . الأسماء : تسلك أسماء (غير الأعلام) في العربيّة المسلك نفسه الذي نجده في اللّغات التّصريفيّة الأخرى وهي تصرّف بغضّ النّظر عن التّناسق والانتظام بنفس الطّريقة التي تصرّف بها الأفعال(13)، وقد قسّمناها إلى أنواع وأصناف (لها

(11) في هذا المدقّق كما في المدقّقات الإملائيّة لجلّ اللّغات الرّومنيّة (langues romanes/romance languages) يعتبر البرنامج "كلمة" كلّ سلسلة من الحروف تحدّدها بياضات، علامات ترقيم (,;:?.!)، علامات الجداول، الخطوط الوصليّة (-)، بعض العلامات الخاصّة(@ & + / $ = #) وكذلك الأشكال المختلفة للاقواس [] {} () والهلالان المزدوجان «».

(12) ويذهب بعض مؤرّخي النّحوالعربيّ ودارسيه من القدماء والمحدثين في محاولة منهم لإنقاذ مفهوم "الكلمة" مذهب المبرّد، إذ يستشهد به عبد القادر المهيري(حوليّات الجامعة التّونسيّة، 1984، ص33-34)، فيقول «[. . .]الكلمة لا تحدّد بحجمها أو بعدد حروفها ولا بقابليّتها لانفصالها عن غيرها؛ فمن الكلمات ما يتكوّن من حرف واحد ومن حرفين اثنين كواو العطف وفائه وكاف الجرّ ولام الإضافة وما ومن ومَن وإن. . . والكلمات التي لا تتجاوز الحرف الواحد يستحيل أن تنفصل بنفسها [. . .]»،للنّقاش، انظر بقيّة التحليل.

(13) بالنّسبة إلى غير النّاطقين بالعربيّة (وحتّى بالنّسبة إلى المختصّين في الدّراسات اللّسانيّة للّغات من غير هذا الصّنف) تبدو هذه التّصرّفات غريبة وتكون نوعا من "التّغرّب اللّساني" ، فالفكرة القائلة بأنّ الأسماء تصرّف (يمكن لأسماء غير العلم أن تتشكّل بمعدّل 100 شكل مختلف نظرا للتّتابع الإلصاقيّ والإلحاقيّ للزّوائد ذات الوظائف النحويّة المختلفة) حسب قواعد صرفيّة لا يمكن قبولها بسهولة.

خاصيّات مشتركة) فأقمنا بذلك قواعد صوريّة للتّوليد تطبّق حسب الصّنف الّذي ينتمي اليه الاسم. ولكنّنا اعتمدنا لتوليد أشكال الجمع منوال تكوين الجموع المعروف في النّحو. وقد أنتج ذلك 119 قاعدة تصريفيّة للأسماء (ممّا يعطينا انطلاقا من جذر معيّن لاسم خال من الشّوائب نفس عدد الأشكال المولّدة).

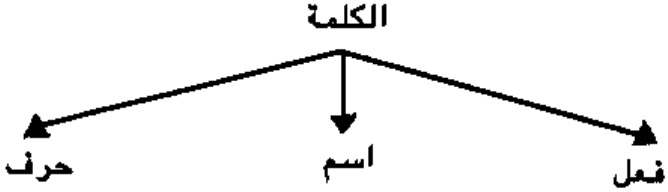
ب- 2-2 . الأدوات : لا نعتبر هذا الصنف إلاّ في حالة استقلاله معجميّا (نعني بذلك أن يكون محدّدا ببياضين # . . . # فاعتبرنا الأدوات الأخرى سوابق أو لواحق حسب الحالة (فمثلا "لِـ" و"كَـ" اعتُمِدا لتوليد أشكال نهائيّة انطلاقا من جذور اسميّة أو فعليّة).

وتخضع الأدوات بدورها للتّصريف ويجب اعتبارها، بغضّ النّظر عن تخمينات النّحويّين العقيمة، أسماء ناقصة أكثر من أيّ شيء آخر. فخاصيّة الأدوات هي أنّها، من بين كلّ العناصر المكوّنة للّغة، تتميّز إحصائيّا بقلّة عدد تصريفها، بل إن بعضها لا يصرّف أبدا (لا يمكن إلصاق شيء به ولا إلصاقه بأيّ شيء، مثل "ثمّ" و "أو").

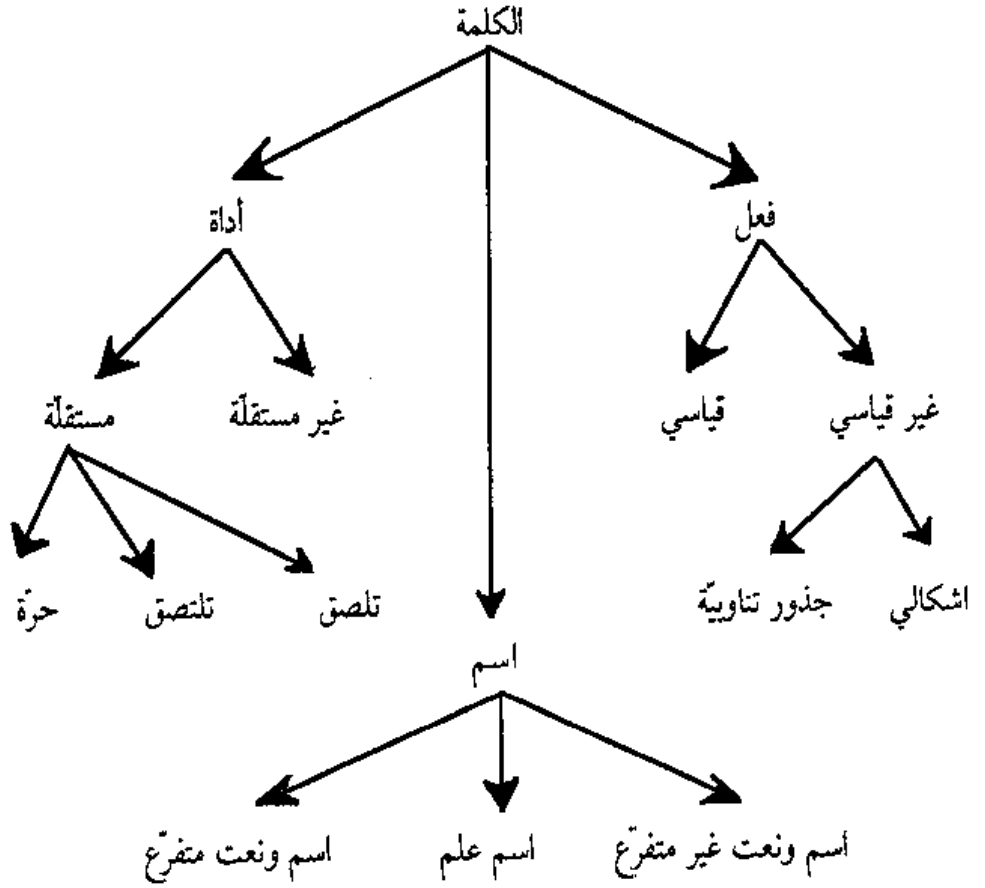
ب-2- 3. الأفعال : إنّ أغلبيّة الأفعال في العربيّة ثلاثيّة، بمعنى أنّ جذورها مكوّنة من ثلاثة صوامت وبعضها -وهي أقلّ عددا- رباعيّة ولكن يمكن ردّها بمراوغة نحويّة إلى جذر ثلاثيّ. وبما أنّ هذه الاعتبارات ليست بذات فائدة فإنّنا وقفنا عند الصّيغ المباشرة للأفعال دون اللّجوء إلى هذه المراوغات. ولكنّ بعض الأفعال الأخرى تذهب إلى أبعد من هذا العدد ويعتبرها بعض النّحاة أشكالا متفرّعة من جذور ثلاثيّة أو رباعيّة وبما أنّ عددها قليل جدّا خيّرنا معالجتها كأشكال غير قياسيّة.

ج . معطيات عمليّة تتعلّق باللّغة العربيّة :

إنّ التّبويب المعتمد وهو نفس التّبويب الّذي يقبله جلّ النّحاة العرب لا يصلح إلاّ جزئيّا، فهو معطى عمليّ أكثر ممّا هو ضرورة لسانيّة حقيقيّة.



يمكننا القول، إذن، إنّ هذا التّبويب ربّما كان ملائما نحويّا (تركيبيّا ودلاليّا) ولكنّه غير ملائم من وجهة النّظر التّصريفيّة، فبعض الأفعال المسمّاة -بحقّ- أفعالا "ناقصة" تتصرّف من هذا المنطلق وكأنّها أدوات، وبعض الأدوات وكأنّها نعوت، وبعضها تبدو كأنّها أسماء عاديّة، وبعض النّعوت كأنّها أدوات ولكنّنا إذا اعتمدنا هذا التّبويب منطلقا لقواعد التّوليد اضطررنا إلى أن ندخل عليها بعض التّعقيدات اللاّزمة :



ج-1 . الأفعال :

من جملة 12 000 فعل تقريبا تحتوي عليها نظريّا اللّغة العربيّة عالجنا 7 000 فعل (نقصد بها الأفعال المستعملة في مختلف الكتب المدرسيّة الّتي وقع جردها(14)) مبوّبة حسب 138 صنفا و98 نوعا تحتوي على جملة توفيقات الأصناف وتركيباتها.

(14) لقد أبعدنا من قائمتنا الأشكال القليلة الاستعمال أو الّتي لها زوج فونولوجي أقرب إلى القبول وذلك اقتصادا في الحجم.

إنّ التمييز بين الأصناف والأنواع مفيد جداً، فقد يكون لفعلين ينتسبان إلى نوعين مختلفين نفس القواعد ولكنّهما يكونان مختلفين من وجهة نظر التكافؤ (valence) مثل التعدية، كما يمكّننا هذا التقسيم من تخفيض عدد قواعد التوليد، فترسم بعض المعلومات الإضافيّة أمام كلّ فعل لتدلّ على الصّنف ونوع القاعدة التي يجب تطبيقها وشروط تكافؤ الفعل. وقد تحصّلنا بذلك في مرحلة أولى بعد إبعاد الأشكال غير المقبولة على ما يقارب 27,000 جذر قياسيّ (وقد لجأنا إلى تغيير بعض الجذور عمدا حتّى يسهل التوليد الآليّ للأشكال وما كان ذلك ليكون لو اعتمدنا الجذور الأصليّة)، وبعد تطبيق بقيّة قواعد التوليد تحصّلنا بعد التثبّت من صلاحيّتها اللّغويّة على ما يقارب الثّلاثة ملايين شكل.

وكان العدد الجمليّ للأفعال الإشكاليّة (التي تحتوي على تغييرات نتيجة التطوّر الزّمني للّغة لأسباب مختلفة مثل القياس والميول الفونولوجيّة)، أي التي وقع إدخالها يدويّا، ضئيلا جداً يقارب 5334 شكلا نهائيّا أي أقلّ من نسبة 0,28 % من مجموع الأفعال المولّدة. ويربط الفعل القياسيّ بـ63 شكلا تصريفيّا بغضّ النّظر عن الحركات (voyelles) والتّعدية. فبعض الأشكال التّصريفيّة لا تختلف إلاّ بوضع الحركات وتكون متشابهة إذا تجاهلناها.

وهذه الأشكال التّصريفيّة الثّلاثة والستّون نتحصّل عليها بتصريف الأفعال في الأزمنة الأربعة (الماضي والمضارع المرفوع والمجزوم والمنصوب) مع الضّمائر الثلاثة عشر (أنا، أنت، أنت، أنتما، أنتم، . . .) وفي الأمر مع ضمائر الخطاب الخمسة.

ج-2. الأسماء : هي الأسماء (غير الأعلام) والأسماء-الصّفات وأسماء المرّة والآلة والمكان والزّمان الخ. . . وهي مقسّمة إلى ثلاثة أصناف (أسماء مشتقّة وأسماء غير مشتقّة وأسماء أعلام). وتحتوي قاعدتنا على 254 31 جذرا للأسماء. وقد كوّنا 119 قاعدة صوريّة لاشتقاق الأسماء أفضت إلى نفس العدد تقريبا من القواعد المكتوبة في نحو لغة برمجة المولّد (langage de programmation du générateur). وقد جمعنا الأسماء حسب القواعد التي تطبّق عليها وحسب ترتيب تطبيقها وقد تمكّنا من تمييز خمسة وثلاثين نوعا من الأسماء فوضعنا أمام كلّ جذع اسميّ علامة تشير إلى القاعدة الواجب تطبيقها على ذلك الجذع حسب عددها (من 1 إلى 35).

ج-3. الأدوات : يوجد في العربيّة ما يقارب 129 حرفا-أداة معروفة، بعضها يلتصق بالجذور (كـ + كلب ← ككلب) وبعضها مستقلّ تماما بمعنى أن لا شيء يلتصق به ولا يلتصق بأيّ شيء (ثمّ)، والبعض الآخر يلصق به بعض الزوائد النحويّة (فـ+ فيـ+ ـه ← ففيه).

ج-4. كان العمل الّذي قدّمناه للمبرمجين لكتابته في لغة البرمجة يحتوي على ثلاث جذاذات : الاولى لقواعد التّوليد الصّوريّة (مع عددها الرّتبيّ) بدون تمييز ما ينبغي أن تطبّق عليه، والثّانية لأصناف الأسماء والصّفات والأدوات تحتوي على العدد وترتيب قواعد التّوليد الّتي يجب تطبيقها، والثّالثة لمجموعات الأفعال وأنواعها بقائمة الجذور إذ تتميّز الأفعال بانتمائها إلى صنف مشار إليه بوجود رقم رومانيّ كما يبيّنه المثال الآتي :

كلب (25)؛

أحمر (19)؛

ضرب (T،45،XII)؛

ثمّ (24)؛

قلم (1).

وهذا يعني أنّ المدخل "كلب" من صنف (25) الّذي يشير إلى وجوب تطبيق القواعد الّتي تنتمي إلى هذا الصّنف كما بيّنا ذلك في جذاذة الأصناف (fiche des types) . وفيما يخصّ فعل "ضرب" (T، 45،XII) فذلك يعني أنّ الفعل ينتمي الى الصنف XII (من الأفعال القياسيّة) والنوع 45 (المناسب لطريقة تصريف الأمر) وهو متعدّ (وذلك يعني أنّه يجب تطبيق قاعدة تصريف اللّواحق الواردة مفعولا به وإضافتها مثل "ضرب"، "ضربه"، "ضربته" الخ).

د. معطيات عمليّة تتعلّق بالمدقّق الإملائيّ : (استعمال المدقّق وخاصيّاته الوظيفيّة)

قد سبق قاعدة المعلومات المعجميّة تكوين نظام قواعد صوريّ في شكل نظام تصريف آليّ للأفعال (7000 فعل) الواردة وللأسماء بنظام السّوابق واللّواحق ومختلف الزوائد التصريفيّة والزّوائد النّحويّة وصيغ إلصاق الحروف للأسماء كما ذكرنا آنفا . . .

د- 1. يمثّل المدقّق في حدّ ذاته أداة استشاريّة وليس أداة تدقيق فحسب ويمكن أن يستعمل معجما إملائيّا (dictionnaire orthographique) على ركيزة إعلاميّة ويمكن استعماله قاعدة نستطيع مساءلتها فيما يخصّ رسم الكلمات المتجانسة (homonymiques) مثل الكلمات المشتملة على "ظ" و"ض" (ظلّ، ضلّ) أو المشتركة الجذور (polysémiques) (عين) أو مشكلة رسم الهمزة وبذلك يظهر لنا بعد آخر لهذا المدقّق، نعني به قيمته البيداغوجيّة كأداة تعليميّة.

إنّ هذا المدقّق غير مشكول ويمكن للنّص المعالج أن يحتوي على علامات شكل فإنّ المدقّق يتجاهلها (دون أن يضيعها) عند عمليّة الإصلاح.

د- 2. هو أوّل مدقّق (حسب علمنا) يعتبر الشدّة أو لا يعتبرها حسب رغبة المستعمل من قائمة "اختيارات" في المدقّق العربيّ. ويقبل هذا المدقق إضافة الى ذلك أشكال الـ"ي" المختلفة ("ي" "ى" أي ياء وألفا مقصورة) أو يرفض (حسب ما يريده المستعمل). وهذا الاختيار المهمّ مرتبط في الحقيقة بمشاكل التّوليد، وبالفعل فإنّ النّظام التّصريفيّ العربيّ لا يوجد فيه أيّ فرق بين الياء والألف المقصورة، بيد أنّ برامج تنسيق النّصوص تعتبرهما سجلّين مختلفين فهما بذلك حرفان مختلفان ومتباينان تماما. ويتبع ذلك مشاكل لا حلّ لها بالنسبة إلى المستعملين وكثير من الضجيج وكثير من الصمت(15) في مستوى قاعدة المعلومات.

د-3. ويضبط المدقّق مختلف مواقع الهمزة على الألف في أوّل الكلمة، وإذ إنّ المستعملين لا يحمّلون أنفسهم عبء كتابة الهمزة على الألف فإنّ البرنامج يقترح (اختيارا) أن يعيرها اهتماما أو أن يتجاهلها.

ولا يحتوي البرنامج-المدقّق المكوّن انطلاقا من القاعدة على محلّل تصريفيّ. فقد سبق الوصفَ اللّسانيّ والنّحوَ عملُ توليد القاعدة، والمحرّك الوحيد الموجود في البرنامج هو عبارة عن جدليّة مكّوكيّة بين القاعدة المعجميّة والأشكال المكتوبة في النّصّ التي تعترض البرنامج المدقّق.

(15) الضّجيج والصّمت مفهومان يخصّان قواعد المعلومات فكلّما أظهرت القاعدة كلمة غير مقبولة ولم يُقصد إدخالها في القاعدة كان ذلك ضجيجا وكلّما تجاهلت القاعدة كلمة وقع إدخالها كان ذلك صمتا.

د- 4. ويبقى حجم القاعدة بعد توليد مختلف الأشكال وبعد إبعاد الأشكال المشبوهة أو الأشكال النّظريّة الصِّرف كبيرا جدّا، وهو ما يجعل استعمالها مستحيلا من قبل البرنامج ولذلك فإنّنا عمدنا الى "ضغطه" بالتّعاون مع كاتبي البرنامج المولّد.

خاتمـــة :

إنّ ضرورات البرنامج المدقّق النّظريّة والعمليّة التي تنتهي بنا في بعض الأحيان إلى إعادة النّظر في بعض البديهيّات أثبتت لنا أنّ الإعلاميّة من أحسن الطّرق لتحسين أمثلة الوصف اللّساني وأنّ التّشكيل(formalisation) ليس محسّنا بديعيّا بقدر ما هو شرط عمليّ. ويبدو لنا أيضا أنّنا نقف عند ميدان تعامل وتكامل ممتاز بين اللّسانيّات والمنطق والرّياضيّات والإعلاميّة.

عبد الرّزّاق بنّور
كلية الآداب بمنّوبة
جامعة تونس الأولى

المراجع :

عبد القادر المهيري : "مفهوم الكلمة في النّحو العربيّ" حوليّات الجامعة التّونسيّة عدد 23، 1984، ص ص31 - 43.

نبيل علي : اللّغة العربيّة والحاسوب، نشر دار تعريب، بيروت، 1988.

Delas et Delac, "in Courtois B. et Silberstein M. "Les dictionnaires électroniques" **Linguistica Communicatio** n° 1, 1989.

Degachi A., **A Morphosyntactic Processor of Modern Standard Arabic**, Ph D. University of Bath. 1991 (thèse non publiée).

Dugas A. "Le verbe dans un dictionnaire éléctronique du français", in **Linguistica Communicatio** n° 1, 1990.

Yoka G.T. "Le vérificateur d'orthographe : des analyses phonétiques et lexicales à l'analyse syntaxique", in : **Le Français Moderne**, Tome LX, n° 2, décembre 1992.