

1- ملخص:

تمتاز اللغة العربية شأنها شأن اللغات السامية بقوة التوليد الصرفي، وفي هذا البحث، نقدم تصورا نعتقد أنه جديدا في نظرتنا إلى الدلالة الصرفية المبنية أساسا على التوليد الخوارزمي الصرفي، ويعتمد هذا التصور على مستويين من الدلالة الصرفية، يمكن التمييز بينهما عند تحليل دلالة التوليد، أولهما الدلالة المصاحبة للجذر، والثاني الدلالة المصاحبة للخوارزميات الصرفية. إن تفاعل هذه المستويات يولد مفاهيم متكاملة، يمكن استيعابها على هيئة مجموعات توليدية أو مشتقات عربية. ولعل الغرض الأساس من هذا المنظور إلى الدلالة الصرفية، يتمثل في التعرف على سبل الربط بين المفاهيم الدلالية والمفردات العربية، فيما يصطلح على تسميته في علم توليد اللغة بالتوليد الخوارزمي الصرفي للمفردات بنوعها البسيط والمركب. يتجلى الهدف من هذه المشاركة في تقديم خوارزميات صرفية للغة العربية، تمكن من توليد أقسام الكلم عامة، والأسماء البسيطة على وجه الخصوص، ومن ثمة الوصول إلى بناء قاعدة بيانات صرفية عامة موجهة أساسا لخدمة المعاجم الآلية للغة العربية.

Abstract:

Arabic, like the other Semitic languages; is characterised by morphological generation .This paper suggests a new outlook on morphological semantics which relies essentially on algorithmic and morphological generation.

This perspective relies upon two levels of morphological semantics that can be distinguished separately when considering semantic generation and analysis: first the morpheme semantic of the root, second the semantic of algorithmic generation .The interaction of these levels yields complete meaning which can be assimilated as generative groups or derivative sets.

The purpose of this approach, then, is shed light on morphological semantics, and try to find out the link between Arabic words (both simple and complex) and their semantics .This is called morphological algorithmic generation of simple as well as complex words, helpful in building paradigms which can be very useful in electronic translation of the Arabic language.

و المركب.

2- الخوارزميات من الدلالة اللغوية إلى الدلالة الاصطلاحية:

قبل أن نتكلم عن الخوارزميات الصرفية في اللغة العربية، لا بأس من الوقوف عند مصطلح الخوارزمية نفسه (algorithme)، والذي لم يعط له ما يستحق من العناية والاهتمام في الأدبيات اللسانية الحاسوبية العربية. إذن، ما هي أصوله الإيمولوجية، وما دلالاته، و ما معناه في اللسانيات و المعلومات؟

إن الخوارزمية أو الخوارزميات ليست مأخوذة من أية كلمة لاتينية أو يونانية، وإنما هي ذات أصول عربية صرفية، وهي تقترن إيمولوجيا باسم العالم الرياضي العربي محمد بن موسى الخوارزمي الذي ألف كتابين مهمين: الأول سماه «علم الحساب»، والثاني عنوانه ب«الجبر والمقابلة»، وبمرور ثلاثة قرون تقريبا، ترجم مصطلح الخوارزمية إلى اللاتينية تحت عنوان (algorismus). وقد نقل المصطلح من مجاله الرياضي الصرف ليستخدم في ميدان المعلوماتية على مستوى بناء البرامج الحاسوبية.

الخوارزميات المعلوماتية هي وسيلة إجرائية، تتجلى وظيفتها في القيام بالحساب في الحاسوب، وأول تعريف للخوارزمية في العصر الحديث هو التعريف الذي قدمه العالم الروسي ماركوف: الخوارزميات هي القواعد الدقيقة، والتي توظف باعتبارها وسيلة إجرائية للحساب، وتمكن في نهاية المطاف من الوصول إلى نتيجة دقيقة انطلاقا من معطيات أولية.

و على هذا الأساس، إن صياغة خوارزمية معينة هي من قبيل الفعل الإبداعي والإبتكاري، أي صياغة الأدوات المعرفية والرياضية والمنطقية التي تقبل البرمجة الحاسوبية، ومن ثمة تترجم إلى خطاطة وظيفية. إن الخوارزميات المعلوماتية تقوم بعدة وظائف من قبيل جمع المعلومات، وفرزها في الذاكرة المركزية للحاسوب، والبحث السريع عن المعلومات، ومعالجة المعطيات، وترجمة لغات البرمجة و الحساب الصوري.

أما الخوارزميات في اللسانيات الحديثة، فقد عرفت تعريفات متباينة ومختلفة نختار منها ما يلي:

تحدد الخوارزمية في النحو الصوري على الشكل التالي: 1

تحدد اللغة الصورية: لص والمعجم: عج، لص هي مجموعة الكلمات أو الجمل التي تنتمي بشكل حر إلى الزوج: 2* (أو عج*) المولد بواسطة أ (أو بواسطة عج) حيث تمثل لمجموعة الكلمات (أو الجمل) المقبولة.

¹ - Maurice Gross , André Lentin , *Notions sur les grammaires formelles* - collection « programmation » ,Gauthier Villers,Editeurs ,Paris 6/1970, p44.

² - M.Gross, opcit -

إذن ، تمثل ل على النحو الآتي : خو* / ل .
حيث ترمز الخطاطة إلى مجموعة الجمل الملغاة.
وتعني الخوارزميات عند البعض تعاقبا محددًا من التعليمات الموصوفة جيدا، مع توفر الخصائص التالية3:

- لا يوجد غموض في أية تعليمية .
 - بعد تنفيذ تعليمية معينة، لا يوجد غموض في تحديد التعليمية التالية.
 - يجري بلوغ تعليمية انتهت ذاتها بعد تنفيذ عدد من التعليمات.
- فيما عرفها آخرون على النحو التالي :4
- الخوارزميات هي مجموعة من القواعد المنظمة في طريقة معينة، تنطلق من القاعدة البسيطة إلى القاعدة المركبة، ثم إلى القاعدة الأكثر تعقيدا.
- هي منظومة من القواعد الصورية المبرمجة في الدماغ البشري ، وتقوم بوظيفة المراقبة بالنسبة للغة على مستويي الإدخال والإخراج.
- هي التقنيات التي يتم بموجها إجراء عمل الحساب على الحاسوب أي سلسلة من العمليات الجزئية المتناسقة ، تتوقف كل واحدة منها على السابقة ، يجري العمل للوصول إلى غاية أو تحقيق نتيجة معينة كعمليات الحساب والجبر على الحواسيب وغيرها .
- نستنتج من خلال التعريفات السابقة الذكر، أن الخوارزميات اللغوية هي مجموعة من القوانين والمبادئ العقلانية المتمركزة في منطقة معينة في دماغ الإنسان، والتي بموجها يستطيع المتكلم اللغوي إنتاج اللغة وفهمها وتوليدها وتحليلها. وفي اعتقادنا أن هذه الآليات الصورية قد لعبت دورا كبيرا وأساسيا في الميدانين الرياضي والمعلوماتي، و من المؤكد أن يكون لها دور وطلائعي ورائد في مجال المعالجة الآلية للغات الطبيعية وهندستها.
- وانطلاقا من النظام الرياضي والمنطقي الذي يحكم برنامج عمل الخوارزميات في الدماغ، كيف يتأتى للقدرة اللغوية استنباط المفردات بنوعها البسيط والمركب من الجذور؟ ثم ما هي المورفيمات التي تضاف إلى البنية النظرية لتكوين المفردة ؟ وهل هناك من إمكانية لحوسبة الصرف و المعجم العربيين . ؟

3- أتمتة الصرف والمعجم :

- ينظر عبد الاله الديوه جي، مفاهيم أساسية حول تقنية المعلومات مجلة عالم الفكر، م18، ع.3، 1987/3.

4- ينظر أطروحتنا، توليد الأسماء من الجذور الثلاثية المعتلة مقارنة لسانية حاسوبية، 1999، ص60.

يعالج هذا المبحث إمكانية حوسبة الصرف والمعجم العربيين، وسوف نركز على مقاربات حاسوبية كان لها إسهام بالغ الأهمية في حوسبة صرفيات اللغات الطبيعية عموماً، واللغات السامية -ومن بينها العربية - على وجه الخصوص.

1.3- الأوتومات محدودة الحالات Finite State Automata

الأوتومات هي عبارة عن أوتومات أو آلة ، تتألف من عدد من الحالات المحدودة . الآلة تنتقل من حالة إلى حالة أخرى كلما قرأ رمز أو كتب رمز. ويمكن تمثيل الأوتومات محدودة الحالات .

تتكون هذه الشبكة من أربع حالات ، الحالات مترابطة عن طريق أسهم ، وكل سهم يحمل رمزا هو عبارة عن عنوان له، هذا الرمز يحدد الحالة التالية التي يجب أن تنتقل إليها الشبكة استناداً إلى رمز، تم إدخاله ، أو تمت كتابته . وقد تولد عن فكرة الأوتومات محدودة الحالات فكرة الأوتومات الثنائية محدودة الحالات Finite State Sranducers . والأوتومات الثنائية محدودة الحالات هي في الأساس أوتومات محدودة الحالات ، تتجلى وظيفتها في ربط المدخلات inputs بالمخرجات outputs ، حيث تأخذ رمزا مدخلا ، وتعطي رمزا مخرجا. إن الأوتومات الثنائية محدودة الحالات تستطيع التعرف على حزمة مدخلة في الوقت الذي تولد فيه حزمة أخرى ، ومن الأسهل للقارئ أن يتخيل الحزمتين كشريطين : شريط يقرأ منه الأوتومات ، وشريط آخر يكتب عليه الأوتومات . ويمكن تمثيل الأوتومات الثنائية محدودة الحالات على نحو ما فعلنا مع الأوتومات محدودة الحالات، عن طريق شبكة محدودة الحالات Finite State Transition Network ، مع فرق طفيف من حيث الأسهم ، ذلك أن الأسهم التي تربط حالات الشبكة الثنائية محدودة الحالات معنونة بزواج من الرموز عوض رموز منفردة . ويبدل العنوان $x : y$ الذي يحمله سهم معين على أن هذا السهم يتم اجتيازه إذا كان الرمز الحالي الذي تقرأ منه الشبكة محدودة الحالات من الحزمة المدخلة هو x ، وأن السهم المخرج الذي تكتبه الشبكة حالياً على الحزمة المخرجة هو الرمز y . تستخدم الأوتومات محدودة الحالات في ضوء المعالجة الطبيعية للغات أساساً للمعالجة الحاسوبية للصرف (computational morphology processing) ، كما تستخدم لأغراض عديدة ، فهي من جهة تعد أداة مرنة ومحايدة من حيث التوليد والتحليل ، بمعنى آخر ، على مستوى التحليل ، يمكن أن تستخدم في استقبال المدخلات ، وعلى مستوى التوليد ، تستخدم في إخراج المخرجات ، ومن جهة أخرى تتميز الأوتومات محدودة الحالات بالبساطة والسهولة من حيث التطبيق الحاسوبي ، وتوفر السرعة والفعالية .

2.3- الصرف ثنائي المستوى Two-level morphology

يهدف الصرف ثنائي المستوى إلى وصف البنية الصرفية للكلمات، وقد ظهر هذا التطبيق إلى حيز الوجود لأول مرة على يد كيمو كو سكينيمي الذي استخدمه لأغراض تتعلق أساسا بالتوليد والتحليل الصرفيين. وهو يقوم على فكرة مفادها أن نفكر في الصرف على اعتبار أنه عملية تحديد ترابط بين التمثيل السطحي والتمثيل العميق للكلمة، يتم ربط أحد المستويين بالآخر باستعمال الأتومات محدودة الحالات. الحزمة المعجمية تتكون من مجموعة من الوحدات الصرفية أي المورفيمات التي تظهر في المعجم، أما الحزمة السطحية فتقابل الصيغة أو البنية التي تظهر عند استخدامنا للغة فعليا.

يتطلب الصرف ثنائي المستوى ربطا بين الحزمة السطحية (المستوى السطحي) والحزمة المعجمية (المستوى المعجمي)، وذلك بالإستناد إلى قوانين خاصة تدعى بقوانين أو قواعد ثنائية المستوى (Two-level rules). هذه القوانين يتم تحويلها إلى أتومات محدودة الحالات ، تمثل علاقات بين المستويين السطحي والمعجمي. تتخذ القوانين ثنائية المستوى الشكل التالي : CP
 \longleftrightarrow LC — RC

حيث يشير الرمز CP إلى ترافق سطحي أي معجمي، ويمثل كل من RC و LC السياق الأيمن (RC) والسياق الأيسر (LC) على التوالي، بينما يقول السهم ثنائي الاتجاه أن سياق CP هو RC و LC. كل من CP و RC و LC تتكون من أزواج من الرموز مثل (x:y) ، الرمز الأول في هذا الزوج أي x يعرف بالرمز المعجمي، ويمثل رمزا يظهر في المعجم، أما الرمز الثاني أي y فهو الرمز السطحي ويمثل رمزا يظهر في الصيغة السطحية. لذلك يعني الزوج x:y هنا أن الرمز المعجمي x يمثل ك y في المستوى السطحي. ويمكن أن يكون الرمزان متماثلين، مما يفيد أن التمثيلين المعجمي والسطحي متماثلان أيضا . وفي هذه الحال، يمكن التعبير عنهما عن طريق كتابة واحد من الرمزین (x تقابل x). بالإضافة إلى ذلك هناك الرمز الصفري 0 الذي يمكن أن يستخدم في زوج الرموز للدلالة على أن الرمز المعجمي ليس له تمثيل في السطح أو أن الرمز السطحي ليس له تمثيل في المعجم. على سبيل المثال، e:0 يدل على أن الرمز المعجمي e ليس له تمثيل في السطح، وهذا يعني أن هذا الرمز يلغى في الحزمة السطحية. وفيما يلي قانون من القوانين ثنائية المستوى :

$$n : m \longleftrightarrow _ p$$

يفهم من هذا القانون أن الرمز المعجمي n يقابل الرمز السطحي m إذا كانا متبوعان بالرمز المعجمي p الذي مثل أيضا على شكل p سطحيًا. هذا القانون يقابل قانون المماثلة الصوتية المسؤول عن تحويل كلمة (ينبغي) إلى (بمبغي) و (impossibe) إلى (impossibe).

وكما سبقت الإشارة إلى ذلك القوانين ثنائية المستوى يتم تحويلها إلى أتومات محدودة الحالات الحزم المعجمية و السطحية يمكن أن يتم اعتبارها شريطي الأتومات. الأزواج التي تقابل الرموز المعجمية والسطحية يمكن أن تعتبر الأزواج التي تمثل عناوين للأسهم التي تربط حالات الأتومات الثنائية محدودة الحالات .

نستنتج من هذا التحليل أن الصيغة السطحية blamed تحلل باعتبارها جذرا معجميا وهو blame متبوع باللاحقة المعجمية ed. وقد تم الربط بين الحزمة المعجمية blame + ed من جهة والحزمة السطحية blamed من جهة أخرى عن طريق القوانين ثنائية المستوى. في هذا السياق ،يمكن أن يكون هنالك قانون يسمح الرمز معجمي بأن يكتب كما هو في السطح دون تحديد أي سياق ،ويتخذ هذا القانون الشكل التالي :

ordChar :OrdChar<—> _

انطلاقا من هذا القانون يقابل الرمز ordChar (ordinary character) أي رمز عادي .هذا القانون يربط بين الرمز الموجودين في زوج الرموز إذا كان هذا الربط غير ممنوع من قبل قانون معين في ساق معين .هذا القانون يمكن أن يستخدم لربط b :b, d :d, m :m, a :a, l :l, c :c علاوة على ذلك هنالك قانون آخر يحلل رمز الفاصل الصوتي “+” الذي يفصل بين الوحدات الصرفية الموجودة في المعجم. رمز لفاصل الصوتي يقابله 0 أي أنه لا يمثل في الحزمة السطحية. ويتخذ هذا القانون الشكل التالي:

+:0<—> _

بالإضافة إلى ما تقدم من قوانين، يوجد قانون آخر هام يسمح بتمثيل الرمز e الموجود في نهاية بعض جذور اللغة الإنجليزية إذا ما أتبع هذه الجذور بأي لاحقة مثل ed و ing، ويتخذ هذا القانون الصورة التالية:

e:0<—> _ e :c+ :0

مفاد هذا القانون أن الرمز المعجمي e في اللغة الإنجليزية يمثل ك 0 في الحزمة السطحية إذا ما:

- 1- سبق بصامت c معجمي يقابله صامت سطحي. الرمز c يقابل أي صوت صامت في اللغة .
- 2- تبع بفاصل صرفي معجمي يتم تمثيله ك 0 في الحزمة السطحية .

3.3- الصرف الأحادي المستوى : One –level morphology

اقترح والتير (walther) (2000، 2000ب) حلا يقضي بوصف استنساخ اللغات الطبيعية ،وتوظيف تقنيات الحالات المحدودة . إن هذه الفكرة ترمي إلى إغناء الأتومات محدود الحالات باقتراح ثلاث عمليات جديدة : المكرر repeat، والقفزة skip، والحلقة ذاتها self loops. تسمح أقواس المكرر بنقل خيط في الداخل بشكل عكسي وتكرر جزءا منه .وتسمح أقواس القفزة بتثبيت

عملية نقل الخيط إلى الأمام خارج أي حرف من حروفها، مثل جزء فقط في بعض أغلفة التقنية المستعملة لتوصيف الاستنساخ ما لم تكن المفردة منسوخة. كان الغرض من عرض هذه المقاربات الصرفية الحاسوبية هو تبيان وضع اللغات الطبيعية الأجنبية في ميدان المقاربة الآلية، و توضيح إمكانية إستغلال تلك التجارب والإستفادة منها في حوسبة الصرف والمعجم العربيين.

4- خوارزميات صرفية للغة العربية : من التحليل إلى التوليد

يعتبر التوليد من الجذور والخوارزميات أو التوليد الخوارزمي الصرفي للمفردات اللغوية العربية من الأمور اللازمة لأنظمة الترجمة الآلية، ونظم توليد اللغة، ويتطلب تنفيذه توافر الوسائل اللغوية والحاسوبية للربط بين معنى معين وكلمة تستوعبه في اللغة الهدف. وعليه فإن التوليد الآلي للمفردات اللغوية العربية أقرب ما يكون إلى التوليد الخوارزمي الصرفي للمفردات، نظرا لوحدة الهدف، والمفهوم اللغوي و التقني.

ولهذا فإننا نحتاج إلى الخوارزميات التي نصنف بموجبها الجذور اللغوية، وبناء على ذلك صنفنا الجذور إلى جذور مطردة (عادية) وجذور نظرية (افتراضية)، مما يفيد أن القدرة المعجمية لدى المتكلم اللغوي تتضمن الجذور المحققة والجذور النظرية(التي تولد بطريقة إعادة الإنتاج)، مما يعني أن المعجم المخزن في القدرة ينشطر إلى شطرين : شطر توليدي ومنتج، يمكن أن نحدد خوارزمياته كما هو الأمر بالنسبة للأسماء والأفعال والمصادر والصفات... الخ، و شطر غير منتج وعقيم مثل الحروف والأدوات وأسماء الأعلام والأسماء المركبة والتعابير المسكوكة والمتلازمات المعجمية والمصطلحات العلمية والتقنية ...

وهكذا يتألف المعجم الآلي للجذور من نمطين من البيانات: قاعدة بيانات الجذور المحققة والمطرده وهي تمثل نسبة 60% من مجموع النظام اللغوي، بينما تمثل قاعدة بيانات الجذور النظرية نسبة 40%.

متى يبدأ المدخل المعجمي في الاشتغال، وما هي الخوارزميات الصرفية التي تفعل تشغيل الجذر وتحريكه ؟

إن الجذر هو نواة و مصدر التوليد، أي مادة تنظيمية لمدخل المعجم، والعلاقة بين الجذر الصوتي والمدخل المعجمي موجهة تتم وفق متوالية من الزوائد والعناصر، مما يعني أن العلاقة بينهما ذات مد وجزر من جهتي اليمين واليسار، ولهذا فالجذر يدل على حدث ممكن التحقق، فيما يدل الجذر على حدث محقق، وبمعنى آخر، إن الجذور ترتبط بالموجود بالقوة، وبينما المدخل المعجمية ترتبط بالموجود بالفعل، وهذا يجرنا بطبيعة الحال إلى الحديث عن المهمل والمستعمل بالنسبة للجذور، وكيفية تعامل الآلة معه. إن الجذور المستعملة هي تلك التي تخضع لقوانين الاستعمال

⁵ - مهديوي 1999، مرجع سابق

اللغوي العربي، والتي احتفلت بها كتب اللغويين والمعجميين والصرفيين والنحويين العرب، مما يتصل بالهوية العربية التي تنحصر في الجذر والوزن والتألف الصوتي... وما دون ذلك فهو إما دخيل أو معرب. أما الجذور المهملة فهي التي تخرق قاعدة من قواعد اللسان العربي، ومع ذلك فإن الحاسوب سيتعامل معها باعتبارها من فصيلة اللغة، لها قواعدها خاصة، لا باعتبارها تشذ عن التوصيف اللساني والحاسوبي.

نظريا نفترض أن الجذر هو العلامة (€)، وهو يتوزع إلى مجموعة من الصوامت الثابتة والمحدودة العدد نمثل لها:

$$€ = R1. R2. R3$$

لكن الجذر لا يتضمن الصوامت وحدها، وإنما يحتاج إلى مجموعة من العناصر لكي يكون متوالية لسانية، يوضح ذلك المعادلة التالية:

$$€ = €R. 1 1. € . R 22 . €R 3 3 . € 4$$

يبدأ الجذر في التشغيل والحركية، عندما نبدأ في ملء الفراغات والفواصل الصوتية، أي عندما يبدأ المحلل النحوي الطبيعي في الاشتغال. هذه الفراغات عبارة عن حروف تسمى ب (les octets)، ولكل حرف رقم، هو عبارة عن سنن، وبموجب هذه الفراغات، ينتقل الجذر من المستوى النظري المجرد إلى المستوى اللغوي المحقق، والأشياء التي تضاف إلى بنية المتوالية الصوتية المجردة غير الدالة (الجذر) هي التي تؤدي إلى التوليد، بمعنى آخر، عندما يشرع المولد الصرفي الطبيعي في العمل، حيث يدل الرمز (1 €) على السوابق التي تكون قبل الجذر، و (3 € + 2 €) على الأحشاء أو الأواسط التي تكون بين حروف الجذر داخليا، وتدل (4 €) على اللواحق التي تكون في نهاية الجذر.

ما هي العناصر التي تضاف إلى الجذر لتوليد الأسماء، والأفعال، والصفات...، وفي أي موقع توضع هذه العناصر؟

قبل الإجابة على هذا السؤال يجدر بنا القول إن الخوارزميات اللسانية في هذا المستوى لا تولد إلا الدلالة العقلية المفردة، أما الدلالة الوضعية المركبة، فهي مرتبطة بالتصديق، والفرق بينها فرق بين التصور و التصديق، وبين المفرد والمركب. بمعنى أننا عندما نكون بصدد المفردة في الذهن، فإننا نكون بصدد التصور النظري أي الدلالة العقلية المنطقية، وعندما ننتقل بهذا التصور إلى العالم الخارجي، نكون في مستوى التصديقات (الدلالات الوضعية المركبة). ولكي يتحقق هذا الانتقال لابد من تدخل المستوى الصرفي الخوارزمي.

ما علاقة اللفظ بالمعنى؟ ثم كيف يتوصل المستوى الصرفي الخوارزمي من الانتقال من المبنى الصوتي إلى مبنى معنوي (كلمة)؟

يقصد بالفهم والمعنى الدلالة الصرفية المصاحبة للخوارزميات اللسانية، وتجب الإشارة هنا إلى أنه لا يمكن أن نبحث في ضوء البحث اللساني الحاسوبي والهندسي ما لم نفهم بعمق المنطق التأليفي الذي يهتم بالتصور والتصديق. إن الوصف المراد به هنا ليس الوصف السطحي الشكلي المطبق في الأدبيات النقدية واللغوية العربية القديمة، ولكنه أعمق من ذلك، إنه بمعنى إنتاج دلالة ما لقاعدة لسانية، أي تلك العملية البسيكولوجية التي تسقط على الخارج، باعتبار الميدان الذي تولد فيه الدلالة، بحيث تطبق القواعد في القدرة اللغوية لتوليد الدلالة، ويمكن أن نوضح ذلك كله على النحو التالي 6 :

الجذر الذي يؤدي إلى توليد الفعل : € — V <

€ — R . 2R . 1 R < 3

الجذر الذي يؤدي إلى توليد المشتقات وهو المعبر عنه في المعادلة التالية:

€ — ND <

€ — R . 2R . 1 R < 3

ما هي العناصر المورفيمية التي تضاف إلى بنية الجذر لتوليد الفعل ؟
 هناك خوارزميات مورفيمية زائدة خاصة تدخل على بنية الجذر لتوليد الفعل وأصنافه، هذه الخوارزميات منها ما يشترك مع الأسماء، ومنها ما هو خاص بالفعالية، ومنها ما هو خاص بالاسمية. لكننا في هذه الدراسة سنتوقف عند الأسماء وحدها، دون أن ننفي التوزيع التكاملي بينها والفعل. إن الفعل في العربية أصناف، هناك الماضي، والمضارع، والأمر، ولتوليد كل صنف من هذه الأصناف لا بد من توظيف خوارزميات صرفية معينة.

وإذا كان الأمر كذلك بالنسبة للأفعال، فكيف تولد المشتقات ؟

تعتبر مادة الفعل قناة للوصول إلى الجذر لتوليد المشتقات (اسم فاعل، اسم مفعول، مصدر، أسماء زمان ومكان، أسماء مبالغة،....) لكن الأمر ليس عاما إلى حد الإفراط، بمعنى آخر، أن القواعد لا تطرد بالنسبة لمعالجة هذه الأصناف المعجمية، لأن هناك حالات خاصة تطرح صعوبات ومعوقات خلال المعالجة نذكر منها على سبيل الذكر لا الحصر، الصفة المشبهة. إن الجذر (م.ر.ض) نولد منه الماضي = [Ø + (مرض)]، لكن السؤال الجوهرية الذي يطرح نفسه هنا هو:

هل تحتاج هذه الحروف المسماة (les octets) (€) الدخول في الجذر أو في أي شيء آخر ؟

لا ينبغي أن تدخل هذه الحروف على الجذر، لأن (€) حرف زيادة، والحالة هذه، لا بد من التمييز بين الجذور البسيطة والجذور المركبة، أي جذور المفردات البسيطة، وجذور المفردات

6- نفسه، مرجع سابق.

المركبة كالأسماء المركبة والتعبيرات المسكوكة ، والمتلازمات المعجمية ، كما ينبغي التفريق بين الجذور الصحيحة والجذور المعتلة ، الأولى هي التي تخلو حروفها الأصلية من حروف العلة ، والثانية هي التي أحد حروفها حرف علة (و، ا، ي) .
وهناك مشكل آخر، لا يقل أهمية عما سبق ذكره ألا وهو هل تعتبر الحركة عنصرا داخل الجذر أم خارجه ؟

انطلاقا من المثال [+ Ø (مرض)] ، يتضح أننا لم نضف أي عنصر إلى البنية النظرية للجذر، بمعنى آخر، ليست هناك أية زيادة حرفية، وكل ما هنالك انصهار للصوامت الثلاثة (م . ر . ض) .
أما بالنسبة للمثال [+م0، و2 (مريض)] ، فقد أضفنا الياء في الموقع الثالث من الجذر، والياء بالنسبة للفعل هي التي تولد الصفة. من هنا يتبين أن لكل مدخل فعلي خوارزمياته الصرفية في توليد الصفة.

هل كل ما قيل عن الفعل والصفة يمكن أن ينطبق على المصادر وأسماء المصادر ؟
يبدو أن الأمر ليس بالبساطة بمكان ، لأن المقولات النحوية لا تكافؤ توزيعيا ، وعدم تكافؤها على مستوى التوزيع يعيق توصيفها على نحو مطرد .لنأخذ مثلا المصادر التي تحتفظ بجميع العناصر التي تدخل في تأليفها وتركيبها ، في حين أسماء المصادر قد ينقصها حرف أو حرفان عن عناصر الفعل .

إذن الفعل [+ Ø (مرض)] أخذ من الجذر (م . ر . ض) ، وليس له اسم فاعل ، ولا اسم مفعول ، بينما الفعل [+ Ø (كتب)] له اسم فاعل [+ 1 (كاتب)] ، واسم مفعول [+ م0، و2 (مكتوب)] ، و لا يمكن أن تأخذ له صفة، لسبب بسيط هو أن الصفة لا تشتق إلا من الفعل اللازم.

يتضح مما تقدم أن توليد أقسام الكلم الأساسية من الجذر تنبني على روزنامة من الخوارزميات الصورية، ذلك أن لكل مقولة وظيفية مولدة من الهيكل الصوتي أي الجذر لها منظومتها الخاصة مع اتفاق في بنية الهيكل الصرفي الأساسية.

فللأسماء البسيطة خوارزمياتها الخاصة التي تولدها من الجذر، كما للأفعال البسيطة خوارزميات أخرى خاصة بها. تتسم خوارزميات أسماء الذوات بكونها تولد بنيات اسمية مغلقة، حيث لا تتفرع عنها إلا بعض البنيات الاسمية الداخلية التي تعتبر من صميم التكوين الاعرابي للاسم، وبالمقابل فإن خوارزميات الأفعال تتصف بالانفتاح على نفسها أولا (أزمنة الفعل وتصريفه مع مختلف الضمائر)، ثم على بنيات أخرى متعلقة بها اشتقاقيا.

ويتجلى الهدف من هذه الخوارزميات الصرفية في أننا نمتد بالقواعد النحوية من أجل استخلاص الصور الموفولوجية ، بهدف تحقيق هندسة لسانية للغة العربية، عبر تحديد الخريطة اللسانية ، ثم إعطاؤها رموزا رياضية ومنطقية مقابل للوحدات اللغوية .

1.4 - خوارزميات التوليد والتحليل الصرفي:

تقوم المعالجة الآلية للغات على ركنين أساسيين وهما التحليل والتوليد، اللذان يؤطران عملية الإنتاج اللغوي، وضمن هذا التصور ينظر إلى الدماغ على أنه آلة لها مدخل ومخرج، ففي المدخل (التحليل) توجد قواعد البيانات التي يتمرس عليه الإنسان، فتتولد لديه المعرفة اللسانية التي ستبني عليها الآلة الخوارزميات اللغوية التي تضبط اللغة في الكفاية، وتتحكم فيها تخزينا واسترجاعا. وفي المخرج (التوليد) ، تتم عملية إنتاج اللغة الانجازية مفردات وجملا ونصوصاً.

1.1.4- خوارزميات التحليل الصرفي

يعتبر برنامج المحلل الصرفي القاعدة الأساس التي تبنى عليها مختلف التطبيقات الحاسوبية المتعلقة باللغة العربية من قبيل أنظمة التصحيح الإملائي والنحوي، وخوارزميات ضغط النصوص وتشفيرها، وبرامج تعليم اللغة العربية لغير الناطقين بها، ورفع كفاءة محركات البحث، ونظم الاسترجاع، ونظم الترجمة الآلية ثنائية الاتجاه .

و من المؤكد أن المعالجة الحاسوبية للصرف العربي ليست جديدة كل الجدة، وإنما هنالك دراسات وأبحاث سابقة مهمة وأساسية نذكر منها على سبيل الإشارة بيزلي 1996، بيزلي 1998، بيزلي 2000، كيراز 2000، الشلبي وأنجليز 1998، بيري زيدوم وعاطف 2001، درويش 2002، وغيرهم كثير. لكن ما يسجل على بعض هذه الأعمال أنها لم تكن شاملة لجميع المستويات اللسانية، أو أنها همشت مستوى من المستويات لسبب من الأسباب، ولناخذ على سبيل المثال الشلبي وأنجليز 1998 ركز اهتمامهما على الأفعال وصيغها، وبيري وزيدوم 2002 اللذين قدما نظاما حاسوبيا لا يستخدم المعجم. في حين اعتنى درويش بجذر الكلمة باعتبارها نتيجة للتحليل الآلي، وتحقيقا لهذا الغرض وظف القواعد اللسانية والإجراءات الإحصائية التي تساعده على معالجة المدونة.

ويرتكز المحلل الصرفي اللغوي في عمله على فحص الخصائص الصرفية والنحوية لقاعدة ضخمة من المفردات العربية المستقاة من تحليل ذخيرة لغوية ضخمة، ولا يعد المحلل الصرفي اللغوي تطبيقا نهائيا، ولكنه أساس لمجموعة من التطبيقات، فهو يدخل في صلب تطبيقات التشكيل والتصحيح والفهم الآلي، ونظم فهرسة المعلومات واسترجاعها، وكذلك تطبيقات التشفير وضغط النصوص. و أهم خاصية تميز المحلل الصرفي اللغوي المبني على الذخيرة اللغوية بتغطية أكبر، وقدرة سريعة على التوسع، وإمكانية تعامله مع النصوص القديمة والحديثة في مختلف التطبيقات الآلية، بالإضافة إلى صغر حجمه في الذاكرة، وإمكانية تناقله بين مختلف أنظمة التشغيل .

وتنقسم خوارزميات التحليل الصرفي إلى أربعة أقسام رئيسية وهي 7:

- 1- خوارزميات التحليل الصرفي القاموسي (المعجمي)
- 2- خوارزميات التحليل الصرفي اللغوي: وتهدف إلى تحديد نوع الكلمة وخصائصها الصرفية، فمثلا كلمة (المحللون) تحلل على الشكل التالي:
 - ال : سابقة أداة تعريف
 - محللون : اسم جمع مذكر سالم
 - ون : لاحقة الجمع
 - حلل : مورفيم الجذر
 - محلل : ساق الكلمة stem.

وقبل تطبيق هذا البرنامج الصرفي، لابد من استحضار العناصر التالية:

- 1- قائمة بجميع الجذور اللغوية ثلاثية أم رباعية (قاعدة بيانات الجذور).
- 2- قائمة بجميع الأوزان الصرفية الأساسية(قاعدة بيانات الخوارزميات الصرفية).
- 3- قائمة بجميع السوابق الممكنة (10-11 سابقة)(مورفيمات الزيادة التصريفية)
- 4- قائمة بجميع الكلمات الجامدة، وهي أقل من 500 كلمة(وتضم الأسماء المشتركة وأسماء الأمكنة، وأسماء الأعلام،...).

3- خوارزميات التحليل الصرفي التبادلي: يشغل هذا البرنامج من خلال إيجاد مجموعة من التباديل من حروف الكلمة المدخلة ، ثم يحاول التأكد من أن الجذر صحيح ، ومن ثم التأكد من أن الحروف الباقية حروف زائدة. فمثلا كلمة (المرتبطون) فسيأخذ الحروف، ويجرب تباديلها ، حيث سيجد (رتب) ، ولكنه (سيجد) من الحروف الباقية حرف الطاء، ولكنه حرف غير زائد ، فسيعلم من أن (رتب) ليست الجذر، وسيأخذ (ربط)، وسيجد أن باقي الحروف كلها مزبدة، و(ربط) هي الجذر. لكن ما يعاب على هذه المنهجية أنها مكلفة بالنسبة للبرنامج المستخدم، حيث تثقل كاهله بمهمات متكررة، يستغرق في تنفيذها وقتا أطول، وتطول المدة أكثر، كلما زاد طول الكلمة عن أربعة أحرف.

4- خوارزميات التحليل الصرفي المصدرية .

ويرتبط المحلل الصرفي بمجموعة من التطبيقات الآلية نذكر منها على سبيل الذكر:

- المصححات الإملائية والنحوية.
- ومولدات المفردات والجمل والنصوص العربية .

⁷- ينظر توليد الأسماء من الجذور الثلاثية الصحيحة، مقارنة لسانية حاسوبية ، دكتوراه قيد الإعداد، جامعة الحسن الثاني المغرب.

- المشكلات الآلية للنصوص العربية، التي تعتمد في الأساس على الخواص الصرفية والنحوية التي ترمز، وكذلك على الإحصاءات المتوفرة من الذخيرة.

- أنظمة الفهم الآلي وحفظ النصوص واسترجاعها.

- تطبيقات التشفير وضغط النصوص.

- الذخيرة اللغوية وتشكل بعد تحليلها وترميزها نواة لكثير من الدراسات الإحصائية اللغوية المختلفة، كما تقوم بدور مساند لبعض التطبيقات التي تستدعي ذلك النوع من الإحصاءات مما يتصل بالتصحيح الإملائي والنحوي وتوليد الجمل.

- النظام الخبير والذي سيقوم بالتعليم من الفريق اللغوي الذي يشتغل على تحليل وترميز الذخيرة اللغوية، واكتساب وتخزين عمله، وسوف يكون كذلك مساندا رئيسا عند الضرورة لتوسيع حجم الذخيرة اللغوية، كما سيعمل على تقليل حجم الفريق اللغوي وتسريع عمله.

وهكذا، يمكن القول إن المحاولات التي أنجزت في هذا الشأن استطاعت أن تتغلب على الكثير من المشاكل التطبيقية في الصرف العربي، لكن ما يعاب على بعضها، أنها لا تعتمد في الغالب على نماذج نظرية لسانية صورية متطورة وقادرة على استيعاب مختلف التطبيقات اللغوية كتعليم الصرف باستخدام الحاسوب، والتدقيق الإملائي، والترجمة الآلية، ولعل هذا كان سببا يحد من شموليتها وإمكانية تطورها بسهولة.

2.1.4- خوارزميات التوليد الصرفي:

تهدف تقنية التوليد الصرفي إلى توليد المشتقات ومزايدات انطلاقا من الجذر، مما يعني أن هذا النوع من المعالجة الصرفية تقوم على عملية التحويل الصرف، تبدأ بالجذر بوصفه مدخلا input، وتمر بالقناة الصرفية الخوارزمية المؤطرة بالقواعد الصرفية والصواتية، وتنتهي بالمرجع output أي المفردة المعجمية،

إذن، لابد لكل مولد أي مشتق إسما كان أم فعلا أن يمر عبر القناة الصرفية الخوارزمية، ولا يوجد عنصر لغوي لا يخضع لهذا الشرط التوليدي.

والتوليد الصرفي يستطيع توليد جميع الأسماء القياسي منها والسماعي، المجرد والمزيد، الصحيح والمعتل.

ينقسم التوليد الصرفي إلى نمطين:

- أ- التوليد الصرفي المعجمي: وهو يتوخى توليد المفردات المعجمية الصحيحة والمعتلة، الثلاثية والرباعية، والمفردة في ضوء هذا النمط من التوليد الصرفي، تكون مجردة من اللواحق.
- ب- التوليد الصرفي النصي: وهو يهدف إلى توليد المفردات كما هي في النص اللغوي الطبيعي، والمفردة في هذا النمط تكون ملتصقة بالسوابق على اليمين، وباللواحق على جهة اليسار، و يوظف هذا النوع في تطبيقات الترجمة الآلية.

وهكذا تعد المعالجة الصرفية الآلية مقوما أساسيا في تحليل النصوص العربية، واسترجاعها، وكذلك في عمليات الإعراب الآلي للجمل العربية، كما يعد المجال نفسه ميذا خصبا للبحث، يتقاسمه علماء اللسانيات، وعلماء الحاسوب، وعلماء النفس، والذكاء الاصطناعي، ورغم حداثة هذا المجال، فقد تمكنت لغات أجنبية كالانجليزية والفرنسية من تطوير برامج مهمة وأساسية نذكر منها على سبيل الإشارة واجهة قواعد البيانات، والتوليد الآلي للحكايات والوثائق و النصوص، والترجمة الآلية، والتعليم بواسطة الحاسوب، ونمذجة العمليات السيكلولوجية وغيرها. أما العربية، فقد أنشأت تطبيقات توليدية وتحليلية للعربية في مراكز بحثية على امتداد العالم العربي بالمغرب وسوريا والسعودية والكويت ومصر والإمارات العربية المتحدة، أقل ما يقال عنها أنها لم ترق إلى المستوى المطلوب، ولم تحقق بعد النتائج المرتقبة.

3.1.4- مراحل التوليد الخوارزمي (الصرفي) :

تولد المفردة بعد خضوع الجذر لسلسلة من التحويلات الصرفية والصوتية يمكن ترتيبها على النحو الآتي :

بعد إدخال الجذر إلى قاعدة بيانات الجذور، يتم اختيار نوع الجذر هل هو ثلاثي أم رباعي، معتل أم صحيح.

تطبق الخوارزميات الصرفية على البنية الصورية للجذر، وذلك بإضافة مورفيمات الزيادة الأصلية أو غير الأصلية، إلى جانب الحركات، وبهذا ينتقل الجذر اللغوي من المستوى الصوتي المجرد المرتبط أساسا بالقدرة المعجمية، إلى المستوى الإنجازي المحسوس أي المفردة أو المدخل المعجمي البسيط.

تفرغ المفردة المولدة في قالب صرفي (وزن/ميزان صرفي) مطابق لها يحدد هويتها بدقة، مما يمكن المستخدم من التأكد من أن الكلمة عربية أو غير عربية.

وبعد الانتهاء من العمليات السابقة، نحصل على قوائم من المفردات النامة البناء

و المرتبة في جداول وخانات على شكل مصفوفات ومجموعات .

4.1.4- قراءة في قاعدة بيانات الأسماء :

تتألف قاعدة بيانات الأسماء من قائمة من الأسماء بنوعها الجامد والمشتق، الصحيح والمعتل، المجرد والمزيد. وهي تضم الأسماء المشتقة مما يتصل بالمصدر (وأنواعه) وإسم الفاعل، واسم المفعول، والصفة، واسم المبالغة، وأسماء المكان والزمان، واسم التفضيل، وأسماء الآلة، والجوامد. ولكل مولد من هذه الصور المعجمية أوزان صرفية تتراوح بين القياسي والسماعي. وتقوم على

أعمدة وصفوف، بحيث تثبت المعلومة بشكل مضبوط ومقنن، يصل عدد الخانات المتضمنة لهذه المعلومات والمعطيات إلى اثنا عشر :

حقل الجذور التي تتفرع عنها الأسماء والصفات.

حقل الجوامد وهو يتكون من المفردة المولدة بشكل صوري، بتطبيق مراحل وخطوات التطبيق الخوارزمي المشار إليها سابقا.

حقل اسم الفاعل.

حقل اسم المفعول.

حقل الصفة.

حقل أسماء الزمان والمكان.

حقل أسماء المبالغة.

حقل أسماء الآلة.

حقل اسم التفضيل.

حقل المصدر وأنواعه.

حقل المختلفات وتضم مختلف التغيرات الصوتية (قلب، إعلال، إبدال...) التي تطرأ على الكلمة مقدمة بشكل صوري قابل للحوسبة .

المتن المعجمي للمدخل، ويقتصر أساسا على تغطية معجمية مختصرة جدا، تشرح المعنى، ويستند هذا الشرح إلى معجمين وهما اللسان والمعجم الوسيط.

5.1.4 نحو بناء نظام خوارزمي لتوليد الأسماء من الجذور وإيجاد المشتقات إلكترونيا :

يندرج بناء هذا النظام الحاسوبي للمفردات العربية في إطار "الهندسة اللغوية العربية" التي تتوخى فتح حوار علمي بين الحاسوب و تطبيقاته و الإنسان العربي بواسطة اللغة العربية بوصفها لغة طبيعية قابلة للمعالجة الآلية، وهذا لن يتأتى إلا بتعميق النقاش الجاد و المثمر في القضايا الأساسية التي تطرحها المعالجة الحاسوبية للنظام اللغوي العربي في سائر مستوياته : الصرفية ، النحوية ، الدلالية ، و المعجمية، و البلاغية من قبيل التوليد الآلي، التحليل الآلي ، الكشف الآلي للنصوص ، الاسترجاع الآلي للمعارف ، الترجمة الآلية ، الفهم الآلي للنصوص ، التعرف على النصوص ، التعرف البصري على الخصائص اللغوية العربية ، و توليف الكلام و تركيبه، وكل ما يتصل بالتحليل والتوليد الآليين . يهدف هذا النظام إلى توليد مفردات اللغة العربية باستخدام تقنيات صرفية وحاسوبية في مرحلة أولى من بناء قاعدة بيانات صرفية ، وفي مرحلة ثانية يتوخى إنشاء قاعدة معرفة صرفية . ولا شك أن إنجاز هذه التطبيقات دفعة واحدة أمر متعسر للغاية، الشيء الذي جعلنا نشرع في مقارنة الأسماء في خطوة أولية في أفق تعميم الدراسة على بقية الأصناف المعجمية الأخرى .

كما يهدف أيضا إلى تقديم خوارزميات صرفية للغة العربية، تمكن من توليد أقسام الكلم عامة، والأسماء البسيطة على وجه الخصوص، ومن ثمة الوصول إلى بناء قاعدة بيانات صرفية موجهة أساسا لخدمة المعاجم الآلية للغة العربية.

إن العمل في إطار "الهندسة اللغوية العربية" يتطلب الإلمام بالعلوم اللسانية بنوعها القديم و الحديث ، و العلوم الحاسوبية التي تهتم بمعالجة اللغات الطبيعية و هندستها .
و إذا تأملنا مسار البحث اللساني الحاسوبي العربي ، فإننا نجد الباحثين و المهتمين بالبرامج الحاسوبية قد غضوا الطرف عن المعرفة اللسانية أثناء المعالجة الآلية للغات الطبيعية، أو أنهم لم يعيروها العناية اللازمة والضرورية، مكتفين بالمعرفة المعلوماتية الصرفة، مما نتج عنه في نقص في النتائج ، دفع بالخبراء العرب في السنوات الأخيرة إلى التفكير في تطوير و صياغة برمجيات عربية ، أقل ما يقال عنها إنها في مهد الطريق ، و هي تبدو متخلفة عن الركب العلمي الذي وصلت إليه برمجيات اللغات الأجنبية .

و لا شك أن هذا الوضع كان سببا من أسباب تعطيل الحوار بين الإنسان العربي بلسانه العربي و الآلة الحاسوبية التي أصبحت تفرض نفسها في جميع الميادين ، و تشكل تحديا من تحديات القرن 21 بامتياز .

و هكذا فإن فتح باب التعاون و التنسيق على مصراعيه بين خبراء اللسانيات و الحاسوبيات لم يعد مطلبا ثانويا ، بل أضحى ضرورة ملحة أكثر من أي وقت مضى قصد النهوض بمستوى البحث الهندسي اللغوي العربي، و جعل لغة الضاد في مصاف اللغات العالمية المنفتحة على التقانة الحديثة و تطبيقاتها المتطورة.

إن المعالجة الآلية للغة العربية تقتضي بالضرورة مراعاة أمرين أساسيين وهما:8
الأمر الأول هو ضرورة الانطلاق من نظرية صورية (خوارزمية) صارمة وواضحة المعالم.
الأمر الثاني هو ضرورة بناء قاعدة بيانات للمفردات اللغوية العربية بنوعها البسيطة و المركبة.
و فيما يتعلق بالأمر الأول فقد وقع اختيارنا النظري و المنهجي على نظرية النحو لتألفي باعتباره نظرية صورية (خوارزمية) تتمتع بمجموعة من المواصفات المرتكزات الأساسية و المفاهيم الإجرائية التي تؤهله لأن يكون نظرية فرعية في اللسانيات العامة و اللسانيات التطبيقية. و نكتفي في هذا الصدد الإشارة إلى بعض الأصول الإستمية و المنهجية التي يقوم عليها النحو لتألفي من قبيل أسلوب تراكم المعطيات و الملاحظة الوصفية و التفسيرية و مبدأ الوصف التصنيفي الذي أثبت نجاه في ميدان البيولوجيا النباتية و الحيوانية و علم الفلك و علم المنطق و الرياضيات الجبرية بالإضافة إلى مبدأي التحليل و التوليد الآليين 9.

أما بالنسبة للأمر الثاني، فيمكن القول إنه الهدف الأساس و القاعدة الصلبة التي يسعى إليها مشروعنا الكبير و المتمثل في بناء قواعد بيانات صرفية للأسماء في اللغة العربية موجهة لخدمة

⁸ - ينظر مقالنا، التوليد الصوري للأسماء في اللغة العربية، المؤتمر الدولي الثاني حول هندسة اللغة العربية و هندسة اللغات، الجزائر، 27، 28

يونيو 2005

⁹ - أطروحتنا، مرجع سابق

المعاجم الآلية للغة العربية والترجمة الحاسوبية من اللغة العربية إلى اللغات الأجنبية أو العكس. و هكذا، فقد تمكنا من إنجاز الشطر الأول من هذا المشروع الضخم من خلال "توليد الأسماء البسيطة من الجذور الثلاثية المعتلة، واليوم نحن بصدد تهيئ الشطر الثاني بتوليد الأسماء من الجذور الثلاثية الصحيحة في اللغة العربية، في أفق بناء قاعدة معطيات كبرى (Macro-Bases de données) تشمل البسيط والمركب من المفردات اللغوية.

1.5.1.4- المفردة في اللسانيات الحاسوبية

تعرف المفردة من منظور حاسوبي بأنها 'متوالية صوتية محاطة ب فراغ من جهتي اليمين واليسار'. وهذا التعريف ينطبق على جميع المفردات اللغوية العربية بما فيها الأسماء و الأفعال و الحروف و الصفات و المصادر الخ، و يخرج منه المتواليات التي تعرضت لتغيرات تركيبية و إضمارية مثل:

أتذكروننا و بمدرستهم، كما يخرج من هذا التعريف أيضا الأسماء المركبة المسكوك منها و العادي مثل:

حيص بيص - أحماسا في أسداس - بصيص أمل... الخ.

فهذه المتواليات تشكل الأرضية الأساسية لقاعدة بيانات الأسماء البسيطة في اللغة العربية الموجهة لبناء المعاجم الآلية للأسماء البسيطة و التي تجيب عن أسئلة ترتبط أساسا بالمورفولوجيا، و الفونولوجيا.

2.5.1.4- مفهوم الأسماء البسيطة

يقصد بالأسماء المتواليات الصوتية بنوعيتها الجامد و المشتق و يندرج ضمن الأسماء البسيطة الجامدة أسماء الأعلام و الأسماء المشتركة و أسماء الأمكنة وغيرها، فيما نعني بالأسماء البسيطة المشتقة المشتقات العربية بما فيها المصدر بأنواعه و اسم الفاعل والصفة، و أسماء الزمان والمكان و أسماء الآلة، و اسم التفضيل، و اسم المبالغة و اسم المفعول.

3.5.1.4- خطة بناء برمجة خوارزمية لتوليد الأسماء وإيجاد المشتقات إلكترونيا :

من أبرز الخصائص التي تطوع اللغة العربية للمعالجة الآلية: الطبيعة الصرفية الاشتقاقية، ذات الإنتاجية العالية و المستمدة أساسا من المكونين الرياضيين: الجذر والوزن. ومن شأن هذا أن يدحض بعض الدعوات 10 التي ترى أن اللغة العربية لغة معقدة وصعبة على مستوى المعالجة الآلية . وبناء على ذلك يصح لنا التساؤل: كيف يتأتى للقدرة الإنسانية أن تستنبط الكلمات (الأسماء البسيطة نموذجا) من الجذور اللغوية، ثم ما هي العناصر التي تتدخل لتجعل من الجذر المعتل (و.ل.د) اسم المفعول (م0، و 2 (مولود) ؟

¹⁰ - هناك العديد من المحاولات العربية وغير العربية التي ترى في العربية غير مؤهلة للمعالجة الحاسوبية لأسباب غير مقنعة علميا من قبيل تعدد صيغ كتابة الحرف العربي، واختلاف تركيب الجملة العربية عن اللغات الأجنبية كالإنجليزية، وغيرها.

للجواب عن هذا الإشكال ننطلق من فرضية أساسية و منهجية مفادها أن الجذر بوصفه علامة صوتية مجردة هو مصدر التوليد و مادة تنظيمية لمدخل المعجم. و لهذا فالعلاقة بين الجذر الصوتي و المدخل المعجمي البسيط (الأسماء البسيطة نموذجاً) لا تعدو أن تكون موجهة تتم وفق روزنامة من القواعد الصرفية و التصريفية هي المعبر عنها بخوارزميات الزيادة . يتجلى دور الجذر أو العلامة الصوتية المجردة في صنع الأساس و البنية الأصلية للمفردة في اللغة العربية، بينما يقوم الوزن الصرفي بتحديد هيكلها العام، وذلك بتوزيع الحركات (الصوائت أو المصوتات) على مختلف حروف الكلمة، كما يقوم بتوزيع خوارزميات الزيادة التي تدخل على مكونات الجذر، و التي تنتج في الأخير الكلمة المولدة، وهذه المورفيمات الخوارزمية هي في الأساس عبارة عن سوابق أو أواسط أو لواحق. هذا التشكيل الرياضي و المنطقي للغة العربية جعلها من لغة انصهارية اشتقاقية في مقابل اللغات الإلصاقية الأخرى (اللغات غير الإعرابية) .

أما الحركات فيتم توزيعها بشكل صارم ومقنن استناداً إلى خوارزميات التطابق بين الوزن و الكلمة المولدة (المنتجة). و المراد بالحركات هنا تلك الوحدات المورفيمية التي تخضع لنظام المحلل الصرفي الطبيعي، و الذي يتولى توزيعها على الحرفين الأول و الثاني بالنسبة للجذر الثلاثي من الكلمة، و الحروف الثلاثة الأولى بالنسبة للجذر الرباعي. أما حركة اللام فيقوم بتوزيعها المحلل النحوي الطبيعي الذي يتعامل مع نظام اللغة بوصفها منظومة من الوظائف الصورية التي لها نظام خاص بها 11.

و على هذا الأساس، فاللغة العربية لغة رياضية يامتياز، و مكونة من أجرومية من الخوارزميات الصورية، دخلها الجذور الصوتية مروراً بالأوزان الصرفية التي تتمتع بقوة الإصهار المورفيمي المبرمج ، و خرجها الكلمات و الجمل.

الجذر (ف ع ل) (سوابق،أواسط،لواحق،حركات)

كلمة = _____

الوزن / الميزان الصرفي

يمثل الجذر في الشكل أعلاه مدخلاً للبرنامج اللغوي المخزن في الكفاية اللغوية للمتكلم. في المرحلة الأولى ، يتم اختيار الجذر المستهدف من العملية ، (جذر ثلاثي أو رباعي). و بعد ذلك تعمل الكفاية اللغوية على تطبيق خوارزميات المطابقة بين الجذر اللغوي و المادة الصورية (ف ع ل). و في مرحلة لاحقة يفعل الوزن عن طريق خوارزميات الإقحام والإدماج التي تقوم بعملية إدراج مورفيمات الزيادة (السوابق، الأواسط، اللواحق) والحركات في البنية النظرية للجذر، بهدف توليد

¹¹ - الحناش محمد، اللغة العربية والحاسوب، قراءة سريعة في الهندسة اللسانية، جامعة الإمارات العربية المتحدة، كلية العلوم الإنسانية والاجتماعية، قسم اللغة العربية، أكتوبر 2002.

الكلمة أو المدخل المعجمي. و يصطلح على هذا النوع من التوليد بالتوليد الخوارزمي الصرفي. و بالمقابل هناك التحليل الخوارزمي الصرفي، و فيه نطبق نوعا آخر من الخوارزميات، وذلك بتفكيك الكلمة إلى عناصرها الأساسية و ما تتألف منه. و لكل صيغة صرفية مقابل معنوي أي دلالي مخزن في الكفاية اللغوية، هذا المكون الدلالي هو الذي يضمن ربط المستوى الصرفي بالمعجم و الدلالة. أما علاقة الصرف بالمستوى الفونولوجي فقائمة بكل تأكيد و لا تستدعي التعليل سيما و أن الأصول المعتلة تشكل مادة خصبة للمتغيرات الصوتية.

يعتبر هذا المشروع استكمالاً لما سبق أن أنجزناه من دراسات وبحوث خلال سنوات (2006، 2005، 2004، 1999) وهو يستفيد أساساً من أعمال سابقة لكل من الدكتور محمد الحناش (شبكة عجمان للعلوم والتكنولوجيا) و يحيى هلال (مختبر المعلومات والعلاج الآلي للغة العربية)، ومرائتي وزملائه، والأعمال المقدمة في كلية تكنولوجيا المعلومات، بجامعة آل البيت، الأردن، ومدينة الملك عبد العزيز للعلوم والتكنولوجيا بالسعودية. ويرمي هذا المشروع إلى وضع برمجة خوارزمية خاصة بتوليد الأسماء، وتزويدها بواجهات استخدام خاصة أي نقلها من الواقع النظري إلى الواقع العملي التطبيقي، كما يسعى القيام بالتحسينات والتطويرات الممكنة باستخراج المشتقات والكلمات والمفردات الناتجة عن الجذر بنوعيه الصحيح والمعتل، والثلاثي والرباعي في اللغة العربية.

ويقوم تشغيل هذا البرنامج بالسماح للمستخدم بإدخال الجذر، والبرنامج يستخرج الكلمة المولدة من هذا الجذر بعد تحريك الخوارزميات الصرفية التي تتدخل في توليد المفردة الإسمية، ومن خلال GUI سيقوم المستخدم بوضع إشارات على الخيارات التي يريدتها، ويقوم البرنامج من خلال ذلك باستخراج الأسماء المولدة من هذا الجذر اعتماداً على تطبيق الخوارزميات.

يهدف هذا العمل إلى معالجة اللغة العربية آلياً مفردات كانت أم جمل بقصد تمكين الحاسوبي بالأدوات الضرورية واللازمة لبناء معاجم إلكترونية، تختلف كلياً عن المعاجم التقليدية الورقية¹²، باعتماد خطة قواعد المعطيات¹³ التي ترتبط بالبرنامج الناتج. كما يهدف إلى بناء مدقق إملائي ونحوي ومشكل آلي بشكل أفضل وكفاءة عالية، ويتوق هذا النظام إلى مواكبة التطورات والتحسينات التي تطرأ على معالجة اللغات الطبيعية الأخرى التي تسير بوتيرة سريعة تنسجم مع عصر التقنية الحديثة.

ويتطلب تنفيذ هذا النظام استخدام إحدى اللغات البرمجية MATLAB أو Visual Java، يتميز النوع الأول بكونه: أداة بيئة تطوير برمجية مخصصة للمهام الحاسوبية، حيث تتوفر على الكثير من

- ينظر مقالنا، التعريف المعجمي بين المعجم الورقي والمعجم الآلي، اللقاء الدولي الرابع للقاموسية، 24/23/22 يونيو

2006¹²

¹³ -Bernard Quemada *Bases de données informatisées et dictionnaires*, Lexique2, PRESSES UNIVERSITAIRES DE LILLE, 1983, p105.

الوظائف والدوال الرياضية المبنية داخليا، والتي تساعد على حل مختلف المعادلات الرياضية ناهيك عن وظائف أخرى. ومما يشرح هذه اللغة لمعالجة العربية آليا أنها تتوفر على مرونة عالية في التعامل مع نظام لغة الضاد بالرجوع إلى قيم ASCII CODE الخاصة بها، كما توفر إمكانيات إحصائية هائلة تذلل علينا صعوبة التعامل مع التطبيقات الأخرى، حيث تحتاج على الأقل إلى عنصرين: الأول خاص بالبرمجة والثاني خاص بالجانب الإحصائي وتقدير النسب.

أما Visual Java فهي لغة برمجة سهلة التعامل وتحتوي على العديد من (Libraries) وال (Build in Function). وتوفر الجافا العديد من ال GUI ذات المرونة العالية في الاستخدام، وقابلة للتعامل مع اللغة العربية بكل يسر وسهولة كما هو الحال مع اللغة الإنجليزية، إذ هنالك طرق للتعامل مع (Strings).

وانطلاقا مما سبق، تم ترتيب مفردات المعجم إلى خمس مجموعات 14:
 للمجموعة المفردات القياسية، وهي التي نرمز لها ب Vs، وهي التي تحتوي على الكلمات التي تم توليدها وتكوينها من سابقة ولاحقة وجذر وواسطة، وهي التي تطرأ على بنيتها الأساسية أي تغيير.
 كلمة ما نرمز إليها ب W من هذه المجموعة يمكن صياغتها:

$$W=(R, P, I, S)$$
 حيث إن:

D - تمثل محمول الاشتقاق (Derivation Predicate).

R - يمثل جذرا من مجموعة جذور اللغة العربية rV .

I - يمثل حشوا من مجموعة أحشاء اللغة العربية Vi .

S - لاحقة من مجموعة لواحق اللغة العربية Vs .

وعلى هذا الأساس نحصل على ما يلي :

$$\{w \mid w \in R, P, I, S, \text{ و } \forall d, P, I, S\}$$

¹⁴ ينظر عبد المجيد بن حمادو، اللغة العربية وشبكة المعلومات، الأنترنت، محاضرة أقيمت بمجمع اللغة العربية الأردني، الثلاثاء 29/1422هـ/22 أيار 2001.

للمجموعة المفردات القياسية المقيدة نرسم لها ب sVd :
 إن داخل مجموعة dV نجد كلمات لا تنتمي إلى القاموس، بالرغم من كونها تتكون من جذر وزوائد
 أي مورفيمات زائدة، ولهذا تم فرض قيد دلالي على هذه المجموعة نرسم له :

$$(R, S, P, I, S)$$

هذا القيد الدلالي يمكننا من تجنب الكلمات غير المقبولة في اللغة أو غير الموجودة في
 قاموس مفردات اللغة مثل أنوقف. وعلى هذا الاعتبار تصبح المجموعة dV الممثلة داخل
 الحاسوب تقتصر على المجموعة المقيدة دلالياً أو محددة sVd حيث :

$$\{ (R, P, I, S) \text{ و } (R, S, P, I, S) \text{ مع } S \in V \text{ و } I \in iV \text{ و } P \in pV \text{ و } R \in rV \} \\ Vds = \{ W$$

مجموعة المفردات المتغيرة التي طرأت عليها تغيرات صرف - صوتية نرسم لها ب tVs .
 ننطلق من العمليتين الأساسيتين الآتيتين :

إبدال حرف بحرف آخر مثل: ن + ت + ث ← ن + د + ث

حذف حرف مثل: ن + ت + ث ← ن + ث .

إذن نسّمّي تغييراً صرفياً صوتياً والذي نرسم إليه ب T تسلسل عمليات مثل التي ذكرنا (حذف أو

$$T = t1, t2, \dots, t3 \text{ (إبدال)}$$

مجموعة المفردات التي طرأت عليها تغيرات صرفية صوتية هي :

$$Vts = \{w \mid \exists X \in Vds \text{ مع } w, w \in (v \setminus vd)\}$$

مجموعة المفردات القياسية المتغيرة:

نرسم أيضاً ب $Vdst$ إلى المجموعة التي نتحصل عليها بتطبيق التغيرات العكسية T^{-1} على

المجموعة Vts داخل مجموعة Vds .

$$Vdst = \{X \in Vds \mid \exists Ti \in T \text{ مع } w, w \in (v \setminus vd)\}$$

مجموعة المفردات غير القياسية :

هذه المجموعة التي لا يمكن صياغتها بالاعتماد على $D(R, P, I, S)$.

نرسم إلى هذه المجموعة ب Ve .

إذن تتكون مفردات اللغة العربية LN حسب النموذج المقترح من اتحاد مجموعة المفردات

القياسية المقيدة بعد استثناء مجموعة $Vdst$ مع المجموعة Vts (المساحة السوداء في الشكل 5).

$$LN = (Vds \setminus Vdst) \cup Vts \cup Ve$$

V+ يمثل كل السلسلات (الكلمات) التي يمكن أن نتحصل عليها من خلال إلصاق الحروف العربية (الألف ياء).

5- خلاصة:

يعتبر مكون الصرف (الخوارزميات الصرفية) في المعالجة الآلية للغة العربية حجر الأساس للمكونات اللسانية الأخرى مما يتصل بالنحو والمعجم والدلالة، وعليه يدور صلب منظومة اللغة العربية. وعليه، فكل تغيب أو تهميش له أثناء المعالجة من شأنه أن يعرقل عمل البرنامج الحاسوبي أو على الأقل يجعل النتائج تتميز بالنقص وعدم الشمولية، بل إن علاقته ببقية التطبيقات الحاسوبية أمر مؤكد للغاية.

ولقد تبين من خلال ما تقدم، أن توصيف العربية حاسوبيا، لا بد وأن يمر في مرحلة أولى من معالجة الخوارزميات الصرفية التي تشكل الأرضية الأساس لبناء مولد صرفي ومحلل صرفي، ومشكل آلي، ومدقق إملائي، و مترجم آلي، وغيرها، ولا شك أن كل ذلك يندرج في إطار تصميم هندسة لغوية عربية تشمل مستويات النظام العربي.

مراجع البحث:

- المراجع العربية:

- ☐ توليد الأسماء من الجذور الثلاثية المعتلة: مقارنة لسانية حاسوبية، مهديوي عمر، أطروحة السلك الثالث، كلية الآداب والعلوم الانسانية، جامعة سيدي محمد بن عبد الله، المغرب، 1999.
- ☐ توليد الأسماء من الجذور الثلاثية الصحيحة، مقارنة لسانية حاسوبية، مهديوي. ع، كلية الآداب والعلوم الانسانية، عين الشق، جامعة الحسن الثاني، المغرب، قيد الاعداد.
- ☐ التعريف المعجمي بين المعجم الورقي والمعجم الآلي، مهديوي. ع، اللقاء الدولي الرابع للقاموسية تونس، 22/21/20 يونيو 2006.
- ☐ التوليد الصوري للأسماء البسيطة في اللغة العربية، مهديوي. ع، المؤتمر الدولي الثاني حول هندسة اللغة العربية وهندسة اللغات، مركز البحوث العلمية والتقنية لترقية اللغة العربية والمجمع اللغوي الجزائري، 28/27 يونيو 2005.

- 📖 اللغة العربية والحاسوب، قراءة سريعة في الهندسة اللسانية، محمد الحناش، كلية العلوم الانسانية والاجتماعية، جامعة الامارات العربية المتحدة، قسم اللغة العربية، أكتوبر 2002.
- 📖 اللغة العربية وشبكة المعلومات، الأنترنت، عبد المجيد بن حمادو، محاضرة أقيمت في مجمع اللغة العربية الأردني، 22 أيار 2001.
- مجلة عالم الفكر، عدد خاص حول الحاسوب، م18، ع3/1987.
- محاضرات الدكتور محمد صلاح الناجم في علم اللغة الحاسوبي، جامعة الكويت.
- المراجع الأجنبية:

- 📖 Bernard Quemada, Base de données informatisées et Dictionnaires, Presses Universitaires De Lille ;1983.
- 📖 Maurice Gross et André Lentin; Notions sur Les grammaires formelles ;collection 'Programmation' ;Gauthier ;Editeurs Paris6/1970.
- 📖 Max Silberztein, Masson ; Dictionnaires électroniques et Analyse Automatique de textes , Masson ; 1993.
- 📖 W.Attalah et Y.Ayach; Introduction Aux Recherches Linguistiques -Arabes Sur Machine ; C.R.A.L.no 23 ; 1973.
- 📖 Les dictionnaires électroniques ; Revue de TAL. vol 44-no 2/2003.