مجلة اللسانيات الرياضية
Journal of Mathematical Linguistics
المجلـد 1 العـدد1/ديسمبر2021
ص ص: 45-69

اللغة العربية والحاسوب؛ قراءة في الهندسة اللسانية العربية

Arabic and computer language; Reading in Arabic linguistic geometry

أ.د/محمد الحناش *

[1] الوكالة الدولية لهندسة اللغات الطبيعية: elhannach@yahoo.com

الاستلام:2020/04/08القبول:2020/05/08 النشر: 2021/12/31

ملخص:

ينحصر الهدف من هذ ا المقال في التعريف بالإمكانات التي يقدمها الحاسوب لمعالجة اللغة العربية بوصفها لغة طبيعية، وخاصة في المجالين البحثي والتعليمي، كما يتعرض للطريقة التي يجب اتباعها لإدخال اللغة العربية إلى عقل الآلة، بهدف إقامة الحوار بين الإنسان والآلة باللغة العربية وقد تبين أنه كلما تطورت تقنيات الحاسوب ازداد قربا من محاكاة دماغ الإنسان في طريقة عمله وفي طريقة تفكيره وتعبيره، مما يفرض وجود دور كبير للغة في هذه المجالات كلها، وعندما نقول معالجة اللغة في هذا السياق فإننا نقصد الهندسة اللسانية، وهذه الأخيرة لا تتوقف عن التقدم والتطور نظرية ومنهجا.

Abstract:

The aim of this article is to define the capabilities offered by the computer to process the Arabic language as a natural language, especially in the research and educational fields. The more computer

* محمد الحناش

technologies developed, the closer it became to simulating the human brain in the way it works and in the way it thinks and expresses, which imposes a large role for language in all these areas, and when we say language processing in this context, we mean linguistic engineering, and the latter does not stop progressing and developing theory and approach .

تزداد هذه العلاقة أهمية بالنسبة للغة العربية التي تنفرد بخصوصيات لسانية صورية جعلت منها، من وجهة نظر الهندسة اللغوية العربية، لغة قابلة للاستجابة بكل يسر للإجراءات الهندسية اللسانية في أغلب المستويات ذات الارتباط بالجانب الصوري وخاصة في بابي الصرف والتركيب.

يتطلب العمل في هندسة اللغة العربية التمكن من نوعين متكاملين من المعرفة: المعرفة اللسانية العميقة وصفا وتصنيفا بمختلف جزئيات النظام اللغوي على ضوء أحدث النظريات اللسانية المعاصرة، وخاصة اللسانيات الصورية، والإلمام بالمعرفة الحاسوبية ذات الصلة بمعالجة اللغات الطبيعية، وخاصة في جانبها البرمجي، وهو ما يعني التأقلم مع التفكير المنطقي الذي تقوم عليه الآلة .فالحاسوب منظومة برمجية منطقية قوامها الخوارزميات الصارمة التي لا تشتغل بالظن أو بالنسبية، ولذلك فإن القواعد التي يجب أن تصاغ لهذه الغاية يجب أن تكون صورية وحاسمة لا تقبل أكثر من تأويل واحد لكل قضية.

إن الجمع بين هذين النوعين من المعرفة هو المحك في التقدم الذي يحققه مجال البحث في الهندسة اللسانية، ولطالما تجاهل واضعو البرامج الحاسوبية الجانب المعرفي اللساني مكتفين بالمعرفة الحاسوبية، وهو ما أدى إلى تطوير برامج عربية يمكن أن نقول عنها إنها ما تزال في بداية الطريق، الأمر الذي انعكس سلبا على صناعة البرمجيات العربية فأخرها مقارنة بالتقدم الذي حصل في البرمجيات الأجنبية، كما أدى إلى تأخير الحوار بين الإنسان

العري بلغته العربية والآلة، هذه الآلة التي أصبحت تفرض نفسها علينا في جميع المجالات. فكيف تم الربط بين مجال الهندسة واللغة ؟

الهندسة فن التحكم في النظم، والحاسوب بشقيه العتادي والبرمجي يقوم على هذا التحكم؛ واللغة نظام معقد متشعب المسالك: كتابة وصوتا وصواتة، وصرفا وتركيبا ومعجما ودلالة وتداولا.

وقد استطاع الباحثون في لغات غربية كثيرة وضع برامج حاسوبية لسانية طبقت فيها جميع الخوار زميات الصورية التي تتعرفها الآلة، وهناك برامج كثيرة تجعل الحوار بين الإنسان الغربي والآلة ميسرا بلغته الطبيعية، نذكر منها: الترجمة الآلية والتوليف الصوتي والتعرف البصري على الحروف والمدقق النحوي والإملائي ...

إلا أن التطبيق على اللغة العربية ما يزال في بداية الطريق، هذا على الرغم من أن هذه اللغة لا تختلف عن أي لغة في العالم فيما يتعلق ببرنامج الكفاية اللسانية، وما تتميز به اللغة العربية عن غيرها يجب أن يدفع بها إلى مقدمة اللغات العالمية في التعامل مع الآلة، فهي تقوم على مكونين رياضيين هما الجذر والوزن، وهما معا غير موجودين في أغلب لغات العالم.

ويتولى الجذر وضع البنية الأساسية للكلمة، ويتولى الوزن وضع هيكلها العام، يقوم الوزن بتوزيع الحركات على مختلف حروف الجذر كما يقوم بتوزيع المورفيمات التي تضاف إلى مكونات الجدر بغرض توليد الكلمات: (سوابق ولواحق وأواسط)

إن هذا التشكيل الرياضي للغة العربية قد جعل منها لغة انصهارية Fusion، خلافا للغات الأخرى التي تعد لغات إلصاقية Ensemblist فيما يتعلق بتوزيع المورفيمات داخل بنية الكلمة الأساس. والقول بالانصهارية يؤدي حتما إلى القول بالطبيعية الرياضية الجبرية للغة العربية.

أما الصوائت التي تعادل Voyelles في اللغات اللاتينية فتوزيعها مقنن بواسطة خوارزميات التطابق بين الوزن والكلمة المنتجة، لذلك لا نضعها – نحن العرب – على الكلمات، لكننا نستطيع قراءتها بدون صعوبة، لأننا نقرأ لغتنا ونتكلمها بالأوزان وليس بالحركات، والمقصود هنا بالحركات تلك التي يتولى المحلل الصرفي الطبيعي توزيعها على الحرفين الأول والثاني من الكلمة في حالة الجذر الثلاثي، والحروف الثلاثة الأولى في حالة الجذر الرباعي، أما حركة لام الكلمة فيسند توزيعها للمحلل النحوي الذي يتعامل مع اللغة بوصفها منظومة من الوظائف الصورية التي لها نظامها الخاص بها..

وهكذا يبدو أن اللغة العربية رياضية بالأساس مكونة من منظومة خوارزميات صورية، من الجذور مرورا بالأوزان التي تتمتع بقوة الإصهار المورفيمي المبرمج وخرجها الكلمات والجمل، وهي كما يلي:

الجذر + (ف ع ل) ز، أ، ل، ح
ـــــــ = الكلمة
الوزن / الميزان

يمثل الجذر دخل البرنامج اللغوي في كفاية المتكلم، في البداية يتم اختيار الجذر المراد تشغيله، (ثلاثي أو رباعي)، ثم تشرع الكفاية في تطبيق خوارزميات المطابقة بين مادة الجذر اللغوي والمادة الصورية (ف.ع.ل)، وفي مرحلة لاحقة يتم تفعيل الوزن/ الميزان عن طريق خوارزميات الإقحام التي تقوم بإدراج الزوائد (سوابق ولواحق وأواسط وحركات) في البنية النظرية للجذر بهدف توليد الكلمة، أما في مرحلة التحليل فإن العملية تكون معكوسة، إذ يتم تطبيق نوع آخر من الخوارزميات تتولى تحليل الكلمة إلى بنيتها الأساسية، أي الجذر.

إن لكل صيغة صرفية مقابلا دلاليا مخزنا في الكفاية، وهذا المكون الدلالي الصرفي هو الذي يضمن ربط المستوى الصرفي بالمعجم والدلالة، وتؤمن ظاهرة الانصهار التي تنفرد بها اللغة العربية حرية الحركة للكلمات داخل الجملة، أما علاقة الصرف بالمستوى الصوتي فلا تحتاج إلى دليل، وخاصة فيما يتعلق بالأصول المعتلة. ونظرا لارتباط كل صيغة صرفية بدلالة محددة تنتظم على شكل حقول دلالية فقد أخذ الصرف نصيبا وافرا من باب الدلالة.

وتتميز الجذور الثلاثية بمرونة واسعة في التحرك داخل البنية اللغوية، خلافا للجذور الرباعية والخماسية.

أما في مجال تركيب الجمل فإننا ننظر إلى الفعل بوصفه دالة Fonction، وأما بقية العناصر فمتغيرات، وذلك كما يلي:

$$\longrightarrow V\ (n1, n2 \ldots nx)\ \ P$$

وقد بني المعجم الإلكتروني لتراكيب اللغة العربية على هذا الأساس (ينظر الملحق) آخر البحث

ونظرا للطبيعة الجبرية لنظام اللغة العربية المتمثلة في منظومة الخوارزميات اللسانية التي تجمع بين مختلف مكونات النظام، فإن الربط بين هذه المستويات أمر بالغ التعقيد لا يمكن أن تقوم به الآلة إلا إذا تم تزويدها بالقواعد اللسانية الصورية، أي بالخوارزميات اللغوية، بنوعيها التوليد و التحليل..

إذا قارنا هذا العمل بمكونات الآلة نجد بينهما تشابها كبيرا في خطوات العمل وفي طبيعة النتائج التي يتوصل إليها البحث في كلا المجالين، وهي تسير وفق ما يلي:

بيانات مشفرة ⟵ ترجمة الخوارزميات ⟵ اللغة الطبيعية

تمثل الخطوة الأولى لغة صورية تتبع تركيبا معينا في بناء الموضوع، وفي الخطوة الثانية يتم ترجمة لغة خوارزميات البرمجة إلى مقدمة لبناء النص باللغة الطبيعية، ثم تأتي في المرحلة الثالثة اللغة الطبيعية التي تعد النتيجة التي يتم من خلالها التأكد من صحة البرنامج من عدمه.

وقد اجتهد الباحثون في الهندسة الحاسوبية في وضع نموذج من الكفاية البشرية على الحاسوب بهدف توفير أسباب الحوار بين الإنسان والآلة. إلا أن نصيب اللغة العربية من هذه البحوث قليل ولا يشمل إلا بعض القطاعات الجزئية التي لا تحتاج إلى خبرة لسانية كبيرة، أما البرامج العربية الكبيرة ، فلا أظن أن البحث قد شرع فيها إلى اليوم، وسنبين أسباب هذا التخلف الذي تعرفه صناعة البرمجيات العربية.

إن امتلاك ناصية الحوار بين الإنسان والآلة باللغة العربية يمثل إحدى الأولويات التي يجب أن توضع نصب عين الباحث اللغوي الخبير بطريقة تصميم الخوارزميات اللغوية وفق نظرية لسانية وصفية يشتغل في إطارها، إلا أن هذا يتطلب العمل في إطار مرجعية علمية ذات طابع لساني صوري.

أما المهندس الحاسوبي فيأتي دوره في المرتبة الثانية، فهو مجرد منفذ للعمليات الحاسوبية اللغوية التي يضعها اللغوي ، ومع ذلك لا يمكن لكل منهما أن يشتغل بمعزل عن الآخر، يجب التعاون المستمر في بناء صرح البرامج اللغوية ذات التوجهات المختلفة: علمية وتعليمية

ومع الأسف الشديد لا توجد لحد الآن لغة برمجة بالحرف العربي، على الرغم من المحاولات التي بذلت في هذا الصدد في بعض مراكز البحوث العالمية، هذا على الرغم من كون اللغة العربية لغة جبرية بالأساس.كما لا

توجد مراكز بحثية يتم فيها التعاون بين المختصين في مجالي الحاسوب واللسانيات الصورية، وقد انعكس هذا التشتت في الجهود سلبا على صناعة البرمجيات العربية.

نظرة تاريخية مختصرة:

بدأ التفكير في الربط بين اللسانيات والتقنيات في أواسط الخمسينات وبداية الستينات عندما شرع الخبراء في وضع برامج الترجمة الآلية، حيث سخرت جميع الإمكانات التقانية لخدمة هذا المجال.

في هذه الفترة كان العلماء يضعون المعدات التقانية لبناء برامج خاصة بهذا المجال البحثي الذي كان ما يزال بكرا في ذلك الوقت، وكانت البحوث تتضمن وضع نظريات في النقل Transfert، وبناء لغات البرمجة، والذكاء الصناعي، وكل مجال من هذه المجالات يتقاطع بطريقته الخاصة مع اللسانيات العامة يقلدها في طريقتها في التعامل مع بنية اللغة البشرية، يأخذ منها ما يفيده في ميدانه.

تبين فيما بعد أن اللسانيات والمعلوميات تتبعان الأسلوب نفسه في بناء النماذج المعرفية لكل منهما، وتوظفان الأدوات الإجرائية نفسها في معالجة اللغة التي كانت تعد الموضوع المشترك بينهما.

بعد هذه الفترة تشكلت أسس ما أصبح يعرف اليوم باللسانيات الحاسوبية التي تعد أحد نتائج الترجمة الآلية ، وصممت محللات تركيبية لهذه الغاية، كما وضعت قواعد صورية تركيبية كثيرة ، كما وضعت خوارزميات لغوية موجهة أساسا لخدمة الترجمة الآلية . وأدى هذا إلى وضع نظريات علمية في إطار اللسانيات الحاسوبية، ومن ثم ارتبط البحث في مختلف فروع علم اللغة بالمعالجة الآلية للغات الطبيعية ثم تطور هذا المجال ليصبح له اسم آخر يعرف به وهو الهندسة اللسانية أو تكنولوجيا اللغات.

وبدأ البحث في الترجمة الآلية في تحديد مفهوم الكلمة، لأنها العنصر الذي كان يعتمد في الترجمة الآلية، إلا أنه ما فتئ أن تطور إلى البحث في المركب (1957)، ومن هناك إلى البحث في الجملة بمختلف تفصيلاتها ، فتعريف الحاسوبيين للكلمة وللمركب وللجملة يختلف عن تعريفات النحاة وغيرهم من اللغويين، وقد ظهرت مدارس مختلفة تنظر إلى هذه المكونات اللغوية نظرة الحاسوبيين، فهم يرون عدم جدوى التعريف إذا لم يكن قابلا للاستجابة لمتطلبات المعالجة الآلية للغة، أي لغة.

ثم بدأت النظرية التوليدية مركبية، ثم ابتعدت عن ذلك لتدخل في دوامة التنظير الذي أبعدها عن الحاسوبيات، وأخذت المعالجة الآلية للغات الطبيعية طابعا هندسيا.

ويفسر البعض تطور المعالجة الآلية للغات الطبيعية بالتعديلات التي أدخلها تشومسكي على نماذجه (1957، 1965، 1972 ...)

ويرى الكثيرون أن السبب الرئيس الذي دعا تشومسكي إلى تغيير نماذجه يتمثل في سعيه الدائم إلى التأقلم مع متطلبات المعالجة الآلية للغات الطبيعية، وخاصة في مجال البحث في المعلوميات النظرية.

ونقل تشومسكى ميدان البحث اللساني في اللغات الطبيعية من مستوى الأداء الذي كان يمثل الشغل الشاغل للبنيويين إلى مستوى الكفاية، فأصبح النحو عنده يعادل الكفاية.

لقد كان تشومسكي وما يزال يعمل بتنسيق مع الحاسوبيين، وقد جاءت جل تحليلاته لبنية النظام اللغوي منسجمة تماما مع متطلبات نسخ الكفاية على الآلة.

أما اليوم فإن نظام المعرفة المتعددة الأبعاد يقوم بديلا عن مفهوم الكفاية كما جاءت في نماذج تشومسكي الأولى.

ويتمثل الهدف الأساس من الذكاء الصناعي ونظام الخبرة في صياغة مناهج تمثيلية للمعارف التي يمتلكها البشر، وهو يوظف أمرين: البرامج والبيانات.

ظهرت نظريات ومناهج لسانية ما بعد التوليدية، نذكر منها نظرية النحو التأليفي الذي أصبح متخصصا في الهندسة اللسانية ذات التوجه الاستقرائي Extension خلافا للنظرية التوليدية التي أصبحت تصنف في إطار التوجه المفهومي Intension بمعناه المنطقي.

تقوم تقنية الاستقراء الحاسوبي على تغذية الآلة بالبيانات المشفرة التي تعد الأساس النظري للتوليد Generation قبل أن تنتقل إلى مرحلة التحليل Analysis. وتقوم تقنية التخزين الاستقرائي على بناء معاجم إلكترونية للغات الطبيعية وفق الاستعمال العادي لكل لغة، مع إبراز خصائصها الصورية في جميع المستويات اللسانية التي صنفتها اللسانيات العامة.

إذ ظهرت عدة مراكز بحثية متخصصة في المعالجة الآلية للغات الطبيعية، وهي على تعددها لا تلتفت إلى اللغة العربية إلا عرضا ، لكن لماذا تخلفت صناعة البرمجيات العربية ؟

سنجيب عن هذا السؤال من خلال تقديمنا للبحث في مجالات المعالجة الحاسوبية للغة العربية

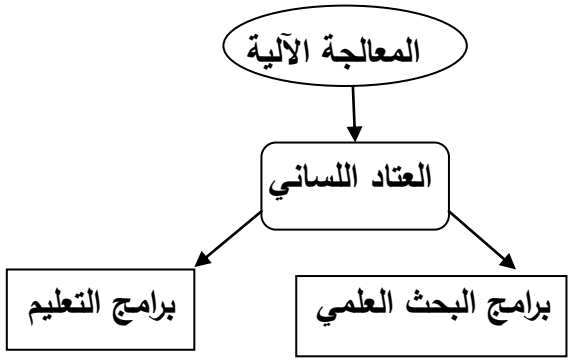
مجالات البحث في المعالجة الحاسوبية للغات الطبيعية: تنقسم المعالجة الحاسوبية للغات الطبيعية إلى ثلاثة مجالات أساسية:

الأول: مجال مشترك، ونقصد به بناء العتاد اللساني.

الثاني: برامج بحثية

الثالث: برامج تعليمية

ويبين الشكل التالي العلاقة بين هذه المجالات:



ونحتاج في المعالجة الآلية للغة إلى ما يأتي:

- العتاد اللساني. - توليد المحارف العربية .
- المعاجم الإلكترونية المتكاملة (أصوات صرف تركيب)
- مولد ومحلل صرفيان - مولد ومحلل نحويان
- المدقق الإملائي والنحوي. - المشكل الآلي

ونحتاج في ذات السياق في البحث العلمي إلى :

- التعرف البصري على الحروف سواء منها المطبوع أو المكتوب باليد
- الترجمة الآلية والترجمة المسعفة بالحاسوب
- التوليف الصوتي. - التوثيق الآلي. - الفهم الآلي للنصوص

وفي جانب التعليم نحتاج إلى :

- تصميم برامج تعليمية على عتاد إلكتروني
- تصميم برامج على الإنترنيت: برامج عالمية

أما برامج العرض مثل: Power Point و Blackboard و WebCT ، وما شابه، فلا يدخل في مجال البحث في اللسانيات الحاسوبية بأي شكل من الأشكال,

مكونات العتاد اللساني:

1-المحارف العربية : من المعلوم أن الحاسوب صنع أساسا بلغات غربية، إنجليزية بالأساس، ولما أصبح تقنية معتمدة في كتابة النصوص باللغة العربية اضطر المبرمجون إلى تصميم برامج تمكن المستخدم العربي من استغلال هذه التقنية باللغة العربية، وفي مقدمتها صناعة الحرف العربي آليا وتنميط لوحة المفاتيح. والموضوع إذن هو ذو شقين:

-**الأول:** يتعلق برسم الحرف العربي على الحاسوب، ويتم ذلك وفق خوارزميات التحكم تضع الحرف في سلمية تسير في الاتجاهين العددين: التصاعدي والتنازلي، تمكن المستخدم من التحكم في حجم البنط، كما تمكنه من تغيير البنط الذي يرغب في الكتابة به.

-**الثاني:** ويتعلق بتوحيد لوحة المفاتيح العربية بين مختلف الدول العربية، تيسيرا لنقل البيانات بين الحواسيب باللغة العربية ونشرها وعرضها وتبادلها، كما ييسر عملية حفظ الوثائق باللغة العربية المدخلة إلى الحاسوب سواء باليد أو بالماسح الضوئي، واسترجاعها وفرزها وتبادلها بين المستخدمين من جميع الدول العربية.

ومنذ دخول الحاسوب إلى الدول العربية قامت منظمة عربية عرفت بمنظمة المواصفات والمقاييس بوضع شفرات المحارف الخاصة بتنميط لوحات المفاتيح الحاسوبية ووضعها في الجهاز رهن إشارة المستخدم العربي، إلا أن المشكل الذي طرح هو تعدد الجهات التي تولت هذه المهمة، كل جهة تسوق لوحة المفاتيح بطريقتها الخاصة، ويذكر المستخدم العربي أول برنامج للتعريب تحت اسم نافذة، وبعده توالت برامج التعريب التي كانت تختلف في تحديد قيمة الأزرار العددية التي تلائم توزيع الحروف على لوحة المفاتيح، ويكفي أن نشير إلى الفرق بين الماكنتوش و أ.ب. م. في أواسط الثمانينات،

وقد بقي نقل البيانات بين الجهازين مستحيلا إلا بعد أن تم توقيع اتفاقية شراكة بين الشركتين في نظام الاستغلال مما أدى إلى الاتفاق حول توحيد لوحة المفاتيح في أواسط التسعينات من القرن الماضي.

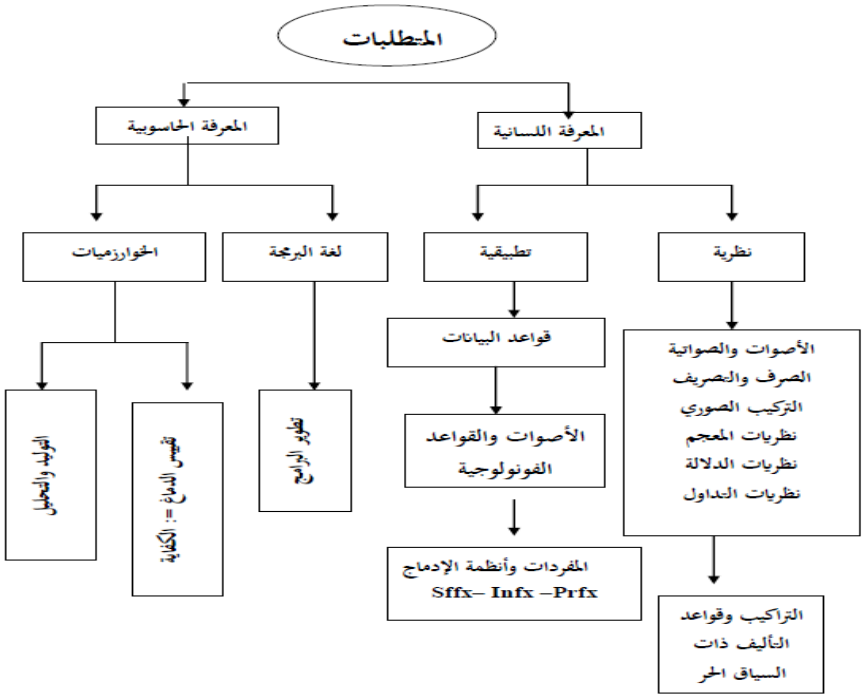
وقد زاد الأمر تعقيدا مع انتشار الإنترنيت، حيث أصبح من الضروري التعامل مع البيانات العربية بشكل موحد، مما دفع الباحثين إلى تطوير لوحة موحدة بلسم Unicode (16 غزوما و 32 عزوما) التي تستوعب جميع الحروف والأشكال العربية وغيرها، فقامت بديلا عن شفرات المحارف العربية القديمة المعروفة باسم: ASMO بمختلف أرقامها: ISO 449، 449، 646، 708 ...

بفضل هذه الجهود التي شارك فيها خبراء لسانيون متخصصون في البرمجة العربية أصبح في إمكان المستخدم العربي اليوم تبادل المعلومات باللغة العربية بين جميع أنواع الحواسيب في العالم، شرط أن يكون قد أضاف إلى بيئة وندوز في حاسوبه شفرة المحارف العربية سواء كانت إحدى إصدارات أسمو المشار إليها أو اليونيكود الذي يستوعب معظم لغات العالم.

2-المعاجم الإلكترونية : المقصود بالمعجم الإلكتروني قاعدة البيانات اللغوية المشفرة، تشمل جميع المستويات اللسانية: الأصوات والصرف والتركيب، بالإضافة إلى بناء معاجم إلكترونية للدلالة، على الأقل في مستواها الصوري الذي يحدد العلاقات المنطقية بين مختلف مكونات المتواليات اللسانية المقبولة في وجهيها الحقيقي والمجازي.

والرسم الآتي يبين الخطوات المتبعة في بناء المعاجم الإلكترونية كما يبين ما تحقق منها في مختلف مراكز البحث اللساني الحاسوبي في العالم، مع ذكر الأرقام التي توافرت لدينا عن كل مستوى، مع التركيز الشديد على

المستويات الثلاثة الرئيسة في البحث اللساني الحاسوبي: الأصوات والصرف والتركيب.



3-برامج البحث العلمي: برامج البحث العلمي في مجال ميكنة اللغات الطبيعية ومن ضمنها اللغة العربية كثيرة ومتنوعة في العالم العربي، ومن المفترض في هذه البرامج أن تقوم على ما تحقق في مجال العتاد اللساني المشار إليه أعلاه. وبما أن هذا المجال لم تتم تغطيته بشكل مرض فلا ينتظر أن تتحقق نتائج كبيرة في مجال الميكنة المتعلقة ببرامج البحث العلمي، ونظرا لتشعب مجالات تطوير البرامج الموجهة للبحث العلمي فإننا سنحصر حديثنا في ثلاثة أنواع منها، وهي: التعرف البصري على الحروف والترجمة الآلية، والتوليف الصوتي.

4-التعرف البصري على الحروف: ويتعلق الأمر بتطوير نظام يتعرف المحارف والهدف منه إكساب الحاسوب مهارة قراءتها قراءة صحيحة سواء

منها المطبوعة أو المكتوبة باليد، تستخدم هذه التقنية عددا كبيرا من خوارزميات التعرف تدخل في مجملها في مجال تقنية معالجة الصور، على اعتبار أن الكتابة كيفما كانت هي نوع من الصور ذات الدلالة الاعتباطية، مما يقوي هامش الافتراضات المتعلقة بتقطيع الصورة إلى كلمات من أجل التعرف عليها، وهناك خوارزميات التجزيء وأخرى يطلق عليها خوارزميات التعرف. ويزيد الأمر صعوبة بالنسبة للغة العربية وجود الحركات التي تقع على مسافة من الحروف، كما أن اشتمال الكلمة العربية على أكثر من ضمير يمثل أكثر من مقولة نحوية يزيد الأمر صعوبة، فلا يستطيع البرنامج التمييز بين بداية الكلمة ونهايتها.

تستخدم هذه التقنية في بعض الأجهزة الإلكترونية التي تشبه مذكرات الجيب، كما تستخدم في بعض مراكز البحث التي أصدرت برامج أولية قادرة على التعرف على الكتابة العربية، أما في اللغات الأجنبية فإن هذه التقنية أصبحت من الأمور التقليدية في البحث.

يتوقف تصميم برامج التعرف الآلي على نجاح برامج رسم المحارف بالحاسوب، كما يتوقف على قواعد البيانات الصرفية التي ترشد القارئ الآلي إلى طريقة تقطيع الكلمات في الواقع الورقي أو الحاسوبي (معالج النصوص).

وما دامت برامج الفئة الأولى المشتركة لم تنجح بعد فإن هذه التقنية لم تنجح إلا بشكل نسبي في التعرف على المحارف المطبوعة، وبشكل ضئيل في التعرف على الخط المكتوب باليد (تنظر ندوة الدار البيضاء 1992).

ومن جهتنا نعتقد أن الأمر يحتاج إلى تطوير كبير في مجال بناء قواعد البيانات العربية التي تيسر على الآلة قراءة المحارف ومطابقتها لما هو مخزن لديها وفق برنامج مبني أساسا على مقتضيات الهندسة اللسانية العربية.

لهذه التقنية نتائج كبيرة على أكثر من صعيد، إذ سيمكن توثيق جميع النصوص التي دونت على مر العصور، كما سيمكنها تقديم خدمة كبيرة لمحققي النصوص العربية الذين يعانون كثيرا في تفريغ المخطوطات على الورق قبل إدخالها للحاسوب، ولا تخفى آثار هذا العمل، إن نجح، على تراث الأمة الذي ما نزال نعاني من قراءته بسبب وضعه المخطوط.

5-التوليف الصوتي: تتمثل هذه التقنية في إكساب الحاسوب مهارة القراءة الصوتية للنصوص المدخلة سواء عن طريق لوحة المفاتيح أو عن طريق القارئ الآلي، بالإضافة إلى إكسابه مهارة تحويل المنطوق إلى مكتوب، وبالنظر إلى الخصوصيات الغرافية للغة العربية التي تتبع نمطا خاصا في الكتابة يتميز بالاقتصار على كتابة الصوامت دون الصوائت، ونظرا للطابع الانصهاري للكلمات في المستوى الصرفي، ونظرا لاعتماد حركة آخر الكلمة في المستوى النحوي، فإن أي تقدم في هذه التقنية يتوقف على اكتمال بناء مكونات العتاد اللساني المشار إليها أعلاه، ولهذه التقنية عدة استخدامات، نذكر منها:

- مساعدة المعوقين على النطق، وهذه شريحة اجتماعية لها أهمية كبرى
- الإملاء الآلي باللغة العربية
- التحقق من هوية المتكلم
- الترجمة الآلية الشفوية وخاصة عن طريق التلفون
- أتمتة بعض المصالح الخدمية مثل قطاع الاتصالات والبنوك وتسيير القطارات معالجة اللغات الطبيعية وغيرها من المصالح ذات العلاقة الآلية بالإنسان

-تدريب متعلمي العربية على النطق الصحيح بأصواتها بحركاتها المناسبة

-الحوار الشفوي المباشر مع الآلة بدل استخدام لوحة المفاتيح أو الفارة

-التلفون المحمول الذي يسير تصنيعه في اتجاه الاستغناء التام عن لوحة المفاتيح مما سيمكن المستخدم ولوج قاعدة البيانات التي يتضمنها ولوجا تاما ومن أي مكان في العالم

إلا أن أيا من مجالات هذه التقنية لم ينجح إلى الآن بشكل كبير في التطبيق على اللغة العربية، والسب هو إغفالها الخبرة اللسانية المشار إليها في باب العتاد اللساني، أما ما تحقق منها فهو الجزء اليسير الذي يعتمد تسجيل بعض الكلمات وجعل الحاسوب يرددها وفق برامج معدة سلفا، وهذا قد لا تكون له علاقة باللسانيات الحاسوبية موضوع هذا البحث.

إن الذي نطمح إلى تحقيقه من وراء هذه التقنية هو أن نكسب الحاسوب مهارة الضبط الصوتي للنصوص المدخلة دون ارتكاب أخطاء إملائية بين الأصوات المتشابهة الصفات والمتقاربة في المخارج، وذلك بهدف الاستعانة بالحاسوب في ميادين تعليمية وبحثية نحن في أمس الحاجة إليها اليوم أكثر من أي وقت مضى.

وقد أنشئت لهذه الغاية مراكز بحث في الأصوات في عدد من الجامعات في العالم نذكر منها مختبر الصوتيات في فرنسا، ومختبر معالجة الإشارة الصوتية في المغرب ومعهد المعلوميات والاتصال في تونس ومعهد الصوتيات في السعودية

6-الترجمة الآلية: يتعامل الحاسوب مع الترجمة بطريقتين:

-**الأولى:** الترجمة المسعفة بالحاسوب، وتعتمد قواعد بيانات مختلفة على شك مخزون مشفر: كلمات وتعبيرات وجمل مع مقابلاتها من اللغة الهدف،

وهي توكل إلى الحاسوب إنجاز ترجمة أولية، يقتصر فيها على ترجمة المفردات التي يتألف منها النص، مما يستدعي تدخل المترجم لإعادة تحرير النص مستعينا بمعاني المفردات التي ترجمها له الحاسوب دون أن يتدخل بشكل كبير في ترجمة التراكيب. كما أن هذا النوع من الترجمة يكون غير قادر على ترجمة المعاني المجازية للمفردات التي تكتسبها في السياق، فهي تقتصر في الغالب على ترجمة المعاني المعجمية للمفردات خارج سياقها التركيبي.

-**الثانية:** الترجمة الآلية وهي التي ترتكز على المعرفة اللغوية الدقيقة في جميع المستويات اللغوية: الصرفية والنحوية والمعجمية والبرا غماتية، إلا أن هذا النوع من الترجمة ما يزال في حاجة إلى تقدم البحث اللساني العام لتكتمل وتنضج وتصبح فعالة. لهذا لا نجد برنامجا واحدا في العالم يقوم بالترجمة الآلية التامة من العربية وإليها، ونرجع ذلك إلى افتقارها لأرضية عتادية لسانية صلبة تعتمد عليها في تطويرها. فالمترجم الآلي يحتاج إلى مدقق إملائي ونحوي يوظفه في عملية الترجمة، كما أنه في حاجة إلى مولد صرفي ونحوي، وهذه لم يكتمل تطويرها بعد.

وأحسن وسيلة للتغلب على مشاكل الترجمة الآلية هي حصرها في قطاعات محددة، يتضمن جميع المصطلحات والتعبيرات المستخدمة في القطاع.

وقد لمسنا من خلال أبحاثنا أبرز النتائج الإيجابية بالنسبة للعرب في النقاط الآتية:

-التفاعل الثقافي

-توسيع المعرفة العلمية للباحثين

-ربح الوقت في ترجمة النصوص المتخصصة

7-المدقق الإملائي والنحوي والمشكل الآلي: لقد جمعنا هذه المكونات الثلاثة تحت بند واحد، وذلك لما بينها من تكامل صناعي برمجي ولساني يحدد الهدف العام منها جميعا. فهي كلها تعتمد المعاجم الإلكترونية التي تحتاج الخبرة نفسها في التصور والبناء. لا يمكن أن نتصور وجود مدقق إملائي لا يقوم على الخوارزميات اللسانية في المستويين الصوتي والصرفي، لأن التدقيق المقصود هو تعرف الحاسوب بنية الكلمة العربية من خلال القواعد التي يضعها اللسانيون، لا من خلال معجم الكلمات التي يكون مزودا بها سلفا. والمدقق الإملائي الحالي يعتمد المخزون المعجمي الموثق سلفا في الجهاز، أكثر من اعتماده على محلل صرفي أو نحوي، وعلى الرغم من أن أغلب البرامج المتعلقة بمعالجة النصوص العربية تذكر من بين خياراتها العادية التدقيق الإملائي والتدقيق النحوي إلا أنها لا تشتغل إلا بما هو مخزن على الجهاز سلفا، ولهذا لا نستغرب إن خطأت كلمات وجملا صحيحة لغويا، والسب في هذا يعود إلى غياب الخبرة اللسانية الحاسوبية في هذه التطبيقات.

الأمر نفسه ينطبق على برامج التشكيل الآلي التي لم تنجح لحد الساعة بسبب عدم اعتمادها المحلل النحوي، وهذا الأخير لم يوضع بعد، على الرغم من أن العديد من الجهات تروج لوجوده، وقد أتيحت لنا فرصة اختبار بعضها فتبين أنها مجرد لغو حاسوبي ليس إلا.

في مجال التعليم: يمكن تقسم برامج التعليم إلى نوعين:

الأول: برامج محلية والمقصود بها وضع برامج تعليمية على أجهزة الحاسوب أو على أقراص مضغوطة، وفتح المجال للمتعلمين للتفاعل مع الجهاز الذي يتحول إلى القيام بدور المعلم، وهنا يكون الجهاز/ المعلم أمام طالب واحد يقوم بدور المدرس الذي يعطي درسا خصوصيا للطالب، وعلى

الرغم من عدم تلبية هذا النوع من التعليم لجميع الأهداف التربوية التي يقوم بها المعلم البشري، فإنه يسهم، وإن بشكل يسير، في عملية التعلم الذاتي.

الثاني: التعليم عن بعد عن طريق الإنترنيت، والمقصود به تطوير برامج توضع على شبكة الإنترنيت بهدف تعليم أكبر عدد ممكن من المتعلمين، إما بشكل متزامن أو غير متزامن، والغاية المرجوة من هذا النوع من التعليم تكمن في توسيع دائرة الحوار المباشر بين أكبر قدر ممكن من المتعلمين تمكنهم من تبادل الآراء عن طريق المساءلة المباشرة وتلقي الرد الفوري، مثال ذلك برنامج Blackboard الذي تستخدمه جامعة الإمارات العربية المتحدة في مشروع الحاسوب المحمول، وقد أدى استخدام هذا البرنامج إلى تحسين العملية التعليمية في الجامعة. تسهم هذه التقنية في ترسيخ مفهوم التعلم الذاتي عن طريق التفاعل بين المتعلم والآخرين من جميع بقاع الكرة الأرضية وهو ما أصبح يطلق عليه التعليم الدولي.

إن وضع أي برنامج تعليمي حاسوبي للغة العربية، في أي من المجالات المذكورة يحتاج بالأساس إلى الخبرتين التربوية وخاصة تقنيات التعليم المعروفة ب(Didactique) واللسانية الصورية، وذلك بهدف مراعاة المستوى التعليمي لمن يوجه إليهم البرنامج، أما الخبرة الحاسوبية فتأتي في المرحلة الأخيرة، أي مرحلة التنفيذ، يضع التربوي المحتوى العام الذي سيقوم عليه البرنامج، ويضع اللساني الخوارزميات اللازمة لهذا المحتوى التعليمي، ثم يأتي الحاسوبي ليترجم الخوارزميات التي وضعها اللساني الصوري إلى أوامر من خلال لغة البرمجة التي يعتمدها. والملاحظ أن أصعب مرحلة هي التي تسند للساني الذي يتولى مهمة بناء الخوارزميات كما يتولى التأكد من صحة النتائج وقابليتها للتجريب وللتغذية الراجعة.

وإذا فحصنا أي برنامج تعليمي مصمم لأي مرحلة من مراحل التعليم، سنجده يحصر خبرته في الجانب التربوي والجانب الحاسوبي، أما الخبرة اللسانية التي تشكل عصبه المركزي فقلما تؤخذ في الحسبان، مما جعل منها برامج غير قادرة على القيام بمهمتها المنوطة بها في عالمنا المتطور الذي أصبحت فيه تقنية التعليم تحتل الصدارة في صناعة البرمجيات التعليمية والبحثية.

ونحن لا نتصور وجود برنامج تعليمي سواء كان للتعليم المحلي أو العالمي ا يقوم على مقتضيات برامج البحث العلمي المشار إليها سابقا، ويكفي الاطلاع على البرامج التعليمية التي تنشر يوميا على شبكة الإنترنيت بلغات أجنبية لنعرف دور البحث العلمي الحاسوبي فيه، فهم لم يشرعوا في وضع البرامج التعليمية إلا بعد أن اكتمل لديهم الوصف اللساني الحاسوبي للغتهم، أما نحن العرب فما يزال وصفنا للغتنا من وجهة النظر الحاسوبية لم يكتمل بعد، ولذلك فإننا نتوقع أن يتأخر دخول التعليم عن بعد والتعليم الذاتي الذي تنادي به جهات كثيرة في العالم العربي إلى مراكزنا التعليمية بمختلف مستوياتها.

وأختم هذه الفقرة بسرد أهم أهداف التعليم بالحاسوب:

- تحقيق التفاعل المتبادل بين الحاسوب / المعلم والطالب
- الإثارة والتشويق والدافعية
- التغلب على الفوارق الفردية
- التعلم لدرجة الإتقان
- جودة إعداد المادة العلمية
- السماح بعرض أنماط تعليمية مختلفة يصعب عرضها بطرق التدريس التقليدية

- التقييم المستمر للطالب
- التعلم الذاتي

والسؤال الذي نطرحه: لماذا تخلفت صناعة البرمجيات العربية؟

كل هذه البرامج التي عرضنا لها لم تصل بعد إلى المستوى المطلوب في اللغة العربية، مع أن هذه اللغة من أكثر اللغات العالمية استجابة للقالب الرياضي سواء بأصواتها أو بصرفها أو بتركيبها، تنطلق الكلمة فيها من الجذور لتفرغ في قوالب صرفية وتصريفية قبل أن تنتقل إلى مكانها المناسب في التركيب الذي تحكمه قوانين صورية مضبوطة. مما يجعل من هذه اللغة بحق لغة جبرية على حد قول المختصين في اللسانيات الحاسوبية.

فما السبب؟ الإجابة: نقص في الخبرة اللسانية الحاسوبية ونقص في فهم الحاسوبي لمتطلبات اللغوي، وانعدام التعاون بين المتخصصين في الاتجاهين كليهما.

ونظرا لما لمسناه من غياب في الخبرة اللسانية التي تعتمدها جل التطبيقات الحاسوبية المدونة أعلاه، قمنا بالعمل في مشروع علمي بالتعاون مع عدد من الباحثين في العالمين العربي والغربي، وذلك في مجالات الأصوات والصرف والتركيب.

ونحيل في هذا الصدد على ملخص المحاضرة التي ألقيناها في السنة الماضية في رحاب هذه الكلية تحت عنوان المعالجة الآلية للغة العربية.

وزيادة في الإيضاح نستعرض مقتطفات من المعجم الإلكتروني للمداخل الصرفية للغة العربية الذي بنيناه تباعا على مر السنين، ويأتي مجهودنا هذا في سياق الأعمال العلمية التي تبذل في العالم العربي لجمع المادة العلمية اللازمة لبناء مولدات ومحللات تتعلق بمنظومة اللغة العربية، فبعد أن استكمل الزملاء في الرياض بتاء قاعدة بيانات الصوتيات العربية ها نحن

أولاء ننتهي من بناء قاعدة بيانات المداخل الصرفية للغة العربية، وذلك بهدف تطوير مولد صرفي ومحلل صرفي للغة العربية، ويعد هذا العمل نقطة البداية لأي عمل لساني حاسوبي تقوم عليه المجالات البحثية المشار إليها أعلاه: التوليف الصوتي، التعرف البصري على الحروف والترجمة الآلية.

كما لا يمكن الحديث عن المدقق الإملائي والمدقق النحوي والمشكل اللغوي إلا بعد استكمال بناء المولد المحلل الصرفي، إذ بدونه يستحيل الدخول في أي مشروع علمي حاسوبي يعود بالنفع على اللغة العربية.

ونحن نعزو الفشل الذي منيت به معظم البرمجيات العربية التي يطلق عليها البرامج التطبيقية إلى غياب الخبرة اللسانية الصورية، فقد اعتمدت التطبيقات السابقة على الخبرة الحاسوبية مع إغفال يكاد يكون تاما للخبرة اللسانية الصورية، قد يعتمد المبرمجون على بعض اللغويين ذوي التكوين التقليدي، إلا أن هؤلاء غالبا ما يكونون غير قادرين على فهم متطلبات الحاسوب فيزودوا المبرمجين بمعلومات أكاديمية صحيحة إلا أنها غير قادرة على الاستجابة لمتطلبات المبرمجين، مما ينعكس سلبا على تطوير البرنامج الموضوع، وهذا بالضبط هو ما فعلته العالمية في برامجها التطبيقية التي تعاني من الفشل على أكثر من صعيد.

إن اللساني الذي نحتاج إليه في وضع برامج الهندسة اللسانية هو الذي المتمكن من مختلف التطورات النظرية التي تعرفها اللسانيات الصورية اليوم، القادر على وضع الخوارزميات اللسانية لمختلف مستويات نظام اللغة العربية. وهذا النوع من الباحثين هم الذين سيتمكنون من الدفع باللغة العربية إلى مصاف اللغات العالمية في الحوار مع الآلة.

لم يعد العمل في حوسبة اللغة مجرد ترف يمارسه بعض الهواة، إنه عصب الأمة والجسر الذي يجب عليها عبوره نحو قرننا هذا الذي نعيشه،

ونعول على الجهات المسؤولة في جمع الخبرات الأكاديمية والتعليمية والمعنوية لتنفيذ هذا المشروع المستقبلي الذي سيرفع من كفاءة اللغة العربية إن في مجال البحث العلمي أو في مجال التعليم الذي أصبح هو الآخر مبرمجا وذاتيا.

الملحق:

واجهة تخاطبية Interface صممت لتوليد أسماء الذات آليا، وهي إحدى الواجهات التخاطبية التي صممت لتوليد مكونات المعجم الآلي للغة العربية، صممت بلغة بيسك. تنقسم إلى مجموعتين من النوافذ: الأولى تتضمن جميع الجذور العربية وعددها يفوق 12،000 جذر (ثلاثي ورباعي)، والثنية تحتوي على مجموع الأوزان الخاصة باسم الذات في اللغة العربية، يفوق عددها المائتي وزن، وبقية الأحياز هي عبارة عن خصائص الاسم المعرب في اللغة العربية: التعريف والتنكير، والتأنيث والتذكير، والتصغير وهلم جرا.

مراجع البحث ومصادره:

أولا؛ باللغة العربية:

1) محمد الحناش، اللغة العربية والتقنيات المعلوماتية المتقدمة، منشورات التواصل اللساني، المغرب 1996

2) محمد الحناش ، المقاربة الحاسوبية للغة العربية، منشورات التواصل اللساني، المغرب، 2002

3) محمد مراياتي وآخرون، نظام اشتقاق الكلمة العربية بالحاسوب، بحث منشور في وقائع مؤتمر اللسانيات العربية والإعلامية، تونس 1989

4) نبيل علي اللغة العربية والحاسوب، منشورات تعريب، الكويت، 1988

5) نشرات وأبحاث على الإنترنيت، أبرزها موقع جامعة مارن لا فاللي بباريس ladl.univ-mlv.fr

ثانيا؛ باللغات الأجنبية:

1. , Z.S. **Harris**, Dunod : Structure mathématique du langage, Paris, 1971

2. Dictionnaires électroniques et analyse automatique de textes, Max Silberztein, masson, Paris, 1993

3. **Maurice** Gross: Méthodes en syntaxe, , Hermann, Paris 1975

4. Les bases de données du LADL: Analyse automatique des langues naturelles, Aspects technologiques, Paris, 1989

5. Jacqueline **LEON**: Traitement automatique des langues, , histoire Epistémologie langage, Paris 2001

6. Eric **Laporte**: Traitement automatique du mot: Etat de l'art, IGM, 2001

7. Unitex, Programme de recherche unicode, IGM, Université Marne la Vallée, Paris 2002

8. Intex; Max **Silberztein**, ASSTRIL 1999-2000, Pris

9. Select pqper on Lexicon-Grammar, Vol. 1-3, LADL,Paris (1973-1999)

10. Jean **Senellart**: Outils de reconnassance d`expressions linguistiques complexes dans des grands corpus. Thèse de doctorat, , Janvier 1999, LADL, Paris.