المحاضرة الثانية: حوسبة مستويات اللغة

1 استخدامات اللسانيات الحاسوبية

إن مجالات استخدام اللسانيات الحاسوبية متعددة وهي قابلة للتوسع والدخول في مجالات أخرى أما أهم المجالات التي تستخدم فيها اليوم فهي:

1- **الإحصاء اللغوي:** ويمكن أن يكون للجذور اللغوية والأسماء والأفعال والمشتقات وغير ذلك.

2- **التحليل الصرفي الآلي:** إن هذه المهمة تفيد جدا الباحثين والدارسين في القرآن والمعاجم والكتب اللغوية وذلك في عمليات الإحصاء للجذور ومشتقاتها واستخلاص نتائج دقيقة تصف أداء المؤلف أسلوبيا.

3- **الترجمة الآلية:** والتي تعتمد على الحاسوب في الترجمة من لغة إلى أخرى وقد وصلت إلى مستوى متطور في هذه الأيام.

4- **الدراسات المقارنة والتقابلية:** هي من مناهج الدراسات اللسانية والاستفادة من القدرة التخزينية للحواسيب وسرعة المعالجة تفيد جدا في هذا المجال.

5- **التدقيق الإملائي والنحوي:** وهي من أصعب العمليات في إكسابها للحاسوب إذ لا يمكن الاعتماد الكامل على الحاسوب في التدقيق اللغوي ولا بد أن يراجعه مختص بعد ذلك.

6- **تعليم العربية للناطقين بغيرها:** إن هذا المجال سيفيد كثيرا جدا من نتائج الدراسات التقابلية بين اللغات والدراسات المقارنة أيضا مما يجعل تعليم اللغة أبسط وأسرع.

7- **تحويل النص إلى كلام والكلام إلى نص:** تعد هذه العملية أكثر العمليات نفعا من تطبيقات اللسانيات الحاسوبية لأنها ستكون مجالا للاستخدام من قبل جميع الناس بينما ينتفع بالعمليات الأخرى بعض الفئات دون البعض.

وننبه إلى بعض الأمور التي يظن البعض أنها من مواضيع اللسانيات الحاسوبية وهي ليست كذلك: "فلا يعد من مجال اللسانيات الحاسوبية تعليم الحاسوب باللغة العربية على غرار ما تقوم به معاهد تعليم الحاسوب في الدول العربية سواء كانت المادة المدروسة تعنى بطريقة استخدام برامج الحاسوب أو تصميم

تلك البرامج وتعديلها أو صيانة الأجزاء الصلبة منه حيث يكون الحاسوب عندئذ مادة الدرس وتكون العربية وسيط التعليم ولغة التخاطب.

ولا يعد من تخصص اللسانيات الحاسوبية ما بات يعرف بمصطلح تعليم اللغة بمساعدة الحاسوب الذي يعنى بثلاث مجالات هي تعليم اللغة الثانية بمساعدة الحاسوب واختبارات اللغة الثانية بمساعدة الحاسوب وأبحاث تعلم اللغة الثانية بمساعدة الحاسوب حيث يعد الحاسوب في هذا التخصص وسيلة من وسائل تعليم اللغة الثانية شأن السبورة وجهاز العرض فوق الرأسي والبطاقات واللوحات وأشرطة التسجيل السمعية والبصرية ويتميز الحاسوب بقدرته على توفير إمكانات كل تلك الوسائل مجتمعة إلى جانب إمكانية التعلم الذاتي الذي لا يستعين فيه الطالب بمعلم.

ذلك لا يعد من صميم اللسانيات الحاسوبية الأبحاث التقنية التي تعنى بآليات تخزين المعلومات في وسائط التخزين الحاسوبية المختلفة التي وثبت وثبات سريعة لتنتقل من القرص المرن إلى القرص الصلب فالمدمج فالذاكرة الضوئية المتميزة بصغر الحجم وضخامة السعة إذ يعد الحاسوب في هذا المجال مجرد وعاء للغة شأن الورقة البيضاء أو سعف النخل وألواح الخشب بادئ أمر التدوين. " الفصل الثالث: اللغة العربية والحوسبة – واقع وتحديات– تعاني العربية اليوم مشكلة وجودية بسبب تأخر أصحابها عن مواكبة التطور الرقمي والتقني وجعله مستوعبا في لغتهم الأم الأمر الذي سيؤدي إلى إقدام الأجيال على الاهتمام باللغات الأجنبية وتعلمها بغية التمكن من التعامل مع الوسائل التقنية الحديثة ويبتعدون عن لغتهم الأم لأنها لا تحقق لهم ذلك ولولا تعهد الله للقرآن والذكر بالحفظ ونزوله باللغة العربية لقلنا إنها اليوم في مرحلة الاحتضار وليس ذلك لقصور منها أو ضعف فيها إنما لتقصير أهلها في خدمتها من أجل أن تخدمهم فيما بعد،

وحوسبة اللغة تعني جعل الحاسوب قادرا على معالجة اللغة إفرادا وتركيبا وممارسة التدقيق الإملائي والنحوي عليها واللغة العربية من هذه الناحية متأخرة جدا فأنت عندما تفعل التدقيق النحوي والإملائي في برنامج تحرير النصوص "وورد" مثلا فإنه لا يطبق أيا من القواعد المعروفة في الإملاء والنحو إنما يعرض الكلمة على القاموس الذي زود به وخزن في الملفات الخاصة بالبرنامج لذلك لا يفرق بين كلمة "المسلمين" و "المسلمون" لأن كلا الكلمتين موجودة في القاموس وتجده يخطئك عندما تكتب "سألتكما" لأنّ قاموسه فيه الفعل سأل وبعض مشتقاته وحالات اتصاله ببعض الضمائر لذلك لا ينتفع بهذه الخدمة كتّاب العربية على النحو الذي ينتفع به كتّاب غيرها من اللغات.

وإن من أهم الواجبات على خدام اللغة العربية والمختصين بها في هذا الوقت العمل على تجاوز العقبات والصعوبات التي تعيق حوسبة اللغة العربية وذلك بوضع حد للفكر المتصلب الجاهلي وابتكار الأساليب المساعدة في الحوسبة والملائمة لخصائص العربية.

2 جهود سابقة في اللسانيات الحاسوبية العربية:

لقد خاض في المجال العديد من الدارسين بعضهم مختص في علم اللغة وبعضهم مهتم أو مختص في علوم الحاسب محب للغة أمته وقدم هؤلاء جهوداً وإسهامات عدة.

إن ما قام به الدكتور إبراهيم أنيس لا يمكن لأحد أن ينكره ولا أن ينكر جهود جامعة الكويت في السبعينات لكن إذا أردنا عرض مراحل التطور لعلم اللغة الحاسوبي في الدراسات العربية المعاصرة أمكن القول بأن كتاب الدكتور نبيل علي "اللغة العربية والحاسوب" الذي صدر عام 1988م يعد أول مؤلف يتناول موضوع اللسانيات الحاسوبية مطبقة على أنظمة اللغة العربية صوتاً وصرفا ونحوا ومعجما مع المعالجة الآلية لهذه النظم اللغوية جميعها.

وبعد نشر هذا الكتاب بسنوات ثمان صدر كتاب د.عبد ذياب العجيلي "الحاسوب واللغة العربية" وهو كما يقول الدكتور نهاد الموسى: "خطوة جزئية إيجابية نحو معالجة مسائل متنوعة من العربية بلغة برولوج وهو يمثل جهدا حميدا في هذا الاتجاه."

و"الواقع إنّ خير دليل على الإسهامات التي يقدمها علم اللسانيات الحاسوبية لمعرفة اللغات البشرية هو الدراسة التي قدمها الدكتور محمد مراياتي بالتعاون مع زملائه العاملين في مركز الدراسات والبحوث العلمية في سورية تلك الدراسة التي تدور حول إحصائية الجذور العربية فقد درس مراياتي الجذور العربية المنتشرة في المعاجم والقواميس العربية القديمة دراسة حديثة معتمداً بذلك على الحاسبات الإلكترونية التي تساعد كثيراً في ضبط العملية الإحصائية والسرعة العلمية فيها وهو ما دفع الدكتور مراياتي لأن يحصي النسب المئوية للجذور الثنائية والثلاثية والرباعية والخماسية في اللغة العربية وقد دفعه أيضاً لأن يحصي الدرجات المئوية التي يمكن فيها للأصوات العربية أن تندمج مع بعضها البعض أو تنفصل عن بعضها بعضاً ثم القوانين التي تحكم هذا الدمج والانفصال والواقع إن هذه الدراسات الإحصائية لجذور الكلمات العربية مهمة بحيث يمكن استخدام نتائجها في الترجمة الآلية من اللغة العربية إلى اللغة الأجنبية الأخرى أو بالعكس ولا سيما من حيث مقابلة المركبات الصوتية العربية مع المركبات الصوتية الأجنبية ومن حيث التحليل والتركيب وقد دعا الدكتور مراياتي هذا الإجراء تنافر الأصوات العربية وانسجامها وإمكانية اكتشاف مثل هذا التنافر والانسجام مبرمجا في الحاسبات الإلكترونية. "

"وآخر هذه المؤلفات في اللسانيات الحاسوبية كتاب الدكتور نهاد الموسى "العربية نحو توصيف جديد في ضوء اللسانيات الحاسوبية" الذي صدر سنة 2000 م ويعد هذا الكتاب أول مؤلف في هذا العلم اللغوي الحديث يصدر عن متخصص في اللغة العربية وعلومها ولذا فهو يمثل نقلة نوعية في توظيف اللسانيات الحاسوبية لخدمة علوم اللسانيات العربية والكتاب كما يذكر مؤلفه أنه محاولة في الانتقال من وصف العربية إلى توصيفها وذلك في ضوء الأطروحة العامة للسانيات الحاسوبية.

وقد اشتمل الكتاب على رؤى حاسوبية حاول من خلالها المؤلف إسقاطها على أنظمة العربية وخاصة النحو والصرف والمعجم إضافة إلى التصويب الأخطاء النحوية والصرفية و والإملائية.

إنّ هذه الجهود التي تمت ضمن هذا الإطار فردية الطابع لكنّ ذلك لم يدم طويلاً إذ سرعان ما أصبحت متعددة الأطراف بعد أن احتضنتها المراكز والمعاهد التقنية والجمعيات الحاسوبية في الوطن العربي وخارجه والمؤسسات والشركات التجارية المحلية. "

3 اقتراحات لحل مشكلة حوسبة اللغة العربية:

قبل البدء في عرض حل مشكلة حوسبة اللغة العربية أذكر هنا نموذجاً اقترحه د.وجدان ليكون منهج اللسانيات الحاسوبية.

اقتراحات د.وجدان : فيما يلي تفصيل يقدم تصوراً تأسيسيا لمنهجية التناول تتعدد فيه المقاربات تبعا لتباين الأنظمة اللغوية الجزئية التي تتكامل فيما بينها وتشكل تلك الأنظمة مجتمعة النظام الكلي للغة العربية

أولاً: توصيف النظام الفوناتيكي:

تؤول العربية في نظامها الصوتي إلى أربعة وثلاثين فونيما ثمانية وعشرون صامتا تتمثل في حروب الأبجدية العربية وستة صوائت تمثلها الحركات الثلاث القصيرة الفتحة والكسرة والضمة والحركات الثلاث الطويلة حروف المد ويعنى في الجانب الفوناتيكي بتوصيف تلك الأصوات آحادا على المستوى الأكوستيكي الفيزيائي...والبادي أن توصيف الأصوات منفردة على هذا النحو سيفضي إلى نتائج دقيقة لا تقبل اللبس وأن حدوداً فاصلة قاطعة ...ستميز كل صوت عن بقية الأصوات.

ثانيا: توصيف النظام الفونولوجي:

تخضع الفونيمات العربية لقواعد فونولوجية تحكم تتابعها في سياق الكلمة أو الجملة وقد يطرأ على الفونيم تغيير في صفته تأثراً بما يسبقه أو ما يليه من الفونيمات ويعني ذلك أن التوصيف الفوناتيكي لأصوات العربية لا يكفي وحده ليمكن الحاسوب من تمييز الفونيمات وأنه يحتاج إلى توصيف آخر لما يعتري فونيمات العربية في سياقها الوظيفي (التضعيف – التجانس والإدغام – ال الشمسية(

ثالثاً: توصيف النظام المورفولوجي:

ينبغي ابتداءً أن نلمح إلى أن توصيف النظام المورفولوجي يهدف إلى تمكين الحاسوب من التعامل مع المورفيمات الوحدات الصرفية التي تشكل التراكيب اللغوية المتنوعة وينبغي ونحن في هذا الصدد التمييز بين مطلبين متمايزين المطلب الأول المتمثل في القدرة على التوليد أي صوغ الأبنية اللغوية السليمة بخطوات إجرائية محددة على نحو شامل مستقص والمطلب الثاني الذي يمكن الحاسوب من التحليل بتعيين المعاني

الصرفية وتفكيك التراكيب اللغوية إلى وحداتها الصرفية الصغرى حين ترد في سياق الجملة أو في سياق النص.

رابعاً: توصيف النظام الإعرابي:

يلزمنا بادئ الأمر تصنيف كلم العربية إلى المبني والمعرب ذلك أن المبني سيتدخل في الحاسوب على هيئة لا تتغير ...وسيفضينا ذلك إلى حصر جهدنا على الاسم المعرب والفعل المعرب.

خامساً: توصيف النظام الدلالي:

يمثل توصيف النظام الدلالي للغة العربية نواة المعالجة الآلية وعموده الفقري ذلك أن جل عمليات المعالجة الآلية للتراكيب اللغوية تستند بشكل من الأشكال عليه وتحيل إليه ويعد عند المعنيين باللسانيات الحاسوبية أعسر المباحث تناولا لتعلق دلالات الألفاظ بداهة بالفهم البعيد المنال عن الحاسوب ...ولعل الأولى ونحن في سبيل إنشاء هذا النظام الدلالي أن نستعين ببنك للنصوص العربية الفصيحة يتم اختيار نصوصه سلفا وفق معايير علمية يرتضيها المتخصصون في العربية.

-وصف خصائص اللغة العربية الخط والتشكيل النحو الدلالة:

تتميز اللغة العربية بالعديد من الخصائص التي تشرفها على غيرها من اللغات ونذكر هنا ما له علاقة بعملية الحوسبة فمنها: أنها لغة معربة، واشتقاقية، غنية بالمفردات، وفيها أكثر من نوع من الجمل، ويمكن التقديم والتأخير لأركان الجملة، والتعبير عن المعنى الواحد بصياغات عديدة، ويكثر في أساليبها الاستعارات والانزياحات اللغوية، ويتميز الخط العربي بالإيجاز في استخدام الحروف فمن ذلك الاكتفاء بحرف واحد عندما يتكرر نفسه متتاليا مثل: "داود" و "يحي" وتحذف ياء المتكلم في كثير من المواضع مثل: "يا رب" و "يا قوم" وكون الحرف له عدة أشكال تختلف في حال الاتصال والانفصال، وأهم من ذلك التعبير عن الصوائت القصيرة بأشكال صغيرة توضع على الحروف ليس كباقي اللغات التي تعبر عنها بمحارف خاصة.

وهذه الخصائص تعطي للغة غنى تعبيريا واسلوبيا يشرفها ويرفع منزلتها مقارنة باللغات الأخرى، لكن بالنسبة للآلة تشكل هذه الخصائص تحدياً وصعوبة في إدراكها ومعالجتها وذلك لأن الحاسوب له خصائصه التي سنعرضها الآن ونوضح لماذا تشكل خصائص اللغة تحدياً له.

4خصائص الحاسوب

السعة الكبيرة: وقد مرت بمراحل عديدة حتى صار بالإمكان تخزين الكم الهائل من المعلومات والبيانات والملفات باختلاف أنواعها على شريحة صغيرة لا يتجاوز حجمها ظفر الإصبع.

المعالجة السريعة: قد تكون هذه الميزة أهم خصائص الحاسوب وهي التي كان لها الدور الأكبر في التفجر المعرفي ولطالما تفاخر الناس على أقرانهم بتطور هذه الميزة في حواسيبهم الشخصية.

الخوارزميات: يعتمد الحاسوب في أداء وظائفه على خوارزميات ومعادلات تنظم خطوات تعامله مع الأوامر والبيانات والملفات وتكون هذه الخوارزميات قد أعدت مسبقاً ووضعت في نظام الحاسب كما يوجد لكل برنامج خوارزميات خاصة به تتناسب مع المهمة التي أعد لها لذلك يمكن وصف نظام الحاسوب بأنه نظام رياضي.

وهناك صفة أخرى للحاسوب ينبغي التنبيه لها كي نتعامل معه بطريقة صحيحة وهي أنه غبي او عاجز عن الفهم. وذلك أنك عندما تعرف الاسم بأنه ما كان مثل رجل فرس حائط فإن الإنسان يقيس عليه كل أسماء العربية التي تعبر عن الموجودات لكن هيهات للحاسوب أن يدرك ذلك ولا سبيل إلا بعملية إحصائية تنص على وصف كل الكلمات الدالة على اسم بأنها اسم وإدراجها في الحاسوب

5 اقتراح المنهج الأنسب:

إن أنسب المناهج في عملية حوسبة اللغة العربية هو المنهج الذي يراعي الخصائص التي ذكرناها لكل من اللغة العربية والحاسوب لأن ذلك سيقودنا إلى وضع الخوارزميات المعبرة عن الوقائع اللغوية والتي يستطيع الحاسوب معالجتها.

والمنهج يجب أن يكون كالآتي:

اتباع ما اقترحه الدكتور وجدان في أولاً وثانياً أي توصيف النظام الصوتي والنظام المعجمي للغة العربية بما يناسب الحاسوب وبعد اطلاعي على جهود محمد مراياتي أرى أن يكون جهده في التوصيف للنظام الصرفي هو المرحلة التالية بالإضافة لما قدمه الدكتور نبيل علي في هذا المجال، أما بعد ذلك فلا بدّ من تزويد الحاسوب بكم هائل من النصوص العربية المشكولة بالكامل القابلة للمعالجة "بنك نصوص لغوية" قبل البدء بالتوصيف الإعرابي النحوي والإملائي والدلالي، وذلك بغية تسهيل عملية وضع الخوارزميات وتوصيفها لتكون مطردة وسهولة كشف ثغرات الخوارزمية لتعديلها وتجاوزها.

ونضرب مثالاً على ذلك بقاعدة التقاء الساكنين إذ سنعطي إرشادا للحاسوب أنه عندما يلتقي ساكنان سيحذف الساكن الأول ونعطيه مثالاً:سعت = سعى +ت

ثم نعطيه أمراً بالبحث عن مواضع التقاء الساكنين في النصوص المدخلة فيه من قبل فيخرج لي كلمة "الضالّين" مثلاً فأعدل على الخوارزمية ذلك بشرط أن لا يكون الحرف الثاني مضعفا. ويخرج لي أيضاً تركيب "عهدي الظالمين" فأضيف للقاعدة إذا كان الساكنان في كلمتين فإنه يحذف لفظا ويبقى رسما. وبهذا يكون توصيف القاعدة شاملاً.

ويلاحظ أني لن أحتاج توصيف حالات التخلص من التقاء الساكنين بالتحريك لأن الحاسوب أصلاً لن يعترض عليها إذ أدخلت فيه خالية من التقاء الساكنين.

وبهذا الطرح أكون قد وفرت وقتي في استقصاء حالات التقاء الساكنين وإدخالها في واحدة واحدة وتجنبت ما قد ألقاه من تناقض وصعوبة في التوصيف والتفريق بين واو الجماعة التي نحركها عند التقاء الساكنين بالضم ونحذفها لفظا أو كتابة ولفظا إذا كانت غير واو الجماعة.

الأمر الذي سيضطرني إلى وضع توصيف للواوات ومتى تكون جماعة أو أصلية أو غير ذلك ثم توضيح الحالات المشتركة، ثم أعود إلى قاعدة التقاء الساكنين، وأضع لها التوصيف المناسب إذا استطعت الانتهاء من الحالات الاستثنائية والخاصة. ويظهر من الذي ذكرته تناسب الطريقة التي ذكرتها مع خصائص الحاسوب وخصائص العربية فقد استفدت من سعة الحاسوب الكبيرة ومعالجته السريعة، وتخلصت من تعدد احتمالات دلالة الكلمة على فعل أو اسم أو ماض وأمر باعتمادي النصَ المشكولَ وكمّا هائلاً من النماذج والوقائع اللغوية مما يجعل وجود تعبير مخالف لما قعدته نادراً جداً

فما هي أسس هذا المنهج؟

أولاً :الاستفادة من الجهود المبذولة في دراسات النظام الصوتي والمعجمي للغة العربية مع التأكيد على وضع رمز لتصنيف الكلمة كاسم أو فعل أو حرف وربما احتجنا بعض التصنيفات الفرعية كضمير واسم إشارة واسم موصول أيضاً كما هو موجود في القواميس الإنجليزية.

ثانياً :اعتماد النص المشكول بالكامل قبل البدء بعمليات التوصيف وذلك لأن الحركات في العربية تعادل حروفا مرسومة في غيرها.

ثالثاً :تخزين كم هائل من النصوص العربية المشكولة بالكامل القابلة للمعالجة بنك نصوص لغوية في ذاكرة الحاسوب.

رابعاً: البدء بالتوصيف الصرفي والنحوي والإملائي مستفيدين من الجهود السابقة والبناء عليها. خامساً : أرى أن توصيف المستوى الدلالي سيظل التحدي الكبير في وجه اللسانيات الحاسوبية وربما يقدم نحو النص مساعدة وتيسيرا في هذا المجال إذا تطور أكثر، لكني أنبه لعملية مفيدة جداً في هذه الناحية بعد أن صار متوفراً عندي كمّ هائل من النصوص العربية المشكولة؛ وهي عملية مقاطعة النصوص مع بعضها البعض مما يتيح للحاسوب معرفة كيفية اصطفاف الكلمات في الأساليب العربية ونماذج عديدة لصياغة العبارة العربية، لنفيد من ذلك في توصيف المستوى الدلالي.

ولا بدّ قبل ختام هذا الفصل من الحديث عن ربط الجهود السابقة والاستفادة من جهود مبذولة في مجالات أخرى حيث أرى أنّ أولاً وثانياً قد صارا جاهزين بفضل الجهود السابقة ولا تحتاج إلاّ بعض التكملات

وتنسيقها معاً لخدمة هذا المنهج ولعل سائلا يسأل: كم سنة تريد حتى ننجز بنك النصوص اللغوية المشكولة؟ ألست تبالغ في هذا المقترح؟

والجواب: أنه جاهز بفضل جهود العاملين على توفير الكتب بطلاب العلم ضمن برامج وتطبيقات تساعدهم على البحث وأخص هنا "المكتبة الشاملة" لأنها تقدم لك النص مشكولا وبدون تشكيل ولا يجهل من سمع باسمها فقط ضخامة الملفات والنصوص التي تحويها فعندما نضم هذا الجهد للجهود السابقة ونطبق ما اقترحته من خطوات سنحصل على نتائج قيمة جدا.