

المحاضرة الثانية

**المعالجة الآلية للغة العربية  
بين الواقع والتحديات**

المهندس علاء الدين صلاح العجاوي  
رئيس مجلس إدارة شركة أنظمة  
المعلومات العربية- القاهرة

الثلاثاء 1 صفر 1422هـ - 24 نيسان 2001م



بسم الله الرحمن الرحيم

يطيب لي بداية أن أتوجه لمجمعكم الكريم ممثلاً في رئيسه الأستاذ الفاضل  
بغاية الشكر على دعوتي للحضور ، وإتاحة الفرصة لي أن أتقدم بهذه الكلمة  
المتواضعة التي أرجو من الله عز وجل أن يوفقني في عرضها بالكيفية التي  
توضح ما بها من أفكار بنيت على تجارب استمرت من عام 1987 في حقل  
التطبيقات اللغوية العربية على أجهزة الحاسوب.

وقد كان لي شرف البدء في هذا الطريق الطويل بإنتاج أول مدقق إملائي  
عربي على أجهزة الحاسوب في ذلك العام ، تلاه عدة منتجات أخرى في المجال  
نفسه تشمل التحليل الصرفي والتدقيق النحوي والمعجم وحتى منتجنا الأخير  
المعربي بالترجمة الفورية للكلمات والعبارات البسيطة.

وقد كان يحركني في هذا كله إيمان راسخ بوجود استخدام التقنيات الحديثة  
لخدمة لغتنا العربية ، وتطوير هذه التقنيات كي تتعامل مع اللغة العربية بالشكل  
الذي تستحقه هذه اللغة التي أثق في أنكم جميعاً تشتركون معي في عشقها.  
ولقد اخترت عنواناً لهذه الكلمة:

### "المعالجة الآلية للغة العربية بين الواقع والتحديات"

وهو اسم يعبر عما يجول بخاطري من ضرورة تكثيف الجهود المشتركة بين  
علماء اللغة والمتخصصين في كافة العلوم الأخرى من أجل تذليل أية مصاعب  
تواجه تعميم الاستخدام الأمثل للغة العربية في كافة مجالات الحياة.

وسوف أتعرض في هذه الكلمة للنقاط الآتية:

أولاً: مقدمة عن تقنيات المعالجة الآلية للغات عموماً واللغة العربية على وجه الخصوص.

ثانياً: الخصائص المميزة للغة العربية.

ثالثاً: حتمية تذليل المصاعب التي تواجه المعالجة الآلية للغة العربية.

رابعاً: مثال لبعض العقبات التي تواجه المعالجة الآلية للغة العربية.

خامساً: أهمية تعميم الاستخدام السليم لحركات التشكيل في الكتابة العربية.

## مقدمة

مع التنامي المتزايد لتطبيقات تكنولوجيا المعلومات في كافة مجالات الحياة، بحيث لم يعد هناك مجال من المجالات يخلو من مثل هذه التطبيقات، ومع التزاوج غير المسبوق بين تكنولوجيا المعلومات وتكنولوجيا الاتصالات، هذا التزاوج الذي أدى إلى استحداث تعبير القرية العالمية (The Global Village) الذي أصبح بحق هو سمة العصر، لم يعد من الممكن لأي أمة أن تقف في معزل عن هذا التطور التكنولوجي المتسارع، وإلا فقد حكمت على نفسها إما بالتخلف عن ركب العصر أو فقدت لغتها ومن ثم تراثها، وستجد أجيالها الجديدة نفسها مضطرة إلى استخدام لغة أخرى لا تبعدها عن ثمرات الحضارة، وإذا كان كلا الأمرين مرفوضاً من أي أمة تعتر بنفسها وتراثها، فإن الأمة العربية تكون مسؤولة أكثر من غيرها من الأمم أمام الله رب العالمين والحضارة البشرية جمعاء عن الحفاظ على لغتها العربية لغة القرآن الكريم.

ولقد تجلت حكمة المولى عز وجل بلختياره وتشريفه للغة العربية بترتيل القرآن الكريم بلسان عربي مبين، فللبناء الرياضي (Mathematical Structure) غير المتكرر للغة العربية أتاح لها القدرة على استيعاب العلوم الحديثة والتطور غير المحدود تبعاً لمقتضيات العصر من دون أن تفقد طابعها المميز ولا سماتها الأساسية.

وتعتبر تقنيات المعالجة الآلية للغات هي البناء التحتي الأساسي (Basic Infrastructure) الواجب توفيره، حيث يتم فوقها بناء كافة التطبيقات الأخرى، ولا تقف تقنيات المعالجة الآلية حالياً عند حد إدخال البيانات والمعلومات واستخراجها (Input / Output)، بل تمتد إلى التطبيقات اللغوية التخصصية من تدقيق إملائي (Spelling Checking) ونحوي (Grammar Checking) وتحليل صرفي (Morphological Analysis) وترجمة آلية (Machine Translation)،

وكذلك تقنيات التعرف الضوئي (Optical Character Recognition) على الكتابات الآلية (Type Setting) واليدوية (Hand written)، وتمييز الحديث (Voice Recognition) وقراءة النصوص (Text to Speech)، وما إلى ذلك من تقنيات أساسية.

ولقد قامت عدة مؤسسات وشركات عربية رائدة بجهود كبيرة لا يمكن إغفالها في مجالات المعالجة الآلية للغة العربية، بحيث أصبح من المتاح الآن الحصول على العديد من التقنيات الأساسية مثل التدقيق الإملائي والنحوي والتحليل الصرفي والقواميس والمعاجم والمترجمات ذات المستوى الأول (Level 1 Machine translators) التي تلبي احتياجات المستخدم العربي، إلا أن الطريق ما زال طويلاً في اتجاه المزيد من التقنيات حتى نساير المتوافر منها في اللغات العالمية الأخرى.

ويشهد العقد الحالي من القرن الحادي والعشرين تدفقاً سريعاً للغاية لتطبيقات تكنولوجيا المعلومات من نظم تشغيل، وبرامج، وتطبيقات، لا يخفى ما لتطويعها للتعامل مع اللغة العربية من أهمية حيوية وأثر إيجابي، ولا يخفى على المراقب المحايد أن الوقت الذي يمضي بين طرح هذه البرامج بلغتها الأصلية التي غالباً ما تكون الإنجليزية، وبين طرحها باللغة العربية وقت طويل للغاية مقارنة مع الإصدارات باللغات الأخرى أوربية كانت أو شرق آسيوية، وفي عصر يحسب فيه الزمن بكسور الثانية وليس بالأيام أو الأسابيع، يمكن الحدس لها في ذلك من أثر سلبي.

وإذا كان حجم السوق العربي الحالي يقف عقبة أمام المزيد من الأبحاث المطلوبة في هذا السبيل، حيث إن حجم الطلب ما زال يقل بكثير عن حجم مثيله في الدول الأخرى (ولعل نسبة الحواسيب إلى تعداد السكان هي أكبر دليل على ذلك، فبينما تبلغ هذه النسبة في الولايات المتحدة وكندا نسبة 4:1 وفي أوروبا

8:1 تبلغ هذه النسبة في العالم العربي حوالي (1:200)، كما أن أسلوب الكتابة العربية الحالي يمثل عقبة أيضاً في نفس الاتجاه، وأيضاً فإن غياب حركات التشكيل في الكتابات الحديثة، وغلبة الشكل الجمالي على القيمة العلمية، وغير ذلك مما سوف نتعرض له بالتفصيل في ورقة العمل هذه كل ذلك يؤثر تأثيراً سلبياً على هذا المجال.

وكما قام أجدادنا الأوائل في العصور الإسلامية الأولى بتطوير أساليب الكتابة العربية حتى تتفق ومقتضيات العصر، فقاموا بإضافة حركات التشكيل، وتنقيط الحروف العربية، فقد أصبح من الواجب على جيلنا الحالي أن يستمر في تحقيق التناغم بين التقنيات المعلوماتية الحديثة وأساليب الكتابة العربية، بحيث تحقق حلاً فعالاً ليس فقط للمشاكل التي قد تعترضنا الآن، بل لما قد يستجد مستقبلاً من تقنيات الله أعلم بها.

وإذا كان الأوروبيون في العصور الوسطى قد قاموا بتغيير أشكال الحروف الغربية حتى تتفق مع تقنيات الطباعة التي استجبت في ذلك الحين، كما لم يجدوا أية غضاضة في استخدام الأرقام العربية (Arabic Numerals) وما زالوا يستخدمونها حتى الآن، فإن ذلك أدعى أن ندرس بكثير من الاهتمام والجدية ما قد تحتاجه أساليب الكتابة العربية من تحديث، بمنظور متفتح يحافظ على التراث قدر سعيه لتذليل العقبات.

وفي ورقة العمل هذه عرض متواضع لبعض المشاكل التي تعترض مسيرة المعالجة الآلية للغة العربية، ولا يدعي كاتبها التخصص الدقيق في المجالات اللغوية، أو قدرته منفرداً على طرح الحلول الشافية، بل يعلن من البداية أنها فوق طاقته منفرداً، بل لنقل أنها استعراض لهموم باحث عربي قضى سنوات طويلة من عمره في مجالات المعالجة الآلية للغة العربية، يعتز بلغته وإسلامه،

ويأمل أن يظل أبنائه وأحفاده ومن بعدهم إلى ما شاء الله على نفس النهج، وهي دعوة لكافة الأفراد والجهات المهمة للتكاتف سوياً والبدء في دراسة هذه النقاط.

### المراحل المختلفة لأي لغة

Lexicon	المفردات
(word list)	
Morphology	الصرف
Grammar / Syntax	النحو
Symantec	الدلالة
Rhetoric	البلاغة

### الخصائص المميزة للغة العربية

في هذا السبيل لابد أن نبدأ بالتعرف على الخصائص المميزة للغة العربية وما أكثرها، فلا يقف الحد عند الكتابة من اليمين إلى اليسار (RTL)، أو تغيير أشكال الحروف بتغيير موقعها من الكلمة (Context Analysis)، بل أن الدارس للغة العربية لا يستطيع أن ينكر العديد من المزايا الأخرى التي لا تجتمع لأي لغة أخرى.

بدايةً ففوة أي لغة تكمن في قدرتها على التعايش مع العصر الذي تعيشه، وعلى التأقلم مع مفرداته، واستيعاب مستجداته، كل ذلك في أسلوب علمي يكفل الدقة في توصيل المفاهيم بأقصى كفاءة، واللغة أي لغة ليست مجرد أسلوب

كتابة، بل هي نسق كامل من المفاهيم والمشارب التي تعبر عن نفسها حديثاً  
وكتابةً وقراءةً بل وتفكيراً، لنقل أنها وعاء كامل لتراث الأمة.

وللغة العربية بناء رياضي فريد لا يتوافر لأي لغة أخرى، فأسلوب الاشتقاق  
الفريد من الأفعال والأسماء بنسق رياضي دقيق يتيح لها استيعاب أي مصطلح  
جديد والتعبير عنه بطريقة تلقائية مبسطة تقترب من العامة قبل الخاصة، بل  
وتطويع مثل هذا المصطلح إلى موسيقاها الخاصة بما يتفق مع الأذن العربية  
السليمة، ويضاف إلى ذلك قوانين رياضية واضحة لإضافة السوابق واللواحق  
(Prefixes and Suffixes) لأي كلمة بما يضيف إلى معناها ولا ينقص منه.

ولو تأملنا فكرة تغيير الموقع الإعرابي للكلمة في وسط الجملة تبعاً للمعنى  
المقصود، وما يعنيه ذلك من تغيير في تشكيل آخرها أو في حروفها، وتأثير ذلك  
على وضوح معناها، لتيقنا كم هي قوية هذه اللغة، وكم تستحق من مجهود وعمل.

وفي عجلة ومن دون الدخول في الكثير من التفاصيل التي لا يسمح بها  
الوقت أحب أن أذكر بعض النقاط التي تميز اللغة العربية عن غيرها من  
اللغات...

أولاً: كإحصائيات مقارنة سريعة بين اللغتين العربية والإنجليزية...

فلن إجمالي عدد الكلمات باللغة الإنجليزية يتراوح بين أربعمئة إلى خمسمئة  
ألف كلمة مستخدمة.

بينما تتعدى الكلمات المستخدمة في اللغة العربية عدة ملايين (أخذاً في  
الاعتبار الكلمات المشتقة والمزيدة).

**ثانياً:** في الوقت نفسه فلن جذور الكلمات العربية أقل بكثير من مثيلاتها الإنجليزية، ولكنها تخضع لجداول حاكمة يمكن من خلالها الاشتقاق المباشر لهذا العدد الكبير جداً.

**ثالثاً:** يمكن حصر حوالي مئة وأربعة وأربعين جدولاً رياضياً سليماً نتعامل تقريباً مع جميع الأفعال العربية من دون استثناء.

**رابعاً:** تمثل تركيبات السوابق واللواحق في اللغة العربية ثراء لا حد له، ويمكن من خلالها التعبير بالكلمة نفسها عن الأفراد والتنثنية وضمائر الفاعل والمفعول والأزمنة والاستفهام ... الخ.

**خامساً:** يمثل البناء النحوي للغة العربية عمقاً دلالياً لا نظير له، حيث تقود حركات الإعراب المختلفة إلى حسم العديد من أوجه اللبس حسماً لا يدع مجالاً لسوء الفهم.

### **أمثلة لبعض العقبات التي تواجه المعالجة الآلية للغة العربية**

لا يخلو الأمر من بعض نواحي القصور في أسلوب الكتابة العربية الحديثة، خصوصاً قد يؤدي إلى الكثير من السلبيات عند محاولة تطبيق التقنيات الحديثة في أساليب المعالجة الآلية للغة العربية، وفي هذه الورقة سوف أتناول بقليل من التفصيل فكرة قلة حروف الحركة (Vowels) العربية، والاعتماد الكبير على حركات التشكيل من فتح وضم وكسر التي قلما تستخدم، وتأثير ذلك السلبي على تقنيات المعالجة الآلية.

فلو قارنا اللغة العربية بلغة مثل الإنجليزية مثلاً، لوجدنا أن الإنجليزية تحتوى على سبعة حروف متحركة أساسية (a, e, i, o, u, y, w) مقارنة بثلاثة حروف عربية (أ، و، ي)، كما أن تركيبات حروف الحركة الإنجليزية كثيرة جداً (مثل ea, ee, ei, ie, oa, ui) مقارنة بندرة تركيب حروف الحركة

العربية، وهذا الأمر مرده أساساً إلى اعتماد العربية على حركات التشكيل، وبكل أسف فإن الكتابات العربية الحديثة لا تستخدم حركات التشكيل إلا فيما ندر.

ينتج عن ذلك تشابه مفردات عربية كثيرة في حروفها الأساسية مع اختلاف بين في معناها ( مثل حروف ك، ت، ب) فقد تقرأ كَتَبَ ( الفعل الماضي) وقد تقرأ كُنْتُب (جمع كتاب)، وبالمثل العديد من الكلمات العربية التي تكتب بالحروف نفسها وتتعدد طرق قراءتها ومعانيها، ومرة أخرى عند المقارنة باللغة الإنجليزية فسوف نكتشف من دون عناء أن نسبة اللبس في العربية غير المُشكَّلة تزيد أضعافاً مضاعفة عنها في الإنجليزية، علماً بأن اللبس يتضاءل جداً في حالة التشكيل بحيث يكاد ينعدم.

وكمثال عملي لنفترض مدققاً نحويّاً يفحص عبارة "كتب المدرسين" هل هي خطأ أم صواب، بالتأكيد سوف يأخذ في الاعتبار حالة الفعل والفاعل وهي خطأ، وحالة المضاف والمضاف إليه وهي صواب، وللمترجم الآلي أهي "Teachers' books" أو "Teachers wrote" ؟، ناهيك عن برنامج قراءة النصوص (Text to speech).

### حتمية تذليل المصاعب التي تواجه المعالجة الآلية للغة العربية

تمتد تطبيقات البرمجيات امتداداً لا يحده سوى الخيال البشري، فمن أول الألعاب التي تخاطب الصغار، وحتى أعقد النظم العلمية والصناعية، يوجد هناك طيف هائل من التطبيقات التي لم تدع مجالاً من مجالات الحياة إلا دخلته، ومع تنوع هذه التطبيقات لم يقتصر أسلوب إدخال واسترجاع البيانات على لوحة المفاتيح النمطية (Traditional Keyboard)، ولا على شاشة البيانات (Monitor/Terminal) والتقارير الورقية (Reports / Hard copies)، بل تعدتها إلى التقنيات الحديثة من التعرف ضوئي على النصوص المكتوبة (Optical Character Recognition) وتمييز الأصوات البشرية (Voice recognition)، واستعراض البيانات على هيئة كلمات مسموعة (Text to speech).

ويوماً بعد يوم يزداد انتشار مثل هذه التطبيقات في كافة المحافل نظراً للمزايا الشديدة التي توفرها لمستخدميها، فمن السهل على سبيل المثال تخيل الوقت والمجهود الذي توفره تقنية التعرف الضوئي مثلاً في حالة الرغبة في تغذية نظام معلومات ما بمئات الصفحات من المدخلات النصية، أو مدى اليسر الناتج عن تطبيق نظام حسابات مصرفي يتيح للعملاء التعرف على أرصدة حساباتهم من خلال الاتصال الهاتفي، أو الأفاق الرحبة التي يتيحها نظام ترجمة فورية للأحاديث اليومية، كل هذه التطبيقات يبني أساساً على تقنيات المعالجة الآلية للغات المعنية.

إذاً فليس أمامنا كعرب إلا أن نتوكل على الله ونمضي قدماً في مجال البحث والتطوير لتقديم تقنيات معالجة آلية للغة العربية، وإذا تقاعسنا للحظة، فسوف تزداد الفجوة التكنولوجية بيننا وبين الآخرين بما ينعكس سلباً على حاضرنا وحاضر الأجيال التي تأتي من بعدنا.

ومن على هذا المنبر منبر مجمع اللغة العربية الأردني وما يمثله من حصن من حصون اللغة العربية فإنني أتوجه بدعوة أرجو أن توافقوني عليها إنني أدعو إلى:

- ١ - ضرورة تكثيف الجهود المبذولة في مجالات الأبحاث والتطوير لتقنيات المعلوماتية من خلال تفعيل آليات التعاون بين علماء اللغة وكافة المتخصصين في مجالات المعلوماتية من أجل تطويع هذه التقنيات للتعامل الكفؤ والفعال مع اللغة العربية.
- ٢ - ضرورة التوصية باستخدام حركات التشكيل العربية في كافة الكتابات، ودراسة ذلك دراسة وافية تعظم الاستفادة منه، وأن يتم اعتماد هذه التوصية من جميع التجمعات المعنية باللغة العربية، والبدء في نشرها والدعوة لها بحيث يتم تطبيقها في المراحل الدراسية

المختلفة ووسائل الإعلام المرئي والمقروء، تمهيداً لتعميم ذلك في  
مختلف الاستخدامات.

