

الحرف العربي والحوسبة

أ.د محمد زكي محمد خضر

الموسم الثقافي لمجمع اللغة العربية

عمان - الأردن

13 ربيع الأول 1423هـ - 5 حزيران 2001م

مقدمة:

اللغة العربية أم اللغات السامية. والآثار المكتشفة عن الكتابة العربية لا تعود إلا لفترات متأخرة من تأريخ اللغة العربية. فاللغة العربية أساساً لغة متوارثة نطقاً قبل أن تتوارث كتابة. وما يعيننا في هذه المحاضرة هو الحرف العربي بشكلية المنطوق والمكتوب وعلاقتها بالحوسبة. وحيث أن حقل الحوسبة الذي يتعامل مع اللغات الطبيعية يدعى بالذكاء الاصطناعي فلا بد من إعطاء مقدمة عن هذا الحقل وتقدمه في العقود الأخيرة وما ينتظره من تطورات في المستقبل.

إن أهمية معالجة اللغة العربية بالحاسوب لم يعد أمر رفاه أو أمراً ثانوياً، بل هو أمر في غاية الأهمية وعليه يعتمد مستقبل اللغة ومكانة العرب في الحضارة الحالية بل ومستقبلهم الإقتصادي والعلمي.

الذكاء الاصطناعي

بعد فترة وجيزة من إختراع الحواسيب، تبين أن بإمكان الحاسوب القيام بعمليات غير الحسابات الرياضية. وأن السرعة التي يقوم بها الحاسب بعملياته يمكن إستعمالها في الكثير من المهام غير الرياضية البحتة. فقد تبين إن هناك إمكانية لإتخاذ القرارات وأن هناك إمكانية لتمييز الأنماط وأن هناك إمكانية للتعرف على الصور. وقد تكونت حقول جديدة تقع تحت مظلة الذكاء الاصطناعي كان من أهمها حقل معالجة اللغات الطبيعية. وقد تطور هذا الحقل ليشمل العديد من

جوانب معالجة اللغات الطبيعية من نواحي الكتابة وتصحيح الإملاء والقواعد النحوية والصرف والمعاني والترجمة بل وحتى البلاغة والشعر.

أما حقل التعرف على الأشكال والصور والنماذج فقد دخل في تطبيقات عديدة مثل التعرف على المواقع والتطبيقات العسكرية والتنقيب عن المعادن والإبحار في المحيطات وسفن الفضاء وغير ذلك. وما التعرف على الكتابة سوى واحد من هذه التطبيقات.

ومن الجوانب الأخرى للذكاء الاصطناعي النظم الخبيرة التي تستخدم في معالجة المرضى وإصلاح الأعطال والتحكم الآلي وغيرها

بعض الوسائل الحديثة في الذكاء الاصطناعي

المنطق الضبابي (Fuzzy Logic)

يعيب الكثيرون على الحاسوب أنه يتعامل مع الأرقام بدقة هائلة بينما الحياة ليست كذلك . فأنت تقول الجو حار وقد يعني ذلك 30 درجة مئوية في عمان أما في دبي فحينما تقول الجو حار فلا يعني ذلك أقل من 40 درجة مئوية. ويعني ذلك أن لفظة حار نسبية وقد تعني هنا ما لا تعنيه هناك. ومثل ذلك كلمات مستعملة بكثرة في اللغة مثل طويل وقصير وثقيل وبعيد وما يشبهها . هذا بالإضافة إلى كلمات أخرى تضاف إلى مثل هذه الكلمات مثل "ما يقرب من " و"جدا " وغير ذلك من الإضافات.

وهكذا وجد حقل جديد للتعامل مع الألفاظ اللغوية العادية بحيث يفهمها الحاسوب ويحسن التعامل معها كما يتعامل معها الإنسان

مثال ذلك : إذا كان هناك مدّة طويلة بين حرفين فعلى الغالب هي حرف سين وسطية

إذا كان هناك حرف سين بثلاث ركزات وفوقه نقطة واحدة فرمما هو حرف شين والثلاث نقط قد تجمعت بنقطة واحدة

هذه القواعد ليست صارمة بحيث يمكن التعامل معه بالطرائق الرياضية التقليدية بل يجب أن تعامل وفق منطق الاحتمالات والمنطق الضبابي

الشبكات العصبية (Neural Networks)

لقد تعلم الإنسان من خلقته التي خلقه الله عليها . فالشبكات العصبية في دماغ الإنسان وفي جسمه تعمل بطريقة غاية في الدقة والتعقيد . فقد وجد أن المعلومات لا تخزن في دماغ الإنسان في موقع معين بل تتوزع على خلايا عصبية عديدة وعندما يحاول المرء تذكر أمر ما فإنه يجمع هذه المعلومات من شتاتها . وهكذا إستطاع الإنسان أن يحاكي ذلك فيستعمل شبكات عصبية إصطناعية داخل برامج الحاسوب تقوم بحل مسائل عديدة خاصة ما يحتاج منها إلى تمرين وتعلم وإستفادة من الماضي .

مثال ذلك إستعمال الشبكات العصبية في قواعد حروف الجر في اللغة العربية(1).

الخوارزميات الجينية (Genetic Algorithms)

تعلم الإنسان مما يحدث في الكون من تطور وتغير دائم وحركة دائمة نحو السمو والكمال في كل شئ والخالق جل شأنه خلق كل شئ فأحسن خلقه . ومحاكاة ذلك أمكن إيجاد وسائل رياضية تحاول الوصول

إلى حلول مسائل متعددة في الحياة العملية ومنها ما يتعلق باللغات الطبيعية وهي حقل جديد بدأ الدخول مؤخرا في معالجة اللغات الطبيعية.

ترميز الحرف العربي على الحاسوب

مر ترميز الحرف العربي على الحاسوب بمراحل عديدة كان آخرها ظهور الرمز العالمي الموحد (unicode) . ولمعرفة تفاصيل ذلك يمكن الرجوع إلى محاضرة المؤلف "الحروف العربية والحاسوب" التي ألقيت في الموسم الثقافي الرابع عشر للمجمع عام 1996 (2)

بدأ مشروع الرمز العالمي الموحد وذلك عام 1988 لمعالجة التداخل بين اللغات المختلفة ومشكلة استعمال الرمز نفسه من قبل أكثر من لغة واحدة . فعند إدخال وثيقة تعمل بلغتين كانت الرموز قبل الرمز العالمي الموحد تؤدي إلى غموض وتداخل بين الرموز . وقد أقر استعمال الرمز العالمي الموحد عام 1991 وقد امتاز الرمز العالمي الموحد بمزايا منها الحسن ومنها السيئ ، فمثلا تم استعمال 16 رمزا ثنائيا أي ضعف عدد الرموز الذي كان مستعملا قبل ذلك وبذلك أصبح الحيز الذي يشغله النص المكتوب بهذا الرمز ضعف ما كان عليه سابقا.

خصصت مواقع للرموز وليس لأشكال الحروف كما أن الرموز كانت للنصوص العادية دون أية إضافة كنوع الخط أو إضافة خط تحت الحرف وغير ذلك وقد استعمل الرمز نفسه للحرف الذي يظهر في أكثر من لغة فحرف الباء بالعربية له الرمز نفسه لحرف الباء بالفارسية أو الأوردية كما أن هذا الترميز يستطيع قبول التشكيل بما يناسب اللغة العربية ولكن رمز الحرف المشكول سيكون مضاعفا أي مكونا من رمز الحرف ورمز حركة التشكيل.

وعلى هذا لم يخلُ هذا الترميز من مساوئ. أهم تلك المساوئ هو الكفاءة فهو كما ذكرنا يحتاج ضعف حجم طرائق الترميز السابقة كما أن استعمال الرمز نفسه للُّغات ذات الأبجديات المتداخلة قد يؤدي إلى التباس بين اللغات فرما كان هناك مجموعة حروف تعطي كلمة ذات معاني مختلفة بين لغتين مثل العربية والفارسية. فقد ترد الكلمة ويراد ترجمتها إلى الإنجليزية وليس هناك ما يشير هل أن الكلمة مراد ترجمتها من العربية أم من الفارسية كما أن مجموعات الحروف المشتركة قد وضعت بشكل لا يخدم كل اللغات بوقت واحد مما يتسبب في صعوبات في وضع تطبيقات لكل لغة على حدة وخاصة في ما يتعلق بالترتيب الأبجدي حيث لا يتوفر هذا الترتيب لكل اللغات المشتركة في الحروف الهجائية كما أن عمليات نقل المعلومات من الترميز القديم إلى هذا الترميز الموحد تعاني من مشاكل عديدة وكذلك من الترميز الموحد إلى الترميز القديم.

يبين الشكلان (1) و (2) جدولين بمواقع الحروف العربية أو حروف اللغات القريبة من العربية في جدول الرمز العالمي الموحد. وهذان الشكلان هما صيغتان معدلتان عام 1996 عما سبق أن صدر عام 1991. (2).

الترميز المتعدد

في ضوء القصور الذي يعاني منه الرمز الموحد برزت هناك حاجة إلى رمز يعتمد على اللغات المختلفة فقد إقترح ما يسمى بالترميز (المتعدد multicode) (3) والذي يحوي على رمز خاص باللغة و8 رموز ثنائية للحروف (أو 16 حرف إن احتاجت اللغة ذلك) وبذلك يمكن شمول جميع لغات العالم عدا بعض اللغات ذات الأبجدية الطويلة كالصينية والكورية واليابانية فيمكن أن تعامل بشكل خاص.

وقد إقترح في هذا الرمز أحد الرموز وهو الرمز وهو الرمز الأخير المرقم 255 (FF بالنظام السداسي عشر) لعملية الانتقال من لغة لأخرى لكي يبين كيف يمكن الانتقال من لغة إلى أخرى.

ورغم كل ذلك فإن مسألة التشكيل في اللغة العربية لم تعالج بشكل مرضي لحد الآن حيث تعامل حركات التشكيل كحروف مستقلة مما يؤدي إلى صعوبات عديدة .

تمييز الحروف العربية

الكتابة هي سجل التاريخ الأنساني ، وقد حفظت لنا وقائع ونصوص آلاف السنين منذ نشأة الكتابة إلى يومنا هذا. وستبقى الكتابة على الورق أو غيره كذلك في المستقبل. ورغم أن هناك من يتصور إن الكتابة الألكترونية على رقائق السليكون في ذاكرات الحواسيب قد تحل محل الورق ، إلا أن ذلك مشكوك فيه على الأقل لحد هذا اليوم.

الكتابة العربية مرت بمراحل عديدة ولم تصل وضعها الحالي إلا بعد تطورات عديدة. فالكتابة العربية قبل الإسلام وفي صدر الإسلام لم تكن منقوطة. (أنظر الوثيقة المبينة في الشكل (3) والتي تبين كتاب النبي صلى الله عليه وسلم إلى المنذر بن ساوي أمير البحرين) لكن التشكيل إضيف للكتابة حتى قبل التنقيط. وقد ظهرت العديد من أنواع الخطوط وتفنن الخطاطون العرب بابتداع خطوط في غاية الروعة والجمال.

وبعد دخول الطباعة إلى العالم العربي ، شاعت خطوط معينة وقل إستعمال الشكل على الأقل في المشرق العربي. وقد أدى ذلك إلى شيوع اللحن في القراءة وضعف التزام القارئ العربي بضبط أواخر الكلمات.

وعند دخول الحواسيب في العقود الأخيرة من القرن المنصرم ، استعملت الطابعات التي تكتب العربية ثم ظهرت الشاشات التي تظهر الكتابة العربية وتحسن جمال هذه الكتابة وأصبح بالأمكان تشكيل النصوص العربية إلا أن التشكيل إعتبر رمزا مستقلا ، شأنه شأن الحروف . وقد إعتبر لبعض الحروف رموز عديدة كالهزمة بحسب وضعها من الكلمة. وعلى هذا فإن استعمال الرموز في الحاسوب إستند إلى خدمة الكتابة كرمز وليس كشكل ولم يكن هذا الإستعمال خدمة أمينة محتوى الكلمة العربية أيضا.

وإذا ما أريد لهذا الوضع أن يصحح ، فإن مقداراً من الذكاء يجب أن يدخل للبرامج العربية التي يتعامل معها المستخدم بحيث تساعده للأقتراب من اللغة العربية السليمة.

تعتمد عملية قراءة النص العربي من قبل الحاسوب على المسح الضوئي للنصوص المطبوعة أو المكتوبة، وهي عملية سهلة ومتيسرة اليوم. لكن الصور التي تنتج نتيجة هذا المسح الضوئي قد تنتج بعض التشويش وما يسمى بالضوضاء التي قد تشوش عملية التمييز في المراحل اللاحقة

خطوات عملية التمييز:

تتكون عملية التمييز من خطوات أولى تلك الخطوات هي عملية المسح الضوئي وذلك بتحويل الصفحة إلى مجموعة من النقط السوداء والبيضاء. ويعتمد عدد هذه النقط على دقة الماسح الضوئي وتتراوح دقة الماسحات حوالي 300 نقطة لكل بوصة أو أكثر من ذلك.

تتضمن الخطوة التالية عمليات هئية تدعى بمرحلة ما قبل التمييز . وهذه المرحلة تتضمن معرفة الموصفات العامة للوثيقة . فهل تحتوي على رسوم وأشكال وهل تحتوي على عدد من الأعمدة أم عمود واحد وهل أصلا كان تصويرها مائلا أم عموديا وهل تحوي جداول .. إلى غير ذلك من الموصفات . وخلال هذه العملية يجري تحديد الأسطر ووضع حدود فارقة بين سطر وآخر وتحديد إرتفاع الكتابة في السطر الواحد وهكذا .

أما الخطوة الأخرى فهي تتعلق بالتعرف على الكلمات والمقاطع والحروف بمختلف أشكالها ومواقعها في الكلمة . ثم استخلاص سمات كل جزء من المقطع وذلك تمهيدا للمرحلة اللاحقة التي تقوم بعملية مطابقة هذه السمات مع ما يعرف من موصفات للحرف أو المقطع الواحد . وهذه أهم وأصعب خطوة في عملية التمييز . وقد أجريت أبحاث معمقة فيما يخص الكتابة المطبوعة . من السمات التي تميز حرفا عن غيره : عدد النقاط ومواقعها وإستقامة الحرف وعرضه وارتفاعه وعلوه عن متوسط السطر والميل وعدد الحلقات المقفلة وشكلها والحجم الكلي ومركز ثقل الحرف ومواقع نقاط الإتصال مع ما قبله أو مع ما بعده وغير ذلك من السمات الأخرى . يبين الشكل (4) مخططا يبين أحد الطرائق المقترحة للتعرف على الحروف العربية بشكل متسلسل وفق بعض موصفاها (4) . أما الكتابة المكتوبة بخط اليد فلا تزال تعاني من إنخفاض في الدقة وكثير أخطاء ولبس بين حرف وآخر نظرا لإن الذكاء الإصطناعي المدخل في هذه البرامج لم يبلغ درجة عالية من النضج والعمق . وتتبع في عملية التمييز وسائل متعددة رياضية ومنطقية وشكلية وقواعد بيانات وغير ذلك . إلا أن ما تجدر الإشارة إليه أنه كلما كانت البرامج أقرب إلى ما يفعله الإنسان في قراءة النصوص كلما كانت الدقة أعلى . ولكن إستنباط ما يقوم به الإنسان ومحاكاته في الآلة ليست عملية سهلة على الإطلاق . وتجدر الإشارة إلى أن بعض مناهج تمييز الكتابة تعتمد تمييز المقطع أو الكلمة أصلا وليس الحرف . ويشبه ذلك ما تدعو إليه بعض المناهج التعليمية للأطفال من تعليم الطفل الكلمات قبل تعريفه على الحروف .

وما يميز اللغة العربية هو تشابه حروف كثيرة مع اختلاف في عدد النقط ومواقعها . ورغم أن هذه الميزة ميزة مفيدة ، إلا أنها قد تكون عائقا في بعض الأحيان وذلك عند عدم قدرة الآلة على تمييز النقطة الضعيفة أصلا أو الخلط بين النقطة والنقطتين والثلاث أو الظن بوجود نقطة نتيجة عدم دقة الماسح الضوئي .. إلى غير ذلك من صعوبات .

ويفيد في هذه الحقل وجود قاموس بالكلمات المتداولة وإعطاء احتمالات لماهية الكلمة المعنية إن كان لها معنى بوجود النقطة أو عدمه وكذلك التفريق بين أن يكون فيها نقطة واحدة أو أكثر . كما أن سياق الجملة قد يفيد في ذلك . لذلك فغن عملية تمييز الكتابة لا تنفصل عن معالجة الجملة العربية ككل ولا عن معالجة الكلمة والمقطع .

المشكل الآلي المقترح

فيما يأتي مشكل آلي مقترح يساعد الكاتب على تشكيل النص الذي يدخل إليه بأقل عناء ممكن . ولنأخذ مثلا على ذلك :

إذا قام المرء بإدخال الآية التالية من القرآن الكريم:

يا : إن الآلة بإمكانها أن تدرك بسهولة أن الياء مفتوحة فلا تقبل العربية أية حركة قبل الألف اللينة غير الفتح كما أنه ليس هناك حاجة لتشكيل الألف وعلى ذلك فيمكن للكاتب أن يكتب يا وتظهر الكلمة يا

أيها : هذه الكلمة كذلك ليس فيها غير احتمال واحد وهو الفتح على الهمزة والشدة المضمومة على الياء والفتحة على الهاء فتصبح أيها وهذه الكلمة من الكلمات التي يجب أن يحتفظ الحاسوب بتشكيلها شأنها شأن الكلمة الأولى يا

الذين : هذه كلمة أخرى ليس فيها سوى نوع واحد من التشكيل : الَّذِينَ

آمنوا : إن هذه الكلمة تحتمل أن تكون آمنوا أو آمنوا فالميم يمكن أن تكون مفتوحة أو مكسورة ولكن سياق الجملة يوحي بأن الكلمة يجب أن تكون فعلا ماضيا وليس فعل أمر فإذا ما احتوى البرنامج على إعراب آلي فإن الكلمة سرعان ما يتبين أن هناك احتمالا واحدا لتشكيلها لا غير. وعلى فرض أن البرنامج لم يصل بعد إلى ذلك المستوى من الذكاء الأصطناعي فإن البرنامج يمكن أن يظهر الاحتمالين معا (آمنوا / آمنوا) يختار الكاتب أحدهما بالضغط على الشكل الصحيح

كتب : هذه الكلمة تحتمل (كَتَبَ / كَتَبَ / كُتِبَ / كُتِبَ / كُتِبَا) وعلى أقل تقدير يمكن أن تظهر أمام الكاتب ليختار أحدها رغم أن السياق يمكن ان يحذف بعض هذه الاحتمالات.

عليكم : تشكيل هذه الكلمة يكاد يكون مباشرا . كما أنه قد يشير فورا إلى أن كلمة كتب التي قبلها مبنية للمجهول وبذلك تصحح الكلمة التي سبقتها مباشرة

وهكذا..

من المثال أعلاه يتبين إن تدريب الآلة على التشكيل الآني ليس مهمة صعبة ولكن تحتاج إلى بحث علمي وقليل من الدعم لنشر مثل هذا البرنامج كي يكون أداة على تصحيح اللفظ الصحيح. فالشكل يساعد القارئ على اللفظ الصحيح آنيا وعلى إمتلاك ملكة للفظ المشكول حتى وإن لم يكن الشكل ظاهرا على الكلمات.

تمييز الحروف المطبوعة:

تعتبر عملية تمييز الكتابة المطبوعة أسهل وأقل تعقيدا من عملية تمييز الكتابة المكتوبة بخط القلم. فالكتابة المطبوعة اليوم ما هي إلا صور متشابهة تماما للحرف نفسه حسب موقعه من الكلمة. فحرف الجيم التي في هذه الصفحة لها الشكل نفسه

تماما إن وقعت في أول الكلمة أو حينما يسبقها حرف غير متصل مثل الواو أو الراء. والجيم نفسها تكون متطابقة مع أية جيم أخرى في الصفحة إن وقعت في وسط الكلمة واتصلت من جهتها. وهكذا. وتتم عملية تمييز الحروف المطبوعة عادة بمطابقة شكل الحرف مع مخزون شكله في الحاسوب ومن ثم تستطيع الآلة التعرف على الحرف بمقارنة شكله مع الشكل المخزون فيها. ويمكن أن تقوم الآلة بخزن أنواع مختلفة من أشكال الحروف التي تدعى Fonts. وقد ظهرت أنواع مختلفة من أشكال الحروف الطباعية وتفنن الخطاطون العرب في رسم خطوط جميلة. ويتوفر على الأنترنت اليوم عدد من المواقع المتخصصة بالخطوط العربية (أنظر الملحق - 1) ويتوفر في الأسواق أقراص مدججة للكثير من أشكال الخطوط العربية الجميلة التي يمكن نصبها على أي حاسوب اليوم والإفادة منها في الكتابة. وعلى هذا فإن برامج تمييز الحروف والتي تدعى أحيانا بالقارئ الآلي تتعرف على الكتابة المطبوعة بالطابعات الحاسوبية الحديثة وقلما تستطيع التعرف على الطباعة القديمة. يبين الشكل (4) كيف تتم عملية تقطيع الكلمات المطبوعة إلى حروف متتالية (5).

تمييز الحروف المكتوبة بخط اليد:

إن عملية تمييز الحروف المكتوبة بخط اليد أكثر صعوبة من تمييز الحروف المطبوعة نظرا لأن الكاتب نفسه لا يعيد كتابة الحرف بالشكل نفسه تماما عند إعادة الحرف نفسه في موقع آخر من الكتابة. كذلك فإن الأشخاص المختلفين لا تتطابق كتاباتهم نهائيا مهما حاولوا ذلك. لذلك فإن عملية تدريب الحاسوب على التعرف على الكتابة اليدوية لا يزال في مراحل التطوير والبحث ولم يصل مراحل التطبيق الفعلي حتى في اللغات الأوربية لحد الآن.

إن عملية تمييز الكتابة المكونة من حروف مقطعة كالإنكليزية مثلا أسهل من تمييز الكتابة المتصلة. لذلك فإن تمييز الكتابة العربية أصعب من تمييز الكتابة الإنكليزية نظرا لأن اللغة العربية لا تكتب إلا متصلة. وقد أجري العديد من الأبحاث العلمية على تمييز الكتابة العربية في جامعات عربية وأجنبية ومراكز أبحاث عديدة. ولا تزال الدراسات جارية في هذا الصدد. إن إحدى الصعوبات المهمة في تمييز الكتابة العربية بالإضافة إلى مسألة فصل الحروف عن بعضها مسألة كتابة بعض الحروف متداخلة عموديا أي بعضها أعلى بعض وقد يؤدي ذلك إلى تداخل الحروف والكلمات مع بعضها. يبين الشكل (5) كيف تتم عملية تقطيع الكلمات المكتوبة بخط اليد إلى حروف متتالية (5).

تمييز الكتابة آليا:

عندما يتم ملاحظة يد الكاتب لحظة بلحظة، يمكن تمييز الحروف بسهولة أكثر من ملاحظة صفحة مكتوبة مسبقا. فالحاسوب يقوم بالتعرف على الحروف أولا بأول وهذه المعلومات التي تنقل للحاسوب هي معلومات مفيدة تسهل

عملية التمييز. لكن الكتابة يجب أن تتم بقلم خاص وعلى لوحة خاصة متصلين بأسلاك. وهذا مجرد ذاتة عائق كبير جدا وسوف لن يكون بديلا لبرامج كفاءة للتعرف على الكتابة بخط اليد.

هناك حاجة للتعرف على الخطوط عدا الغرض العادي وهو تمييز الحروف المكتوبة بخط اليد وذلك لنقلها الى خط مطبوع ومن ثم قابليتها للمعالجات الآلية المختلفة كالتشكيل والتصحيح والأعراب والترجمة وفهم النص واتخاذ إجراء معين نتيجة النص وغير ذلك من الأغراض. ومثل هذه الحاجات يمكن أن تكون توثيق توقيع شخص ما أو التأكد من أن الخط يعود لشخص ما أو حدوث تحريف ما في مخطوطة معينة.. إلى غير ذلك من الأغراض المختلفة.

إن هناك تطبيقات عديدة لتمييز الكتابة منها عملية الفرز الآلي لرسائل البريد وذلك بالتعرف على العناوين وتوجيه الرسائل إلى الوجهة التي قصدها مرسل الرسالة. ومنها قراءة الصكوك المصرفية وصرفها آليا بعد التأكد من صحة التوقيع وقراءة المبلغ الذي تحويه.

كما تجدر الإشارة إلى أن تمييز الكتابة لا تستدعي تفكيك الكتابة إلى حروف بل هناك محاولات للتعرف على الكلمات جملة واحدة بدل تفكيكها إلى مكوناتها حرفا حرفا . ولا شك بأن هذه المهمة أصعب من سابقتها.

الصوتيات والحاسوب

ينتج الصوت نتيجة حدوث تموجات في الهواء ولا ينقل إلا خلال وسط مادي سواء كان هواء أو مادة صلبة. والأصوات الكلامية تنتج في جهاز الصوت في الإنسان من خلال الهواء الخارج من الرئتين . فالرئتان تعملان كمنفاخ للهواء الذي عند مروره عبر الحنجرة والحلق والأنف والفم يحدث الصوت. ويتعرض الهواء لتضييق وتعديل في مساره فنتج أصوات الحروف المختلفة. والصوت هو الوسيلة الطبيعية التي تنقل من خلالها اللغات الطبيعية. وتتم عملية الكلام بعدد من المراحل يمكن تلخيصها بما يأتي (6):

- 1 – مرحلة الترميز الدلالي بأن يصوغ المتكلم ما يريد قوله بلغته
- 2 – مرحلة الترميز القواعدي وذلك بأن يحدد المتكلم ما سيقوله وفق قواعد لغته
- 3 – مرحلة الترميز الصوتي (الفونولوجي) وذلك بأن يصوغ الكلام بسلسلة من المورفيمات وفق قواعد لغته
- 4 – يرسل الدماغ تعليماته إلى أعضاء النطق مبينا لها الحركات اللازمة لكل مورفيم
- 5 – قيام أعضاء النطق بالحركات المنتجة للأصوات

6 - تنتقل الأصوات في الهواء

7 - تدخل الذبذبات أذن السامع الداخلية والوسطى

8 - تنتقل الذبذبات من الأذن إلى الدماغ من خلال حوالي 30 ألف خيط عصبي

9 - يقوم الدماغ بفك رموز الصوت إلى فونولوجيا

10 - يقوم الدماغ بفك رموز الأصوات وفق اللغة التي يعرفها

11 - يقوم الدماغ بفك قواعد اللغة إلى معاني (ومفاهيم) الدلالة

يبين الشكل (7) إنتقال الرسائل الشفوية من المتكلم إلى السامع وبالعكس. كما يبين الشكل (8) الجهاز النطقي لدى الإنسان أنواع الفونيمات

فونيمات أولية أو أساسية وتنقسم الى نوعين

(أ) الصوامت

ومنها تتكون جزئيات الكلام . والكلام الإنساني هو حقيقة سلسلة صوتية متصل بعضها ببعض اتصالا وثيقا ولا وجود للأصوات المنفردة أو المنعزلة خارج السياق. فالجمل ليست تتابعات صوتية منفصلة ولا توجد فجوات بين الأصوات التي تتكون منها الكلمات كما لا توجد وقفات بين الكلمات المنطوقة.

وتقسيم الأصوات الى صوامت وحركات هو الشائع لكنه ليس حاسما فهناك أصوات كلامية ذات حالة وسطية بين الصامت والحركة مثل الهاء المجهورة ، ويرجع ذلك أساسا الى صعوبة رسم الحدود بين الصوامت والحركات وذلك لإختلاف المعايير المستخدمة في تعريف كل منها.

تتفاوت اللغات في عد الصوامت والحركات فيها وتبلغ أعلى نسبة من الحركات للصوامت في اللغة الفنلندية حيث تصل 40% أي 8 حركات من أصل 21 فونيم

يقسم علماء الأصوات الصوامت حسب طريقة النطق الى سبعة أصناف هي الوقفات والمركبات والجانبيات والمكررات والأنفيات والأحتكاكيات ثم أشباه الحركات. وتقابل الوقفات ما كان يعرف عند الأقدمين من علماء العربية بالحروف الشديدة وتمتاز اللغة العربية بأن الأصوات الوقفية فيها تتوزع على خمسة مخارج ونادرا ما يلاحظ ذلك في لغات أخرى.

والوقفات العربية هي الهمزة وتعرف بالوقفة الحنجرية والقاف والكاف أما الوقفات الأسنانية – اللثوية فتشمل الضاد والطاء والذال والتاء . أما حرف الباء فهو وقفي شفهي مجهور مرفق.

أما المركبات ففي العربية صوت واحد هو الجيم . أما الأصوات الجانبية فتشمل اللام المرفقة والمفخمة

أما المكررات فتشمل الراء المكررة والراء غير المكررة . والأنفيمات تشمل الميم والنون والأحتكاكيات تشمل الهاء والحاء والعين والحاء والشين والغين والسين والزاي والصاد . اما التاء فتنتطق بوضع طرف اللسان بين أطراف الثنايا ونحو ذلك تنطق الذال والطاء والفاء بمثلث بين الأسنان. وفي العربية حرفان هما الياء والواو وهما أشباه حركات . يبين الشكل (9) الأجدية العالمية للأصوات كما يبين الشكل (10) رموز الأصوات العربية ويبين الشكل (11) الأصوات العربية الصامتة بحسب نطق المتخصصين وقراء القرآن الكريم.

(ب) الحركات

إن وصف الحركات أصعب بكثير من وصف الصوامت . وقد توصل المختصون الى تقسيم الحركات إلى ثمان حركات معيارية أولية وهي تمثل كما في الشكل (12) وتقع الحركات العربية الفتحة والكسرة والضمة ومواقع الحركات العربية منها وفي هذا الشكل تمثل الخطوط العرضية (الأفقية) مستويات ارتفاع النقطة العليا للسان في الفم . أما الخطوط الطولية فتمثل جزء اللسان الذي تقع فيه النقطة العليا ونظرا لضيق الفراغ المحصور بين النقطة العليا للسان وبين الحنك سميت الحركات التي تقع على هذا الخط ضيقة أما الخط الأفقي السفلي يكون اللسان مسجى في قاع الفم فتكون سعة الفراغ المحصور بين جزء اللسان والحنك فتسمى الحركات متسعة. أما الخط الطولي الأيمن فيمثل الجزء الخلفي للسان فتسمى الحركات خلفية . أما الخط الأيسر المائل فيمثل الجزء الأمامي من اللسان فتسمى الحركات أمامية.

فونيمات ثانوية

وهي عبارة عن ملامح صوتية إضافية تؤثر في الأصوات الكلامية ومن هذه الملامح: النبر ووظيفتها التعبير عن التوكيد أو الأنفعال وهناك وظيفة أخرى للنبر في بعض اللغات للنبر داخل الكلمة. ويقع في اللغة العربية في مثل " الطلاب يكتبون دروسهم " ويقع النبر في " بو " من يكتبون أو في الكاف في كلمة " كتب " وهكذا يمكن استنتاج أن النبر في الكلمة العربية يعتمد على عدد المقاطع في الكلمة النغمة والتنغيم وهي عبارة عن إضافة اللحن للكلام . ويستعمل التنغيم للتمييز بين صيغة جملة وأخرى مثل :

ما أحسنَ زيدًا!

ما أحسنُ زيدٍ؟

ما أحسن زيدُ.

(جـ) الطول أو الكمية أو المدة : وفي اللغة العربية نميز درجات كثيرة من الطول ربما خمسا أو سنا مثل الحركات القصيرة (مثل كَتَبَ) والطويلة (مثل كاتب) والطويلة جدا (مثل صحراء وهدوء وبرئ) والمديدة (مثل ضَّال)

(د) المفصل : وهو سكتة خفيفة بين كلمات أو بين مقاطع كلمة معينة مثل : هذا / كتاب / زيد أو مثل السكتة بين ها / إذا أو كـ / تاب

تحليل الأصوات العربية

تحتوي الأصوات بصورة عامة على تردد أساس يختلف وفق المتكلم. وتنحصر قيمة هذا التردد للرجال بين 80-200 دورة بالثانية وللنساء بين 150-450 دورة بالثانية بينما هي بين 200-600 دورة بالثانية للأطفال.

تقسم الدراسات الحديثة الأصوات إلى قطع صوتية مكونة من أنواع مختلفة ولكل منها طول مختلف يتراوح بين حدود معينة.

ويتم الرجوع إلى قاموس يحتوي على ترجمة معاني القطع اللفظية. إلا أن المشكلة الرئيسة تكمن في المراحل الانتقالية بين القطع الصوتية المتتالية. يبين الشكل (14) لفظة كلمة " كَتَبَ " وكيفيات تقسيمها إلى فونيمات أو دايفونات مختلفة خاصة للصوامت.

وقد أجريت العديد من الدراسات حول الصوامت المستعملة في اللغات الأوروبية. أما الصوامت الخاصة باللغة العربية كحروف الضاد والطاء والقاف فقد لقيت هي الأخرى اهتماما خاصا من قبل الباحثين مؤخرًا.

أما تمييز الأصوات فيحتاج إلى تحليل للإشارة الصوتية ومعالجة كل حيز من الترددات على حدة لمعرفة الخواص الدقيقة لكل مكون من مكونات الصوت وتعتبر عملية تقسيم الصوت الى قطع متتالية من أصعب هذه المهام نظرا لاختلاف مواصفات الصوت بين متكلم وآخر بل واختلاف نطق المتكلم نفسه وسرعة كلامه ودقة نطقه وكيفية ربط الكلمات بعضها البعض.

وتجري عملية التمييز بعد عملية التقطيع وهي لا تقل صعوبة عن سابقتها للأسباب نفسها. إن عملية تمييز الأصوات لا تقل تعقيدا عن تمييز الحروف المكتوبة بخط اليد نظرا لأن كليهما مختلف بين شخص وآخر بل وللشخص نفسه بين حرف وحرف آخر.

إن إحدى التطبيقات المهمة المؤملة من تمييز الأصوات العربية هو تطوير وسيلة إملاء آلية بحيث يتم تحليل الصوت وتحويله إلى كتابة صحيحة. وبالطبع فإن تخلص الصوت المنطوق من شوائبه وإضافة ما لم ينطق بشكل كامل أو صحيح إلى الكتابة هو أمر في غاية الأهمية.

إن المصاحف اليوم بما تحويه من علامات وقف وترتيل هي أمثلة على ما تحويه اللغة من مكونات لفظية غير ظاهرة للعيان عند تدوينها بالحروف فقط. فهناك الحركات الخاصة بالشكل وهناك العلامات الأخرى اللازمة لضمان دقة اللفظ.

إن النظرة السريعة للحركات في اللغة العربية توحى بوجود 3 حركات هي الفتحة والضمة والكسرة. إلا أن الحقيقة أن هناك عددا أكبر من ذلك نتيجة تعاقب الحركات مع حروف العلة (الألف والواو والياء) قبلها وبعدها وكذلك وجود الشدة على حرف العلة . فالكسرة قبل الواو المفتوحة (مثل سوى) هي غير الكسرة قبل الياء المفتوحة (مثل عليا)

وتجدر الإشارة كذلك إلى أن بعض الحروف العربية لها أكثر من فونيم كاللام المرققة والمفخمة والراء المرققة والمفخمة كما أن الياء والواو هما حرفات صامتتان وهما حركات في الوقت نفسه . أما همزة فهي حرف صامت بينما الألف هو حركة فقط.

كما أن بعض الحروف تنقلب إلى حروف أخرى وفق موقعها في الكلمة وحركتها مثل أسفراً (سين) وأسدلّ (زاي) ويسطّع (صاد) .

هل نحن بحاجة إلى دراسة الفصحى والعامية أيضا؟

نعم ! إن الحاجة إلى دراسة تحليل اللغة العربية الفصحى أمر في غاية الأهمية نظرا للحاجة إلى ذلك في فهم الكلام العربي المنطوق وتحويله إلى نص مكتوب أو ترجمته إلى لغة أخرى سواء كان ذلك ترجمة فورية أم غير فورية.

أما فهم وتحليل اللهجات العامية فهو من الناحية العملية لا يقل أهمية عن فهم الفصحى وذلك لأن الحاسوب يمكن أن يكون وسيلة لتعليم الفصحى، وإن عملية فهم العامية واستنباط ما يقابلها من فصحى هو أمر في غاية الأهمية. فالمتكلمون بالعامية اليوم هو أكثر بكثير ممن يتقن اللغة الفصحى. لذلك فيمكن أن يكون الحاسوب وسيلة فعالة للقضاء على العامية من خلال التصحيح الفوري للحن بالفصحى أو استبدال العامية بالفصحى أو تصحيح الأخطاء اللفظية في تشكيل أو آخر الكلمات أو غير ذلك من المعالجات اللفظية للغة العربية. وكل هذا يحتاج إلى دراسات وأبحاث للغة العربية وكيفية برمجة أصواتها. يلاحظ أن هناك أصوات في اللهجات العامية غير مستعملة في الفصحى مثل " بيت " العامية غير " بيت "

الفصحى. أو "لوح" غير "لوح" الفصحى. أو "بيض" غير "بيض" الفصحى وهي غير الكسرة في كلمة "بيض" جمع أبيض . وهكذا

تركيب الأصوات العربية

كما سبق وأن تبين أن الأصوات العربية يمكن أن تفهم من الحاسوب فإن الحاسوب يمكن أن يكون أصواتا عربية صحيحة. وبهذا الصدد فإن تكوين مثل هذه الأصوات في غابة الأهمية نظرا لأنه وسيلة لتعليم اللغة العربية الصحيحة نطقا وتشكيلا وتداولها وهو وسيلة للعودة إلى الفصحى بدل العامية.

وتحدر الإشارة إلى أن بعض الباحثين يشيرون إلى أن مخارج الحروف الشائعة اليوم في بعض البلدان العربية مثل الضاد والطاء والجيم وغيرها بعيدة عن النطق العربي الصحيح كما وصفه علماء اللغة العربية في القرنين الثاني والثالث الهجري. وقد يكون الحاسوب وسيلة هامة في تصحيح هذا النطق وإعادةه إلى سالف عهده بل وإلى توحيد النطق الصحيح بين العرب جميعا إذا ما أحسن استخدام ذلك.

أهمية الدراسات الإحصائية للغة العربية

لا شك بأن هناك مهام كثيرة يجب أن يقوم بها الباحثون العرب في حقل اللغة العربية نحوا وصرفا ودلالة وبلاغة ونطقا من قبل المتخصصين باللغة العربية وفي حقل برمجة اللغة العربية من قبل أخصائيي الحاسوب ومن قبل علماء اللسانيات فيما يتعلق بعلاقة اللغة العربية مع اللغات الأخرى ترجمة ومفاهيم. إن كثرة هذه المهام تفرض عليهم جميعا أن ينتقوا المهام الأكثر أهمية ومن ثم ينتقلوا إلى ما هو أقل أهمية بمضي الوقت. ونظرا لسعة اللغة العربية وتعقد نحوها وصرفها وبلاغتها ونطقها، فإن اختيار ما هو أكثر شيوعا للبدء به ثم الانتقال إلى ما هو أقل شيوعا يجعل المهام أيسر وأقرب إلى التطبيق العملي.

إن مما لا شك فيه أن اللغة العربية المتداولة في الحياة اليومية من مدرسة وسوق وصحف ومحطات فضائية لا تشكل سوى نسبة قليلة من مجمل الألفاظ والمعاني الدلالية العربية المدونة في أمهات الكتب الخاصة باللغة العربية. لذلك فإن نقطة البداية يجب أن تكون من المتداول وليس من المعاجم العربية. وهذا لا يتأتى إلا من خلال دراسات إحصائية لما هو متداول اليوم تتطور شيئا فشيئا لترداد دائرتها ولتقوم بتصحيح الخطأ ثم تحسين المتداول .

يبين الجدول (1) إحصائية بسيطة على عدد الحروف العربية التي قمنا بها على أحد الكتب والتي تبين عدد الحروف المكونة للمقاطع (من ناحية الرسم) وهي تساعد على فهم تكون الكلمة العربية رسماً ومن ثم تعطي فكرة عما يجب على الباحث في حقل تمييز الكتابة العربية فهمه ومن ثم تمييز الكلمة العربية حرفاً حرفاً أو مقطعاً مقطعاً.

ترابط الدراسات مع بعضها

إن الدراسات الخاصة باللغة العربية والحوسبة مترابطة مع بعضها البعض وكل منها يعتمد على الآخر.

فالمخزون اللغوي المستعمل في حقل ما يمكن أن يستعمل في حقل آخر. ودراسة الدلالة من السياق يمكن أن يصحح كثيراً من الأخطاء عند تمييز الكتابة العربية أو فهم كلام منطوق بسرعة .. وهكذا. لذلك فإن نشر الأبحاث وعقد المؤتمرات وتعاون الباحثين ونشر المطبوعات وتعارف الباحثين مع بعضهم البعض ، كل ذلك مما يثري المعرفة في هذا المجال ويزيد من سرعة الوصول إلى أهداف خدمة اللغة العربية العاجلة منها والآجلة.

خطر ترك الأبحاث عن اللغة العربية بيد غير العرب

أصبحت كثير من المتعلقة باللغة العربية والحاسوب ذات طابع تجاري. وقد أدركت ذلك الشركات الكبرى ومراكز الأبحاث الغربية ، لذلك فهي تقوم بأبحاث ثرية متعلقة باللغة العربية وتنتج برامج تباع لمن يتعامل مع اللغة العربية على الحاسوب أو عبر الأنترنت. وهذه تنشر ثقافات وتثبت قواعد وأسس لها ما بعدها وما يستند إليها وتوجه التعامل مع الحاسوب وجهات ليست في صالح العربية ولكن هي في أحسن أحوالها وجهات ذات أهداف تجارية. وقد تشيع لهجات عامية أو إنحراف عن القواعد النحوية أو انحراف في معاني ألفاظ عربية عن معناها المعتمد في اللغة.

لذلك فإن خدمة اللغة العربية عبر التقنيات الحديثة لا تتم إلا بأيدي أبنائها من خلال مراكز أبحاث رصينة أو جامعات عربية تفهم اللغة وتدافع عنها وتحمل همومها.

المصادر:

- 1- " Zeki Khedher, " Use of Neutral Network in Arabic text Transliteration Mohammed Information Conference on Recent Trends in Computer Science Applications and Systems, Philadelphia Univesrsity, Amman, Jordan, July 14,199-13
- 2- الحروف العربية والحاسوب : أ.د. محمد زكي خضر ، المحاضرة السادسة ، الموسم الثقافي الرابع عشر ، 22 حزيران 1996.
- 3- Mudawwar, Multicode : A Truly Multilingual Approach to Text Encoding, Mohammed .Computer, Vol. 30, No. 4, April 1997 IEEE
- 4- of Hexagonally Sampled Printed Arabic Fakhri Khellah abd Sabri Mahmood, Recognition .and Engineering, Vol. 19, No. 4A, Oct. 1994 Characters, The Arabian Journal of Science
- 5- Amin, Arabic Character Recognition, Handbook of Character Recognition and Adnan Document Image Analysis, pp 397-420, Word Scientific Publishing Co. 1997
- 6- محاضرات في اللسانيات : الدكتور فوزي حسن الشايب ، وزارة الثقافة الأردنية 1999
- 7- الأصوات العربية : دكتور كمال محمد بشر - مكتبة الشباب 1987.
- 8- International Conference CATAEE-99, Oct 19-20, 1999 Amman, Jordan, pp 190rd3