١٠٤ تمييز الكتابة العربية آليا

تعد عملية تمييز الكتابة آليا (القراءة الآلية) إحدى العمليات الرئيسية لإدخال النصوص اللغوية، والتي يتزايد استخدامها بمعدل مطرد لتوفيرها الهائل في الجهد اليدوي، وإسراعها من معدل تغذية البيانات، علاوة على إلغاء احتمال الخطأ البشري. بتوفير الأساليب الآلية لتمييز النصوص المكتوبة تكتمل الحلقة الأوتوماتية لمعالجة الكتابة «دخلا» وإخراجا».

يمكن تصنيف نظم تمييز الكتابة آليا إلى عدة أنواع رئيسية هي:

- نظم تمييز الكتابة المنسوخة بنمط «أو فونط» كتابة واحد
- نظم تمييز الكتابة المطبوعة مع تعدد أنماط أشكال الحروف
- نظم الكتابة اليدوية، وتنقسم بدورها إلى نظم التمييز اللاحق (أي ما بعد تمام الكتابة)، ونظم التمييز الفوري (أي المصاحب لعملية الكتابة اليدوية والتي يتم إدخالها باستخدام قلم ضوئي يتحرك على لوح حساس خاص).

تواجه عملية التمييز الآلي لكتابة «العربية» عدة مصاعب أهمها:

(أ) التشبيك، الذي هو بطبيعته غير مستغرق لكل الحروف في كل الأحوال، مما يؤدي إلى وجود فواصل داخل الكلمة الواحدة (مثال: «مزروع»).

(ب) تقارب بين أشكال الحروف، مثل حرفي «الفاء» و«الغين» في وسط الكلمة.

(ج) يضاف إلى ما سبق، بالنسبة للخط اليدوي، الفوارق الفردية الشاسعة في أسلوب الكتابة، وميل بعض الأفراد إلى إضافة الذيول و«الشطحات» ( ) التي لا

دلالة لها، وتركيب الحروف (الحَرَم)، وإزاحة النقاط وعلامات التشكيل عن أماكنها الصحيحة بالنسبة للحروف الحاملة لها. والإسراف في اختزال شكل الحروف (حسن)، والترخص في قواعد التشبيك (الحسن)، إلى غير ذلك.

في مقابل ذلك، هناك بعض خصائص للكتابة العربية تعد مؤازرة لعملية التمييز الآلي، أهمها في رأي المؤلف :

(أ) التنقيط : كميته وموضعه حيث تحصر احتمالات التمييز في نطاق أضيق من الحروف التي تقسم وفقا لخصائصها تنقيطها.

(ب) تميز أشكال تيجان الحروف العربية، مما يوفر سمات فارقة يسهل التعرف عليها آليا

(ج) علاقات التشبيك التي تمد نظام التمييز الآلي بقرائن مفيدة لحصر نطاق افتراضاته في نطاق الحروف التي تتسق مع علاقات التشكيل رهن التمييز.

يمكن تلخيص عملية التمييز الآلي للكتابة العربية في عدة مراحل هي:

المرحلة الأولى : تمييز السطور كل على حدة، وذلك بعملية مسح رأسي تحدد الفواصل الرأسية ما بين السطور، وتفرق بين هذه الفواصل، وتلك التي تفصل ما بين الحروف، ونقاطها، وعلامات تشكيلها.

المرحلة الثانية : (وهي تمثل أحد الفروق الأساسية بين نظم تمييز الكتابة العربية والكتابة اللاتينية) فك التشبيك لتحليل سلاسل الكتابة المتصلة إلى حروف منفصلة. ويتم ذلك بعملية مسح أفقي لتمييز خانات الفراغ الفاصلة بين الكلمات، وتلك التي ترد داخل الكلمات نفسها(كما في كلمة «مزروع» مثلا)، ثم تمييز مقاطع التشبيك الأفقية التي تصل بين الحروف المتتالية. تتم عملية التمييز المذكورة باستخلاص الملامح الرئيسية الثابتة (أو المستقرة في حالة الخط اليدوي) للحرف الجاري تمييزه، مما يساعد على حصر نطاق التمييز في حيز أضيق. تتضمن هذه الملامح الرئيسية عدة خصائص شكلية، مثل عدد النقاط ومواضعها، وجود حلقات مغلقة في شكل الحرف (كحرف الطاء، والفاف مثلا) أو عدم وجود أي منها، طول الحرف وعرضه، وصعوده أو هبوطه عن خط التشبيك الأفقي المتصل. وبالنسبة للكتابة اليدوية هناك أيضا خاصية رفع القلم أثناء كتابة الحرف أو عدم رفعه (قارن، مثلا، بين طريقة كتابة حرف «هـ» و حرف «لا»). بناء على هذه الملامح

يتم تحديد المجموعة الأضيق التي يندرج تحتها الحرف الجاري تمييزه تمهيدا لتحديده بصورة قاطعة في المرحلة التالية .

المرحلة الثالثة : تحديد نمط الحرف، وذلك باستخدام عدة طرق، من أهمها:

(أ) تقسيم الإطار المستطيل للحرف، بعد تكبيره داخل ذاكرة الحاسوب، إلى خانات (خلايا) مربعة أو مستطيلة، ومقارنة أجزاء الشكل المحصورة في الخلايا الصغيرة بعدد محدود من النماذج النمطية التقريبية لتحديد أقربها إلى شكل الجزء المحصور بكل خلية. يحدد نمط الحرف الجاري تمييزه بدلالة مصفوفة الأشكال النمطية التقريبية المناظرة لخلايا تقسيمه.

(ب) تتبع مسار شكل الحرف، وذلك بتقطيعه إلى أجزاء طولية صغيرة يتم تقريبها إلى أقرب نمط من الخطوط المستقيمة، الأفقية، أو الرأسية، أو المائلة بزاوية ٤٥° في الاتجاهات الأربعة (٦٩). يتم تحديد نمط الحرف بدلالة مصفوفة الكميات المتجهة، والتي تمثل سلسلة الخطوط المستقيمة النمطية التي تم تقريب شكل مسار الحرف لها.

المرحلة الرابعة : مقارنة النمط الذي تم تحديده في الخطوة السابقة لشكل الحرف الجاري تمييزه مع أنماط الحروف المعيارية المخزنة في ذاكرة النظام الآلي. بالنسبة للكتابة المطبوعة، خاصة المنسوخة بالآلة الكاتبة، تتم عملية المقارنة بدقة عالية. أما في حالة الخط اليدوي فتتسم عملية المقارنة بالمرونة، وعدم التحديد القاطع، حيث يتحسس النظام الآلي طريقه في تمييز شكل الحرف على ضوء عدة بدائل محتملة، وفي ظل تداخلات كثيرة من حروف أخرى مشابهة، ترتكز عملية المقارنة على قياس الفرق بين نمط الحرف الجاري تمييزه والأنماط المخزنة التي تنتمي له، وذلك باستخدام معادلات رياضية وأساليب إحصائية تساعد على التقريب والتوفيق.

المرحلة الخامسة : إذا تمت عملية المقارنة بنجاح، اختتمت عملية تمييز الحرف، ليتنقل نظام التمييز الآلي إلى الحرف الذي يليه. أما إذا فشلت عملية التمييز، فيلزم افتراض عدة بدائل له باستخدام بيانات إحصائية وقرائن لغوية وسياقية لتخمين الحرف المبهم، أو حصر نطاق الافتراضات في أضيق الحدود في انتظار مزيد من القرائن التي قد يمدنا بها السياق القادم.

إن تمييز الكتابة، خصوصا بالنسبة للخط اليدوي، لا يمكن أن يعتمد فقط على المعطيات

التي تستخلص من الشكل الفيزيائي للحرف (٩١)، بل يجب الاستعانة بإحصائيات عن تتالي الحروف والمقاطع الصوتية لبنية الكلمة، ومعدلات انحياز علامات التشكيل للحروف، ومواضعها داخل الكلمات والجمل. يعد التحليل الصرفي الآلي للكلمات العربية إحدى الوسائل الأساسية للحكم على صحة الكلمات بعد تمييزها، أو تقييم مدى وجاهة افتراضات الحروف البديلة في حالة فشل عملية التمييز على أساس الخصائص الشكلية فقط ،حيث تحدد الصيغ الصرفية مواقع حروف الزيادة وحروف الجذر الأصلية بصورة يمكن الاستفادة منها في تضييق نطاق الاحتمالات.

في بعض جوانبها، تعد عملية تمييز الكتابة اليدوية بطريقة فورية، (أي مصاحبة لعملية الإدخال) أسهل من تمييزها اللاحق بعد إتمام إدخالها، إذ في الحالة الأولى يمكن للنظام الآلي متابعة حركة يد الكاتب، واتجاه خط الكتابة وتسلسل اطراده، والتعرف بشكل قاطع على مواضع رفع القلم وتنزيله، وجميعها قرائن مفيدة تضاف إلى تلك التي سبق ذكرها.

يعد تمييز الكتابة العربية المطبوعة بالجمع التقليدي أو الضوئي (الإلكتروني) أصعب بكثير من تلك الخاصة بالكتابة العربية المنسوخة بالآلة الكاتبة، حيث تتسم الكتابة المطبوعة بالتغيرات الحادة في أشكال الحروف ، والمرونة في طرق تشبيكها وتداخلها وتراكبها، وذلك تحقيقا للقيم الجمالية للخط العربي. تزيد هذ العوامل من صعوبة تمييز النصوص العربية، حيث على النظام الآلي في هذه الحالة أن يتعامل مع تعدد أشكال الحروف، وتباين أحجامها، وتراكمها فوق بعضها، وتداخلها مع بعضها، وتغير أحجام علامات التشكيل المعدلة للحروف ، وأوضاعها بالنسبة لهذه الحروف، وهلم جرا . وعلى ما يبدو، تحتاج عملية التمييز للنصوص المطبوعة إلى اتخاذ وحدة كتابية أكبر من الحروف كأساس لعملية التحليل الشكلي، وهو ما يذكرنا بالمسائل المتعلقة بتداخل الأصوات في نظم تمييز الكلام المتصل (انظر الفقرة ٧: ١١ من الفصل السابع).

يحتاج تحديد الوحدة الكتابية المذكورة إلى تحليل دقيق للأصول والمبادىء والاصطلاحات المتبعة في الخطوط العربية الجمالية، وذلك تمهيدا لوضعها في قواعد رسمية تراعي الحساسية السياقية العالية التي تتميز بها الخطوط العربية (ويمكن استخدام نفس هذه القواعد في العملية العكسية لتوليد هذه الخطوط آليا)، بتطبيق هذه القواعد يمكن استنتاج جميع التركيبات الممكنة لأشكال الحروف العربية في مواضعها المختلفة داخل

الكلمات ـ بعدها، يتم تطبيق أساليب تصنيف الأنماط على هذه التركيبات لاستخلاص الوحدات الكتابية المندمجة مع بعضها والتي يجب التعامل معها كوحدة شكلية واحدة، تمثل هذه الوحدات الأبجدية الرمزية التي يتعامل معها نظام التمييز الآلي، ويوضح (شكل ٤ : ٩) بعض عينات من هذه الوحدات.

شكل (٤ : ٩) بعض الوحدات الشكلية المندمجة للخطوط العربية المختلفة

نظرا لاختلاف أشكال الحروف العربية في أنماط الحروف المختلفة (نسخ، رقعة، ثلث، كوفي، . .)، يفضل أن يقوم نظام التمييز الآلي بالتعرف على نمط الخط تلقائيا، لإمكان ذلك يجب تحديد حد أدنى من السمات الفارقة التي يمكن على أساسها التفريق بين نمط خطي وآخر، ومن خلال عملية مسح سريعة لجزء من النص المكتوب، يلتقط النظام الآلي هذه السمات الفارقة، والتي على أساسها يقوم بتحديد نمط الخط المستخدم.

فيما يخص عناصر التشكيل، تواجه عملية تمييزها آليا بعض المصاعب الفنية نظرا لصغر حجمها، وتداخلها مع نقاط الإعجام أحيانا، وتغير وضعها بالنسبة للحروف الحاملة لها، في المقابل، هناك كثير من خصائص التشكيل المتنظمة يمكن استغلالها للمعاونة في عملية التمييز المذكورة، من أمثلتها: عدم جواز وضع علامات السكون على الحرف الأول للكلمة، وانحصار التنوين على الحرف الأخير منها، وعدم تتالي علامتي سكون، وما شابه. بجانب ذلك يمكن للمعالجة الآلية للصرف والنحو أن تلعب دورا حاسما في التمييز الآلي لعلامات التشكيل، وذلك على مستويين:

المستوى الأول : الاستهداء بهذه المعالجات في حالة فشل نظام التمييز الآلي في تحديد

علامات التشكيل بشكل قاطع.

المستوى الثاني : فصل علامات التشكيل عن عملية التمييز الآلي، وإحالة عملية تحديد عناصر التشكيل لمعالج التشكيل الآلي، والذي سبق ذكره في الفقرة (٣: ١١) من الفصل الثالث، ويستعين المعالج المذكور بالمعالجات الصرفية والنحوية، في حالة اللبس، أي وجود أكثر من احتمال لتشكيل بعض الكلمات، يمكن الرجوع لنظام التمييز الآلي لاختيار أقربها على ضوء علامات التشكيل الفعلية التي تتضمنها الكلمة موضع اللبس.

المصدر: نبيل علي، اللغة العربية والحاسوب.