أثر تحول وسط الكتابة العربية الى ألكتروني في تفاقم أزمة الإنقطاع الحضاري

الاستاذ الدكتور عبد المنعم صالح أبوطبيخ

عميد كلية علوم الحاسوب / الجامعة التكنولوجية- بغداد

الدكتور عبداللطيف علي حسين

رئيس قسم علوم الحاسوب / كلية التربية /الجامعة العراقية

خلاصة:

الكتابة هي من أعظم انجازات البشرية التي مكنت الامم من التواصل مع الأمم الأخرى، وكذلك بين اجيالها عبر ازمانها المتعاقبة بوساطة التدوين. ومهدت ظروف اقتصادية، سياسية، اجتماعية، وتقنية لتطور اساليب الكتابة وتعددها. إلا أن تطور اساليب الكتابة لم يكن خاليا من المساوئ تماما، بل حمل في طياته مخاطر كثيرة من أهمها ابتعاد الاجيال اللاحقة، المتمتعة بالاساليب المتقدمة في الكتابة، عن تراثها المكتوب باساليب كتابة قديمة. أطلق على ظاهرة ابتعاد الاجيال الناشئة عن نتاج امتها التاريخي الثقافي والمدني اسم "الانقطاع الحضاري".ويحدث الانقطاع الحضاري ،في عصرنا بعد ظهور التدوين الالكتروني.

وهذا البحث يعرض أثر ظهور التدوين الالكتروني في تفاقم أزمة الانقطاع الحضاري عن التراث العربي والاسلامي بالتحديد، وسبل المعالجة من خلال تحويل الكتابة المطبوعة والمخطوطات الى الوسط الالكتروني باستخدام تقنية تمييز الحروف الضوئي (Optical Character Recognition – OCR)، ثم إخضاعها للتدقيق اللغوي والنحوي، يدويا وآليا.

يعرض البحث المحاولات العديدة لحل هذه الازمة والمشكلات التي واجهتها، ويقترح الأساليب والمتطلبات الداعمة لحلها، من ذخائر لغوية وأدوات معالجة اللغات الطبيعية ودراسات لسانية حاسوبية وغيرها، في محاولة لطرح رؤية لأبعاد المشكلة وآفاق الحلول المتاحة و المستقبلية.

مقدمة

كان نشوء الكتابة احد أعظم انجازات البشرية التي مكنت الامم من التواصل بين اجيالها عبر ازمانها المتعاقبة بوساطة التدوين. ومهدت ظروف اقتصادية، سياسية، اجتماعية، وتقنية لتطور اساليب الكتابة وتعددها. إلا أن تطور اساليب الكتابة لم يكن خاليا من المساوئ تماما، بل حمل في طياته مخاطر كثيرة من أهمها ابتعاد الاجيال اللاحقة، المتمتعة بالاساليب المتقدمة في الكتابة، عن تراثها المكتوب باساليب كتابة قديمة. أطلق على ظاهرة ابتعاد الاجيال الناشئة عن نتاج امتها التاريخي الثقافي والمدني اسم "الانقطاع الحضاري". إن اهمية التركيز على أثر التغير في اساليب التدوين عبر الزمن يأتي من كون التدوين هو الوعاء الذي تحتفظ فيه الامم بنتاجها الحضاري.

حدث الانقطاع الحضاري مرات عديدة عند تغير وسط التدوين، بشكل واضح عند انتقال الكتابة من الوسط الحجري الى الجلود، وعند انتقال الكتابة من الجلود الى الورق، ويحدث في عصرنا بعد ظهور التدوين الالكتروني، وهذا البحث يعرض أثر ظهور التدوين الالكتروني في ظاهرة الانقطاع الحضاري عن التراث العربي والاسلامي بالتحديد. ويحاول البحث طرح رؤية لابعاد المشكلة وآفاق الحلول المتاحة و المستقبلية.

مفاهيم اساسية كما تم التعبير عنها في البحث:

الثقافة : هي طيف منِ المعتقدات الدينية والقيِم والنظم والتشريعات الحاكمة واللغة والإرث، التي تشكِل بالنسبة للأمة شخصيتها وهويتها بين الأمم، وهذا ما نعبر عنه بثقافة الأمة، أو الثقافة المجتمعية.

الحضارة : المظاهر التي تعبر بها الأمة عن ثقافتها، وتعمل من خلالها على حماية هذه الثقافة وصيِانتها وتوريثِها للأجيال وعولمتها بين الأمم.

المدنية : وهو عبارة عن :مظاهر حضارية مادية نشأت عن تفاعل بين العلوم التجريبية وتطبيقاتها من ناحية وبين منتجات الأمةالحضارية من ناحية أخرى،وهي مرتبطة غالبا بالمجالين الصناعي والاقتصادي،[1].

الانقطاع الحضاري المشكلة والمسببات

فكرة الانقطاع الحضاري أو الحداثة التامة تعبر دائما عن نظرة مجتمع إلى مجتمع معين وهي نظرة خارجية لا تفسر لنا حركة الواقع واستعداداته القابلة للتفتح في المستقبل ونمو الجديد في أحشاء القديم. كما لا تستطيع أن تقدم لنا تصورا مقبولا لموضوعية التاريخ وتاريخية الحضارة،[2].

تطور أدوات الكتابة

تنوعت المواد التي استخدمها الانسان في كوسط للكتابة عليها وأبرزها، ،[3]:

1) **الطين:** أولى علامات الاتصالات البصرية من الكتابة هي رموز طين وضعت لأغراض المحاسبة في بلاد ما بين النهرين في الألف التاسع قبل الميلاد، [11]. دونت اللغة السومرية بالرقم الطينية. و في عصر السلالات تقدمت الكتابة مراحل أخرى بحيث أصبحت وسيلة لتدوين شؤون الحياة المختلفة والسجلات الرسمية . ومن أقدم الألواح والرقم الطينية المكتشفة هي تلك التي اكتشفت في أور والتي تعود إلى حوالي.(2700 ق.م). ظلت الكتابة المسمارية مستعملة في التدوين حتى بعد انتهاء آخر الأدوار الحضارية في العراق القديم وسقوط بابل في عام 539 ق.م. وقد وصلت إلينا رقم طينية من الفترات المتأخرة مؤرخة بما يقابل سنة 150 ق .م حمل اسم كاتبها و هو مساعد لأحد الكهنة، [4].

بقيت الكتابة المسمارية مستعملةً في التدوين عبر مسيرة من الزمن تنوف عن ثلاثة آلاف سنة . واستخدم الخط المسماري لتدوين لغتين هما السومرية والأكدية. واللغة الأكدية تفرعت إلى لهجتين هما البابلية والآشورية . ولم ينحصر استخدام الخط المسماري في بلاد وادي الرافدين، بل خرج من موطنه الأصلي وانتشر إلى مناطق أوسع، فتأثرت به الأقوام التي عاشت على أطرافه والبعيدة منه فاقتبست الكثير منه وتغذت من معينه، [5].

2) **الجلد:** وكانوا يسمونه الرق، والأديم، والقضيم، والفرق بينها غير واضح من النصوص والروايات نفسها، ولكن المعجمات تجعل الرق الجلد الرقيق الذي يسوى ويرّقق ويكتب عليه، وتجعل الأديم الجلد الأحمر أو المدبوغ، وتجعل القضيم الجلد الأبيض يكتب فيه، وقد ورد ذكرها كّلها في الشعر الجاهلي. كتب الآراميون لغتهم على الجلود، [3].

3) **القماش:** وهو إما حرير وإما قطن، ويطلقون على الصحف إذا كانت من القماش: المهارق، مفردها المهرق، وهو لفظ فارسي معرب، ،[3].

4) **النبات:** وأشهر أنواعه العسيب، وجمعها عسب وهو السعفة، أو جريدة النخل إذا يبست وشذبت من خوصها. وقريب من العسيب: الكرنافة وجمعها كرانيف، وهي أصول السعف العراض اللاصقة بالجذع، وقد ورد أن الوحي كان يكتب في عهد الرسول صلى الله عليه وسلم على العسب والكرانيف. ومن ذلك الرحل ويكون من الخشب الذي يوضع على ظهر الجمل ليركب. ومن الخشب أيضًا الروسم (الخشب العريض) والسهم والعصا، ،[3].

5) **العظام:** وأشهر أنواع العظام التي كانوا يكتبون عليها: الكتف واللوح والأضلاع.

6) **الحجارة:** وهي حجارة بيض رقاق يكتب عليها، وقد كتبت عليها آيات القرآن الكريم على عهد الرسول صلى الله عليه وسلم ، ومن تمام الحديث عن النقش على الحجارة أن نشير إلى النقش والكتابة على البناء،فقد ورد أن رسول الله صلى الله عليه وسلم، لم يدخل الكعبة يوم الفتح حتى أمر بالزخرف فمحي وأزيل، وأمر بالأصنام فكسرت. وقد روى ابن الكلبي أنه أخذ علمه بأنساب العرب مما وجده على جدران كنائس الحيرة، ،[3].

7) **الورق:** يمكن تتبع صناعة الورق إلى حوالي 105 م، عندما صنع (تساي لن Ts'ai lun) ، وهو مسؤول في المحكمة الإمبراطورية في الصين، صفحة من الورق باستخدام التوت ، ومخلفات القنب و ألياف أخرى. في ترحاله البطيء توجه فن صناعة الورق نحو وسط آسيا ،فوصل سمرقند في 751م. وفي 793م، صنعت الورقة الأولى في بغداد خلال عصر هارون الرشيد، ، العصر الذهبي للثقافة الإسلامية التي جلبت صناعة الورق إلى حدود أوروبا بحدود القرن الحادي عشر. و بحدود القرن الرابع عشر أنشأت مصانع الورق في أوربا، [14].

استخدم العرب الورق في الكتابة. ذكر ابن النديم أنه رأى أوراقا من ورق الصين فيها كتابة بخط يحيى بن يعمر المتوفى عام 90 هجرية، إضافة إلى أن عثمان بن عفان عزم على كل رجل معه من كتاب الله شيء أن يذهب به إليه، وكان الرجل يجيء بالورقة والأديم فيه شيء من القرآن. وقال عمرو بن نافع مولى عمر بن الخطاب " فاستكتبتني حفصة بنت عمر مصحفًا لها، فلما بلغت إليها حملت الورقة والدواة " وغير ذلك من الأدلة. ولقد كانت لهذه المواد المكتوبة ألفاظ عامة يطلقونها عليها ليدّلوا على المكتوب، وما كتب عليه معا، من ذلك الصحيفة. وقد تكون جلدًا أو قماشا أو نباتا أو حجرًا أو عظمًا أو ورقا،[3].

8) **الكتاب:** وهي لفظة قد تكون أعم من الصحيفة، وأكثرمنها شيوعًا، ،[3].

9) **الزبور:** وجمعها زبر، وقد يراد بها الكتاب الديني، و تطلق أيضًا على غيره من الكتب، [3].

العوامل المؤثرة في تطور الكتابة تاريخيا

ارتبط تاريخ الكتابة بعوامل أساسية، [3]:

الأول: عامل حضاري قائم على العلم، ونشوء التدوين،وأسلوبه (الحروف والرموز).

الثاني: عامل مادي – فني متعلق بتوفر مواد الكتابة من البردي، والرق والورق فيما بعد،وتهيئتها لصناعة الكتاب.

الثالث : ثقافي – سياسي، كان الحراك السياسي والتطور العلمي والتفاعل الحضاري يطور وسائل التواصل بين الشعوب، وله تأثير كبير على تفعيل العوامل في أعلاه. ومثال ذلك التطور الذي جرى في عصر الدولة العباسية،فقد حدث التطور لعاملين، [7]:

الأول : سياسي، هو انتصار المأمون ذي الاتجاه الخرساني على أخيه محمد الأمين ذي الاتجاه العربي يناصره البغداديون ، وقد نهبت الدواوين التي كتبت بالجلود في أيامه ، ولعل ذلك بسبب الحرب .

الثاني : اجتماعي ،وذلك أن صناعا من الصين عملوه بخرسان من الكتان على مثال الورق الصيني، وهو تقليد للصناعة الأصلية التي قال عنها ابن النديم إنه (يعمل من الحشيش، وهو أكثر ارتفاع البلد)، [9]، إي أن ثمنه مرتفع لجودته.

أثر تبدل وسط الكتابة مع الاستدلال بالقضايا التاريخية

أ. التحول من الحجر الى الجلد

في أواخر العصر الآشوري الحديث، فرضت العوامل الاجتماعية تغيرا في المادة المستخدمة للكتابة عليها لمسايرة التطور الحاصل في الجوار بعد قدوم الآراميين للمنطقة، فقد (كان لملوك الآشوريين كاتب آرامي يكتب على الجلد أو لفائف البردي أو ألواح خشبية أو العاج الى جنب كتبة المسمارية الذين يكتبون على ألواح الطين)، [12].

وفي العصر الآشوري الحديث (911–612 ق.م) والعصر البابلي الحديث (626–539 ق.م) أخذت الكتابة الآرامية المكتوبة على الرقع الجلدية تحل محل الكتابة المسمارية على الرقم الطينية، [7].

في قصة تاريخية له، كتب هيرودوت أن الآشوريين من بلاد ما بين النهرين (قبل الميلاد القرن 8) قد فضلوا جلود الحيوان على الألواح الطينية وكتبوا على جلود الأغنام والماعز الخالية من الشعر. العملية البسيطة المستخدمة في وقت مبكر خلقت صعوبات في تجفيف الجلد الرطب، لجعل الجلد ورقة ناعمة وشريحة خالية من التجاعيد، تنتقص من استخدامها مما يقود لاستخدام ورق البردي كونها وسط الكتابة السائد في جميع أغلب منطقة الشرق الأوسط، لتجعل مملكة مصر مركز الإنتاج الرئيسي.

ب. التحول من الجلد الى الورق

ذكر ابن النديم في كلام مركز أدوات الكتابة وعوامل تطورها خلال العصور، والبلدان التي ظهرت فيها، فأشار إلى إن الناس كتبوا على الطين زمن آدم، ثم كتبت الأمم على النحاس والحجارة قبل الطوفان، وبعدها على الخشب وورق الشجر، وكتب أهل مصر في القرطاس الذي يعمل من قصب البردي، وكتب الصينيون في الورق الذي يعمل من الحشيش، وقد فسر عوامل التطور بالرغبة في الخلود وحاجة الوقت،[9].

والشيء المهم هو أدواتها في العصر العباسي، وقد اشار ابن النديم إلى تطور مرت به بقوله، [9]: (فأما الورق الخرساني فيعمل من الكتان – ويقال انه حدث في أيام بني أمية، وقيل في الدولة العباسية، وقيل انه قديم العمل – وقيل :انه حديث وقيل إن صناعا من الصين عملوه بخراسان على مثال الورق الصيني فإما أنواعه :السليماني الطلحي، النرحي، الفرعوني، الجعفري، الطاهري، أقام الناس ببغداد سنين لا يكتسبون إلا في الطروس، لأن الدواوين نهبت في أيام محمد بن زبيدة وكانت في جلود، فكانت تمحى ويكتب فيها، قال :وكانت الكتب في جلود دباغ النورة، وهي شديدة الجفاف، ثم كانت الدباغة الكوفية تدبغ بالتمر، وفيها لين) ، فالتطور الحاصل هو انتقال

الكتابة من مادة الجلود إلى الورق، وهو تطور حاسم في الثقافة، ويحتمل إن يتسرب هذا التطور من مصدرين مصر والصين، والذي حدث انه اقبل من الصين، لأن الدولة العباسية التي قامت في خراسان وارتبطت باهلها، وكانت اقرب إلى الصين موقعها وتأثراً، واقرب الاقوال صحة انه حدث في الدولة العباسية، لأن الجاحظ ينقل في رسالته (في الجد والهزل)، [8]،التي أهداها إلى محمد بن عبد الملك الزيات جدلا بين أنصار الجلود وأنصار الورق ، مما يعني أن التطور قريب عهد بزمن تأليف الرسالة ، ولو حدث أيام بني أمية لاستقر الناس على صنف منهما خلال هذه المدة التي لم تكن قصيرة واختفى الجدل حول الصنف الأخر، بل إن الناس قد استعملوا الصنفين أيام الجاحظ ، مما يدل إن التطور من مراحله الأولى ، وقد حدث التطور لعاملين:

العامل الأول سياسي، هو انتصار المأمون ذي الاتجاه الخرساني على أخيه محمد الأمين ذي الاتجاه العربي يناصره البغداديون ، وقد نهبت الدواوين التي كتبت بالجلود في أيامه ، ولعل ذلك بسبب الحرب .

والعامل الثاني : اجتماعي ،وذلك أن صناعا من الصين عملوه بخرسان من الكتان على مثال الورق الصيني، أي أنه تقليد للصناعة الأصلية التي قال عنها ابن النديم (إنه يعمل من الحشيش، وهو أكثر ارتفاع البلد)، [9] ، والمقصود إرتفاع ثمنه لجودته ، ويبدو إن صناعته تطورت زمن ابن النديم .

لكل من الورق والجلود مسميات تعارف الناس عليها ، ولكونه صناعة مستوردة ، أطلقت عليه أسماء دخيلة ، منها "الطامور والطومار "ويعني الصحيفة ، وعده بعضهم عربيا، ذكره الجاحظ بقوله: رب كلام قد ملأ بطون الطوامير ، قد عرف جملته وما فيه الضرر منه بسحاءة أو حرف تبين من ظهره، ومعلوم إن الذي يمزق بسرعة ، وتنتشر منه أجزاء ، والذي تبين حروفه من ظهره هو الورق لخفته ، والدفتر يعني جماعة الصحف المضمومة والدفتر واحد حيث تطلق اليوم على الأوراق البيضاء المجموعة بحجم صغير ، والقدماء كانوا يطلقونها على مجموعة الكراريس التي تضم إلى بعضها ، لتؤلف بعد الكتابة عليها كتابا صغيرا فهي أشبه بالدوريات ، وقد روى الجاحظ حديثا على لسان الزيات أحد أنصار الجلود يشير الى ذلك (قلت لي، ليس لدفاتر القطني أثمان في السوق وان كان فيها كل حديث طريف ولطف مليح وعلم نفيس، ولو عرضت عليهم عدلها في عدد الورق جلودا ثم كانت فيها كل شعر بارد وكل حديث غث ، لكانت أثمن ، ولكانوا عليها أسرع)، ورخصها في السوق يعود إلى صغر حجمها ومضمونها فأوراقها قليلة والعمل المبذول فيها وهو الكتابة يتطلب وقتا أقصر وجهدا أقل ، ومضمونها لا يحتوي علما أو مادة علمية عميقة ، وإنما هي قصص وأحاديث تسلية ونوادر ومجموعة أشعار أو دواوين صغيرة ، وقد روى الجاحظ حديثين عن أبي عمرو بن العلاء قال : وما دخلت على رجل قط ، ولا مررت ببابه فرأيته ينظر في دفتر وجليسه فارغ اليد إلا اعتقدت انه أفضل منه وأعقل، [8].

اختفت الأدوات القديمة للكتابة في العصر العباسي وفي زمن الجاحظ باستثناء الجلود، ونلاحظ في (رسالة الجد والهزل) للجاحظ إن الاستعمال اقتصر على الورق الصيني والكاغد الخرساني والجلود، وقد ندر استعمال الصيني لأرتفاع ثمنه وبقي الكاغد الخرساني والجلود يتنازعان السيادة، والكاغد الخرساني يعد تطوراً خاصاً في الثقافة ولكل منهما أنصاره وخصومه وقد عكست رسالة الجاحظ، جدلا بين أنصارهما فالجاحظ نصير للورق ومحاوره محمد بن عبد الملك الزيات الذي أهديت الرسالة إليه نصير للجلود، ولكل حججه، إذا عرضت بمعزل عن ظروفها ونظر إليها على وفق تصور وتطور المجتمع المعاصر، بدت سطحية. إلا إن الكاغد والجلود في عصر الجاحظ أعمق دلالة،

لأنها تعبر عن موقف حضاري وتصدر عن واقع ثقافي متطور، يجب إن ينظر إليه من خلال حديث أنصارهما، فقد قيل :(على الجلود يعتمد في حساب الدواوين وفي الصكاك والعهود وفي الشروط وصور العقارات، وفيها تكون نموذجات النقوش وفيها تكون خرائط البر،[3]، و جوانب إدارية تتضمن سجلات الجند ومرتباتهم وعقارات الأرض والضياع وصكوك الممتلكات والعهود والاتصالات، ويكون الاهتمام بهذه الجوانب في المرحلة المتقدمة من الحضارة أو المتطورة، أي في بداية نشؤ الدول عندما تكون بحاجة ماسة إليها لتنظيم شؤونها، أو في مرحلة متقدمة من تطورها عندما تزدهر حياتها، وتتشعب ميادين عمله. والدولة العباسية في عصر الجاحظ قطعت شوطا في التطور إلا أنها لم تنضج إلا في القرن الرابع. فلا عجب إن يدافع الزيات عن استخدام الجلود التي يعتمد عليها رجال الدولة في تنظيم الشؤون المالية والادارية، والجلود –في نظره (اصلح للجرب ولعفاص الجرة وسداد القارورة)،[8].

ج. التحول من الكتابة اليدوية الى الطباعة الآلية

لقد أعاقت الحاجة إلى النسخ باليد انتشار المواد المكتوبة . ولكن مع اختراع مطبعة غوتنبرغ في عام 1454، أصبح من الممكن إنتاج الكتب بسرعة وبتكاليف زهيدة على نطاق واسع، [13].

التحول من الورق الى التدوين الالكتروني

لم يعد بإمكان القائلين بعدم قدرة الكتاب الألكتروني على منافسة الكتاب الورقي في الاستخدام، التمسك برأيهم. لقد كانت المقاييس تدور حول رخص ثمن الكتاب الورقي مقابل جهاز الحاسوب، وخفة حمل الكتاب، حتى أن القاريء يحمل كتابه للقراءة في فراشه بخلاف الحاسوب، وأن القراءة في الكتاب لا تؤذي عين القاريء مثلما تفعل شاشة الحاسوب. ولكن الحال قد تغير منذ ثمانينات القرن العشرين، بتغير ميزات أجهزة الحاسوب الألكترونية، و بتغير قدرات أدوات الكتابة الألكترونية، [14].

فبدلا من الاوراق الثابتة للكتاب الورقي، يمكن للكتاب الألكتروني أن يوفر الصفحات المتحركة والمتغيرة، المرتبطة بشبكة متشعبة من الصفحات الأخرى ذات العلاقة. ويمكن القاريء من التفاعل مع النص بوساطة معالج النصوص أو قواعد البيانات النصية، تعديلا ، مقارنة نصوص، بحث عن نص، إضافة هوامش، والعديد من المزايا التي يصعب حصرها. إن ذلك يضع مصير الأعمال المكتوبة على الكتب الورقية محل تساؤل كبير. واللغة العربية تفتقر إلى جهد كبير لتحويل التراث الهائل من النصوص الورقية الى نصوص ألكترونية.

إن إولى المحاولات لتحويل المطبوعات الورقية كانت بتصويرها وخزن صورها ألكترونيا. ولكن ذلك لم يكن يصل بصور صفحات المطبوعات الورقية، الى مستوى الكتاب ذي النص الألكتروني المكتوب بمعالجات النصوص الألكترونية.

تحويل الكتابة المطبوعة والمخطوطات الى الوسط الالكتروني

مما سبق يتبين أن الحل يجب أن يقترن باستخدام تقنية تمييز الحرف الضوئي. (Character Optical–OCR Recognition) وتحويل النص المطبوع الى نص الكتروني.

لقد وصلت العديد من أنظمة التعرف الضوئي على الحروف العربية (AOCR) مرحلة متقدمة، ولكن دقة وأداء هذه النظم لا تزال غير مرضية بالمقارنة مع أنظمة التعرف الضوئي على الحروف اللاتينية، [15].

هذا الواقع هو نتيجة لأسباب عديدة مثل: تفاصيل زخرفية لخطوط الكمبيوتر، والعدد الكبير من أشكال الحروف التي تتغير وفقا لموقع الحرف في الكلمة. وهذا يؤدي إلى مشكلة صعبة، وهي العثور على الطرائق الأكثر كفاءة لاستخراج الميزة للتمييز،[15] [16].

الخصائص التالية للغة العربية المكتوبة يجعل الاستخراج مشكلة متحدية:

- الأحرف العربية المكتوبة تأتي في أحجام مختلفة غير متناسبة في نفس حجم الخط، الأمر الذي يجعل الحروف تحتل مناطق لا يمكن التنبؤ بها داخل الكلمة.
- يمكن أن يظهر الحرف نفسه في أشكال وأحجام مختلفة تماما داخل نفس الكلمة، مما يجعل الاستخراج المعتمد على خصائص حرف معين غير كفؤ.
- الأحرف يمكن أن تظهر في اتجاهات مختلفة، لأن الأحرف العربية في الكلمة ليست الانحياز وأنها يمكن أن تمتد فوق أو تحت خط.
- وجود النقاط والهمزة (ء) في أشكال مختلفة فوق الحروف. يؤثر حجم الخط إلى حد كبير في الحروف وخصائص النقاط والأشكال.
- (على سبيل المثال النقاط الثلاث تكون مرتبطة في الخطوط صغيرة)
- خطوط عربية مختلفة تشكل تغييرا جذريا في رسم وخصائص الحرف مع أنواع الخط المختلفة. ويلاحظ أيضا أن حروفا مختلفة في خطوط مختلفة قد تظهر متماثلة.

تطوير برمجيات اللغة العربية

إن تطوير برمجيات معالجة اللغة العربية ألكترونيا هو أمر محوري، من أجل توفير أدوات برمجية تؤدي آليات كانت تنجز يدويا، حيث أن ذلك المجهود اليدوي لن يستوعب الحجم الكبير الذي تتصف به المجموعات الوثائقية اللغوية الحديثة. وفيما يأتي أهم هذه الأدوات البرمجية:

1) برنامج محلل الصرف العربي (Arabic Morphological Analyzer)

يؤدي برنامج المحلل الصرفي مهمة مسح مفردات اللغة العربية باختلاف انواعها التاريخية والجغرافية، وتوليد التصنيفات الدلالية المعجمية الممكنة (Lexical Semantic Specification) لكل كلمة يتم تحليلها صرفيا. تتصف آلية التحليل الصرفي العربي بقابلية زيادة تدريب النظام على نوع محدد من النصوص بهدف زيادة دقة التحليل الصرفي. كما تتصف ايضا بقابلية اكتشاف و تصحيح الأخطاء الهجائية الشائعة. و تمتاز بالتعامل بمرونة مع الكلمات المدخلة، سواء كانت بالحركات الاعرابية أم بدونها.

عملية الصرف فى اللغة العربية هي المنظومة المسؤولة عن بنية الكلمات بالتحليل والتكوين، وتعتمد في مدخلاتها على منظومة المعجم التى تزودها بالجذور والفروع (الجذوع)، التى تطبق عليها التحويلات الصرفية لتكوين كلمات

اخرى. تعتمد في تحديد معانيها على منظومة الدلالة لتحديد المعنى الوظيفى المراد بناء الكلمة على وفق قالبه، ثم يأتي دور منظومة النحو التى يعتمد عليها في تعيين وظيفة الكلمة في سياق الجملة و تحديد حالتها الإعرابية.

إن لمنظومة الصرف وظائف يمكن إجمالها كالآتي:

أ. تركيب الكلمات بطرق مخالفة للاشتقاق كالنحت و المزج والتركيب.

ب. تحديد المتغيرات التى تحدث في بنية الكلمة استنادا الى حالتها الإعرابية.

ج. بناء الكلمة بالاشتقاق.

د. تحديد و اجراء بعض الطرائق الصرفية عند حالات التأنيث والإضافة و ما شابهها.

إن منظومة الصرف هي منظومة حيوية بين مجمل منظومات اللغة العربية لانها ترتبط مع باقي عناصر المنظومات في النحو والصوت و الدلالة ومعجميا ، و لذلك، فان بناء آلية التحليل الصرفي يجب ان يكون محور تقنيات معالجة اللغة العربية، لتبنى عليها تقنيات اخرى هامة للتحليل الصرفي، للتشكيل الآلي، العنونة الصرفية، والتعامل مع اللبس (Ambiguity) و إزالته.

2) برنامج العنونة الصرفية للنص العربي (Arabic Tagger)

توفر آلية العنونة الصرفية للنص العربي إمكانية تمييز و/ او عنونة أجزاء من الكلام، وهي بذلك عملية ضرورية لتوليد أجزاء الكلام التي تمثل خصائص الصرف من الكلمات البسيطة للنص، بدون اعتبار السياق.

الى جانب ذلك تعد العنونة اللغوية ذات اهمية بالغة لمهام المعالجة اللغوية، حيث تعد العناوين الناتجة بيانات مدخلة لجميع معالجات الحاسوب في عملية التقطيع النحوي الذي يعد بحد ذاته، خطوة نحو فهم اللغات والترجمة الآلية.

المنظومة النحوية هي المنظومة المسؤولة عن ترتيب و تنظيم الكلمات ومكوناتها في سياق الجملة، و مدخلاتها تأتي من منظومة الدلالة التي تعطي لقواعد النحو، المعاني العامة والوظائف الدلالية لمكونات الجملة وهي بذلك تمثل البناء المنطقي للعبارة اللغوية.

ان منظومة الصرف و منظومة المعجم تغذيان منظومة النحو بالمفردات، بعد تنفيذ المعالجات الصرفية المناسبة لها. فيتم إدخال نص عربي أولي مثل نصوص الجرائد و بعد ذلك، تقوم آلية العنونة الصرفية بتوليد مجموعة من العناوين (Tags) لكل كلمة، لتحدد صفاتها النحوية.

إن لمنظومة العنونة الصرفية وظائف يمكن إجمالها كالآتي:

أ. تعيين الترتيب الافتراضي للكلمات في سياق الجملة.

ب. تنفيذ التحويلات البنائية على بنية الجملة، مثل الإضمار و الحذف، التأخير و التقديم ، و ما شابهها.

ج. إضافة التعديل النحوي الذي يحدد وظيفة كل كلمة في سياق الجملة، مثل عملية التشكيل الاعرابي، ومطابقة الفعل و فاعله، و ايضا التعديل الناجم عن التعديلات التركيبية.

ترتبط منظومة العنونة الصرفية بتطبيقات مهمة مثل :

أ. تقنية التشكيل الآلي، حيث يكون مخرج عملية العنونة الصرفية مدخلا لعملية التشكيل الآلي.

ب. المحلل النحوي في اللغة العربية (Arabic Syntax Analyzer)، فتقنية العنونة الصرفية هي بداية أساسية لانشاء أي محلل نحوي، فيتم إدخال نص عربي أولي مثل نصوص الجرائد، لتقوم آلية العنونة الصرفية بتوليد مجموعة من العناوين لكل كلمة، تمثل صفاتها النحوية المحتملة في سياق الجملة.

3) التشكيل الآلي (Automatic Diacritization).

هي تقنية مبنية على أساس تقنية التحليل الصرفي ، وتتمتع بخصائص وإمكانيات التعامل مع النص العربي. و كذلك التعامل مع الكلمات الأعجمية المكتوبة بحروف عربية. وتتميز بإمكانية التعامل مع الاختصارات السائدة والأرقام بصيغها المتنوعة و ارجاعها إلى الصورة النصية. و حل مشاكل ضبط الاواخر مثل مشكلة التقاء الساكنين بين كلمتين متتاليتين، و بما يناسب النص العربي.

هذه التقنية هي تشكيل آلي لحروف كلمات اللغة العربية داخل النصوص المكتوبة، وتصمم وتبنى لمعالجة النص العربي كاملا، بضمنه، الأرقام والاختصارات والأسماء الاعجمية المستخدمة. يتم إدخال نص عربي أولي، فتقوم هذه الآلية بتشكيله على وفق نمط محدد، و من هذه الأنماط: (تشكيل إملائي، تشكيل صوتي، ...الخ).

ترتبط تقنية التشكيل الآلي بعدة تقنيات أهمها تقنية التحليل الصرفي العربي، وتقنية النحو الصوتي العربي. وتعد تقنية التشكيل الآلي للنص من تقنيات البنية التحتية لعدد من تقنيات معالجة الصوت العربي مثل تحويل الكلام المنطوق إلى الكلام المكتوب .

4) محرك البحث العربي (Arabic Engine search)

هو محرك بحث عربي لاسترجاع الملفات والبحث عن المعلومات خلال :

أ. البيانات الإلكترونية.

ب. المستندات.

ج. قواعد البيانات.

د. مواقع الإنترنت.

محرك البحث هو نتاج الأبحاث في مجال المعالجة الطبيعية للغة (NLP) مخصص للعربية ، ويفترض أن يتميز المحرك بالمرونة وسعة الاستخدام.

يستخدم محرك البحث العربي تقنية المعالجة الطبيعية للغة، لتصحيح الأخطاء الإملائية، كما يتيح البحث باللغة العربية، باستخدام جذر الكلمة أو أحد اشتقاقاتها خلال كم كبير جدا من البيانات، ليعطي نتائج بحث دقيقة.

من خلال محرك البحث العربي، يمكن رسم صورة كاملة و دقيقة لكل المعلومات المتوفرة في نظم الملفات، مواقع الويب، قواعد البيانات، أو قواعد المعرفة.

يوفر محرك البحث العربي إمكانات و خصائص اهمها :

أ. فهرسة البيانات كبيرة الحجم بسرعة فائقة.

ب. التصحيح الاملائي للأخطاء في اللغة العربية.

ج. استخدام جذر الكلمة او احد مشتقاتها باستخدام تقنية التحليل الصرفي.
د. استرجاع موضوعات مماثلة او ذات صلة بموضوع البحث.
هـ. البحث بستخدام المترادفات اللغوية.
و. إمكانية التعامل مع مصادر مختلفة للبيانات مثل:

1) نظم الملفات
2) الأسطوانات المدمجة
3) الكتب التفاعلية
4) شبكات الحاسوب المحلية
5) قواعد البيانات
6) المواقع الإلكترونية

تقنيات معالجة اللغة الطبيعية

توفر تقنيات اللغة البشرية (Human Language Technology) (HLT) الإمكانية للوصول إلى المعلومات بطريقة أكثر طبيعية، وتدعم طرقا أكثر فعالية بتبادل المعلومات والسيطرة على نموها المتزايد. وهناك أيضا حاجة متزايدة لتوفير وصول سهل إلى أنظمة المعلومات ثنائية اللغة ولعرض إمكانية تداول المعلومات التي تحملها بطريقة مجدية و ذات معنى. إن التطور في تقنيات اللغة البشرية تقدم الوعد بوصول شامل للخدمات والمعلومات بطريقة آنية مباشرة (on–line). إن حقل تقنيات اللغة البشرية يغطي مدى واسع من النشاطات بهدف تمكين المستخدم من التواصل مع الآلات باستخدام مهارات الاتصال الطبيعية.

تتضمن نشاطات البحث والتطوير ما يأتي:

1) ترميز الكلام Speech Coding.
2) تمييزالكلام Speech Recognition.
3) توليد الكلام Speech Generation.
4) تفسير الكلام Speech Interpretation.
5) اللسانيات الحاسوبية Computational Linguistics
6) استرجاع المعلومات Information Retrieval(IR).
7) الترجمة الإلية Machine Translation.
8) مراكز الاتصال بقابلية بالكلام Speech Enabled Call Centers.
9) تعليم اللغة بمساعدة الحاسوب Computer Aided Language Learning (CALL).
10) تنقيب الأفكار Opinion Mining.
11) دعم ذوي الاحتياجات الخاصة Disabled Support .
12) أنظمة الرقمنة Digitization Systems .

ولكون اللغة العربية يستخدمها أكثر من ثلاثمائة مليون عربي [10]، ويستخدمها اكثر من مليار ونصف مسلم للتعبد، فإنها من أكثر اللغات أهمية في العالم. بالإضافة إلى أن اللغة العربية لها خصائصها المتفردة والمتحدية، يجب توجيه عناية خاصة لخدمة هذه اللغة المهمة. ليس فقط لإحداث خطط عمل محددة لمختلف التطبيقات البرمجية ذات العلاقة بالمنطقة، ولكن أيضا من اجل ضغوط

طلبات المجتمع للحفاظ على الهوية العربية .هناك جهد ملحوظ أنجز في تطوير تقنيات اللغة العربية، هي من ضمن المواضيع التي بحثت حديثاً حتى من قبل غير الناطقين بالعربية، في مجالات منها :

1) تمييز الكلام العربي .
2) تحويل النص العربي إلى كلام .
3) الترجمة من العربية إلى الانكليزية وبالعكس .
4) استخلاص المعلومات باللغة العربية .

بالرغم من أن بعض أدوات تقنيات اللغة العربية ومواردها هي موجودة حالياً فأن هذه الأدوات والموارد تبقى قاصرة في إتاحتها وجودتها بالمقارنة مع مثيلاتها من الأدوات بما يمكننا أن ندعوها لغات المستوى الأول (Tier– 1 Language)، التي تسيطر على الشبكة المعلوماتية مثل الانكليزية ، الفرنسية، واليابانية . يمثل هذا القصور صعوبة بالغة تؤثر في تبادل معلومات كفوء، الذي بدوره يؤثر على الحكومات والتعاملات التجارية ، والمجتمعات العربية بشكل كبير. ان كل الأمم تهتم بإدامة لغاتها وخدمتها بشكل جيد في عهد الرقمنة، و لايمكن توفير تقنيات اللغة العربية بصورة وافية، إلا بإتاحة كمية وافرة من موارد اللغة (Language Resources –LR) .

تقنيات معالجة الصوت العربي

إن احتواء المجموعة الوثائقية اللغوية العربية على مقاطع صوتية يستوجب توفر تقنيات معالجة الصوت لتداول المخزون اللغوي الصوتي، لإتاحة إمكانات تحليله ، تصنيفه، و تركيبه كما يحدث في النصوص المكتوبة في الجزء الآخر من المجموعة الوثائقية .

أ.تقنية تحويل النص العربي المكتوب إلى كلام منطوق (TTS) Text–To–Speech

يجب ان يتميز نظام توليد الكلام المنطوق آليا من النص العربي المكتوب بما يأتي:

1) جودة الصوت الخارج بما يماثل الصوت البشري الطبيعي.
2) احتواء تقنية تحويل النص الى صوت منطوق على تقنية تشكيل (ضبط) النص العربي آليا بدقة، والتعامل مع المصطلحات، الارقام ،التواريخ ، و الأسماء الأجنبية.
3) سرعة توليد الكلام بما يكافئ سرعة النطق البشري.

تقنية التوليد الالي للكلام المنطوق من النص العربي المكتوب (Text–To–Speech) توفر التفاعل الصوتي مع العملاء من خلال قواعد بيانات متجددة متاحة لهم باستخدام. تعتمد الفكرة الاساس لهذه التقنية على تحويل النص المكتوب إلى كلام مسموع، يتيح للمستخدم اختصار الوقت في قواعد البيانات التي يستغني عن الصوت البشرى لاعادة تغذية قاعدة البيانات الجديدة. فيتيح بذلك دعم التطبيقات المهمة والحلول الفعالة المبنية على تقنية توليد

الكلام المنطوق من النص المكتوب. العديد من التطبيقات و الحلول المهمة مبنية على تقنية توليد الكلام المنطوق من النص المكتوب. فصناعة المعلومات العربية تحتاجها على نطاق واسع، و من مجالاتها:

1) انظمة استعادة المعلومات باستخدام الصوت (Information Voice Retrieve systems)، من خلال مواقع ويب أو من خلال الخدمات الهاتفية.
2) انظمة القراءة الالية لشاشة الحاسوب الداعمة لاستخدام المكفوفين والاميين، ...الخ، والكثير من التطبيقات والحلول الاخرى.
3) صناعة الكتاب الالكتروني التفاعلي.

ب. تقنية التعرف الآلي على الكلام المنطوق Automatic Speech Recognition

يتم تطبيق تقنية التعرف الآلي على الكلام المنطوق على الكلمات المفردة الصغيرة لتوجيه الأوامر بواسطة الصوت، لتمكن المستخدم من توجيه بعض كلمات الاوامر للتقنية، فتتعرف عليها وتنفذ مضمونها.

تدخل هذه التقنية في استخدامات عديدة و متنوعة أهمها:

1) تعليم المكفوفين و لمساعدة الطلاب في التحصيل الدراسي.
2) أنظمة استعادة المعلومات باستخدام الصوت.
3) الاستعلام الهاتفي.
4) الاستفسار عن مواعيد رحلات الطيران والقطارات.
5) متابعة أسعار البورصة.
6) البحث في دليل الهاتف.
7) طلب نتائج الامتحانات.
8) الكتاب الالكتروني التفاعلي .

المجاميع اللغوية والبيانات التجريبية corpus

إن أهمية المجاميع اللغوية لبحوث اللغة و اللسانيات ، تأتي بالتوازي مع أهمية البيانات التجريبية (Empirical Data) لهذه البحوث.

البيانات التجريبية تمكن اللغوي من جعل العبارات الموضوعية منضوية ضمن التصور المعرفي (Cognitive perception) للغة، فضلا عن تلك التي هي ذاتية (Self–perception)، أو تلك المعتمدة على أساس الرؤية المعرفية الفردية الخاصة. لا يمكن للدراسات اللغوية و اللسانية الاعتماد على الحدس او التخمين او عينات صغيرة من بيانات اللغة، بل تحتاج هذه الدراسات الى بيانات تجريبية مخزونة في قواعد بيانات نصية كبيرة، كما في منهج الدراسات المعتمدة على المجاميع اللغوية (Corpus–based Studies).

تستخدم مناهج الدراسات المعتمدة على المجاميع اللغوية لدراسة مدى واسع من مختلف المواضيع في اللسانيات. و لكونها تحتوي كما كبيرا من النصوص، فان المجاميع اللغوية تمكن باحث اللسانيات من تحليل اللغة

ضمن السياق اللغوي (Contextualize)، و هي مناسبة جدا لمناقشة خواص اللغة في اللسانيات، من المنظور الوظيفي لمكوناتها.

سهلت تكنولوجيا المعلومات إمكانية خزن عدد كبير من النصوص في مجاميع لغوية، و إمكانية تحليل خصائص هذه النصوص، من الناحية اللغوية و اللسانية. اتفق باحثو اللسانيات على مختلف مذاهبهم على أهمية وجود المجاميع اللغوية كمورد هام في جداول أعمالهم.

بسبب أهمية المجاميع الوثائقية في دراسات اللغة واللسانيات فإن الحاجة للمجاميع الوثائقية تبرز في المعاجم (Lexicography) ، قواعد اللغة (Grammar) ، علم الدلالة (Semantics)، والمعالجة الطبيعية للغة (Natural Language Processing) (NLP) ،والدراسات اللغوية الأخرى.

الاستنتاجات

تشتت كم كبير من الجهد المنجز لتقنيات اللغة العربية ، بسبب غياب خارطة طريق واضحة واسعة الرؤية وانتابه كثير من الثغرات الجدية في جانبي بناء تقنيات اللغة العربية و هما:

1. البنى التحتية المعلوماتية Information Technology Infrastructure.
2. التطبيقات البرمجية Software Applications.

إن خارطة الطريق هذه، مطلوبة لصياغة أسس علمية للقرار على تمويل حقل البحث والتطوير .

إن مثل هذا الجهد يجب أن يبدأ بما يأتي :

1. بدراسة واقع حال تقنيات اللغة العربية
2. تحليل متطلبات موارد اللغة.
3. تحديد استيعاب التطوير وتحقيق التوافقية (Interoperability) في المكونات والمعايير.
4. القرار على تحديد وتطوير حد أدنى من أدوات معالجة اللغة وتوفير موارد اللغة .

المصادر العربية:

1. مصعب الخالد، "نظرة في تعريف الثقافة والحضارة والمدنية"، شبكة الألوكة، 2011/10/15.
2. محمد نور الدين جباب،"إشكالية الهوية والمغايرة في الفكر العربي المعاصر"، أطروحة لنيل درجة دكتوراه دولة في الفلسفة، جامعة الجزائر- كلية العلوم الإنسانية والاجتماعية، 2006.
3. د. أمان الدين محمد حتحات،"مراحل تطور الكتابة العربية حتى نهاية القرن الثاني الهجري"، جامعة الإمارات العربية المتحدة، وحدة المتطلبات الجامعية العامة، برنامج اللغة العربية، 2009.
4. أندريه بارو – سومر :فنونها وحضارتها، ترجمة وتعليق د . عيسى سلمان وسليم .طه التكريتي – وزارة الإعلام، بغداد1977
5. د .وليد الجادر،"صناعة التعدين، حضارة العراق"، الجزء الثاني، بغداد،1985.
6. د.عبد الحسن حسن خلف ،"الكتابة ومكانتها الحضارية عند العرب الدولة العباسية"، جامعة بغداد، مجلة كلية الاداب / العدد ٩٧، 2011.
7. احمد عزيز سلمان، م.م .عباس زويد الجبوري،"أختام منبسطة من موقع تل الولاية في المتحف العراقي"، مجلة مركز بابل للدراسات الانسانية،2013.
8. . محمد طه الحاجري، "مجموع رسائل الجاحظ"، دار النهضة للطباعة، بيروت 1982.
9. أبو الفرج محمد بن إسحاق بن محمد الوراق النديم ، "الفهرست" ، تحقيق د.أيمن فؤاد سيد ، مؤسسة الفرقان للتراث الاسلامي،لندن، 2009.
10. ايون أون كيونغ، "أفضل منهج لتعليم اللغة العربية لغير الناطقين بها من وجهات نظر علم اللغة الاجتماعي". مجلة الأستاذ العدد (201)، 2012.

المصادر الاجنبية:

11. Olson D. R.،" The Sage Handbook of Writing Development: The History of Writing" ، Sage Publications (CA) ،2009.
12. M. L. West،"The East Face of Helicon : West Asiatic Elements in Greek Poetry and Myth"، Clarendon press،oxford، 1997.
13. Peter Damerow،" The Origins of Writing as a Problem of Historical Epistemology"، Cuneiform Digital Library Journal 2006:1.
14. Jay David Bolter،"Writing space: Computers، hypertext، and the remediation of print"، Lawrence Eralbum Associates Inc.،2001.
15. Kabbani، R. "Selecting most efficient Arabic OCR features extraction methods using Key Performance Indicators"، Dept. of Artificial Intell.، Damascus Univ.، Damascus، Syria
16. Liana M. Lorigo"Offline Arabic handwriting recognition: a survey "،IEEE Computer Society، Volume:28 Issue:5.
17. Omar Al-Jarrah.، Samer Al-Kiswany. B. Mohamed Fraiwan. Hani Khasawneh،"A New Algorithm for Arabic Optical Character Recognition"،AIKED'06 Proceedings of the 5th WSEAS International Conference on Artificial Intelligence، Knowledge Engineering and Data Bases،2006.