المحاضرة الثالثة

التحليل الصرفي للغة العربية باستخدام الحاسوب

الأستاذ مأمون الحطاب
المدير العام لدار حوسبة النص العربي
الأستاذ حسان عبدالمنان
دار حوسبة النص العربي

السبت 15 محرم 1417هـ - 1 حزيران 1996م

ملخص

تعرض هذه الورقة تجربة دار حوسبة النص العربي لبناء محلل صرفي باستخدام الحاسوب من خلال حوسبة القواعد التي يبنى عليها التحليل الصرفي، دون الحاجة إلى استخدام المعجم.

لقد جرت محاولات عدة حتى الآن لتوظيف الحاسوب في خدمة اللغة العربية، خاصة في مجال الصرف، وقامت شركات حاسوب وأفراد وجامعات ببناء محللات صرفية ودراسات حول الموضوع، وتعتمد البرامج الموجودة حالياً على المعجم. حيث تم إدخال مادة المعاجم كلياً أو جزئياً للرجوع إليها عند تحديد جذر الكلمة واشتقاقاتها.

وعملنا هذا يعتمد على نظريتنا أن في اللغة العربية ما نسميه "أنساقاً داخليةً" تحكم تركيب الكلمة من حيث توزيع الحروف واستخدامها، وهدفنا أن نصل إلى القواعد التي تحكم هذه الأنساق، ثم بناء أسسٍ رياضية يمكن حوسبتها واعتمادها لمعالجة "التحليل الصرفي".

وقد أمكن وضع قواعد لكل وزن صرفي على حدة، بحيث تتم عملية التحليل الصرفي دون استشارة المعجم، لكننا لم نجد اضطراداً لقاعدة معينة يمكن أن تنطبق على الأوزان الصرفية بشكل عام.

ومع وجود القواعد المذكورة للقيام بالتحليل الصرفي آلياً فإنه لا غنى عن استشارة المعجم في النهاية وذلك لأمرين لم نتمكن من حوسبتهما وهما: تمييز العربي من غير العربي، وتحديد المستخدم من غير المستخدم من صيغ تصريف المادة آلياً.

تعريف المحلل الصرفي

يقوم علم الصرف على هيكلية الكلمة، وقد دل تصرف هذه الهيكلية ذات المعنى الواحد، أو المعاني القريبة في أمثلة كثيرة، على أن الكلمة تبقى ذات أصل واحد لا يتغير ضمن اشتقاقاتها المختلفة، فعلم أن أصل الكلام المتسق تحت نمط واحد من التفكير والطريقة يعود إلى أحرف، تبنى منها جملة حروف تكون

كلمات.

والمقصود بالمحلل الصرفي أنه برنامج حاسوب يقوم بعدة وظائف آلياً، لمعرفة تداخل المباني والمعاني، وأبرز هذه الوظائف:

١ -يحلل الكلمة إلى الجذر الثلاثي أو الرباعي أو الخماسي، مبيناً الأصل الذي عليه الكلمة، ومبيناً الزيادة التي أحدثت في هذه الكلمة.
٢ -يبين ما زيد على الكلمة من تأثيرات خارجة عن الكلمة نفسها، وتطرأ هذه الزيادة كما هو معروف في بداية الكلمة أو نهايتها، وتكون أحرفاً عارضة أو ضمائر متصلة.
٣ -يبين الوزن الذي عليه الكلمة، أمجردة كانت أم مزيدة.
٤ -يتمكن من صياغة أمثلة كثيرة على وزن معين من الأصل المعروض عليه، وذلك لأمرين: تعليمي، أو صياغة دلالية في النص.
٥ -يرجع الحروف المنقلبة إلى أصولها، مدركاً مواضع ذلك.

فوائد هذا البرنامج ومجالاته

ولتتضح أهمية هذا البرنامج، نذكر بعض المجالات المستفادة منه، وهي على اختصار كالآتي:

1- خدمة المعاجم وكتب النحو والصرف

أ- من حيث الأصالة:

دور المحلل الصرفي في تقريب هذه الفائدة أنه يمكننا من فهرسة الألفاظ من النصوص الكثيرة، وفهرسة أوزانها وأساليبها، مما يزيد المعاجم غنى وقوة.

وهذا يقرب لنا المسائل والدلائل في طرق التأصيل عند المتقدمين، وعند المدارس المتبناة في عصورهم، ويقرب لنا المعاجم المتأخرة الخالية من الاجتهاد إلا ضمن ما كتب ممن سبق.

ب - من حيث طرق التأصيل:

ليس من شك أن طرق التأصيل عند المتقدمين في عرض المعلومات، والتعليل، تختلف باختلاف المدارس المتبناة عندهم أولاً، وبما يمتلكه كل منهم من قدرات التعليل العقلية.

فواضح مثلاً أن التعليلات التي أوردها سيبويه في كتابه في مبحث أوزان اللغة، لا علاقة لها بالمادة اللغوية، وإنما هي حس لتقريب المسائل، قَلَّ من يؤتاه، لذا تجد المتأخرين – وهم ذوو شأن – عجزوا أن يحسنوا إبداعاً كإبداعه، وعولوا في كتبهم على حروفه، وقَلَّ من انتقد منها شيئاً.

والمحلل الصرفي قد يكون حكماً على مثل هذه التعليلات، إذا أوتي من صحيح النصوص ما يبني به قاعدة وأساساً يحاكم به المادة الأولى التي بنى عليها سيبويه أو غيره تعليلاً.

جـ– من حيث الاختلافات:

وجدنا أن تحليل الكلمات وبيان الأصول، وكيفية فلسفتها قد اختلفت عندهم، فوضع الهمزة في حروف العلة عند الخليل بن أحمد لم يرق لآخرين، ففصلوها ، ووضع الألف في باب الألف أحياناً عند ابن فارس لم يُرض، بل جعل في حرفي (الواو والياء) على ما قدر من انقلابها وإعلالها.. وهكذا.

ثم وجدنا أن أوزان الكلمات في المعاجم وكتب النحو والصرف قد اختلف في قسم كبير منها، لاختلافهم في أصول هذا العلم، والسبب الذي من أجله يحكم فيه على أصالة الحرف أو زيادته ، وأصل هذا العلم اجتهادي، لا علاقة له بأصل اللغة إلا ما ألجأنا إليه التبويب والترتيب ومعرفة بعض الصيغ التي لها استخدام نحوي.

فوضع لهذا كله منطق من الاجتهاد لترتيب الكلمات والمواد، وأكثر قواعد هذا المنطق كان شكلياً، وقلبياً أحياناً، إذ القواعد التي وضعوا تختلف باختلاف في التطبيق، أو اختلاف في القواعد نفسها، أو اختلاف بين القاعدة ومن لم تكن عنده بعض هذه القواعد... لكنها في النهاية تنصب لمعرفة الأصل والزائد من الحروف في ثلاثة أمور: (الاشتقاق، عدم النظير، كثرة زيادة الحرف في ذلك الموضع المخصوص).

إذ أصل فلسفة الأوزان قام على أن أصل الكلمة ثلاثة أحرف ما لم يعرض لها عارض يمنع، وفكرة الثلاثة جاءت من الغالب في الكلام ، فما زاد على الثلاثة الأحرف ينظر فيه، فإن حذف في اشتقاقاته، علم بزيادته، فإن لم تكن له اشتقاقات يعلم منها، هذا وكان الحرف الزائد على الثلاثة مما يزاد في العادة، وعلمت زيادته في كلمات كثيرة مبنية على القاعدة الأولى، حكم عليها بالزيادة.

وكذا تصرفوا في الكلمات التي فيها أكثر من لغة، في إحداها يظهر الزائد، وفي الأخرى يظهر في موضع الأصل، فحملوا هذا على الزائد، وأخرجوه من ظاهره إلى أنه زائد ، وكذا حكموا في الكلمات التي إن حكموا على بعض أحرفها بالأصالة، لم يكن لهذه الكلمات نظير، وإن حكموا على هذه الأحرف بالزيادة كان لها نظير، فيحملون الأوزان على ما كان لها نظير... وهكذا في سلسلة من المنطق الصرفي الذي وضع لحل مشكلة الترتيب والتبويب.

ومن الطبيعي أن يكون بين المنظرين خلاف، لا سيما أن هذه القواعد موضوعة وغير منظورة، أي: لا تعتمد على كثرة الشواهد، والأدلة الحسية، اعتمادها على البرهنة العقلية.

لذا وجدنا أن الخلاف الواقع بينهم أدى إلى اختلاف حكم المادة بين كثير من المعاجم من جهة، وبين المعاجم وكتب النحو من جهة أخرى، ووجود المدارس المختلفة آنذاك له دور آخر أيضاً في زيادة الخلاف.

وهناك نماذج كثيرة من الخلاف بين اللغويين والنحويين في زيادة النون وأصالتها من الكلمة الرباعية أو الخماسية... تظهر مثلاً أن ما حكم بزيادته سيبويه، حكم بأصالته في موضع آخر من كتابه، وما حكم بأصالته، حكم الصاغاني والفارابي وغيرهما بزيادته ، وما حكم سيبويه بزيادته حكم الصاغاني والفيروزآبادي وغيرهما بأصالته ، وما حكم سيبويه والصاغاني وابن عصفور وغيرهم بأصالته، حكم الفارابي والفيومي بزيادته... إلى كثير من الأحكام المختلفة بينهم.

فمثل هذه الخلافات يجعلنا المحلل الصرفي في حل منها، إذ إنه يقضي عليها بالتزام القاعدة والترتيب الواحد، دون أن يكون هناك تناقض في مواقفه.

ويمكن له أن يدخل على الكلمة كيفما كانت وعدت عندهم، فاختلافهم في تعيين زائد أو أصل لا يعيقه عن أداء عمله واستخراج الفائدة التي من أجلها استحدث.

2- مجال الموسوعات

وجود المحلل الصرفي عامل مهم في إنتاج كثير من تطلعات العاملين نحو الموسوعات، لأنه بسهولة يمكن أن يرتب أجزاء عشرات الآلاف من الصفحات ضمن الفائدة الموجهة فيه.

ولنضرب على هذا مثالاً بموسوعة الحديث النبوي الشريف:

نقول: لو أردنا أن نعمل في هذا المجال يدوياً لكلفنا من العمل عشرات من السنوات، فهذا فريق من المستشرقين يتزعمهم (ونسنك) قدم" المعجم المفهرس لألفاظ الحديث النبوي" لتسعة كتب فقط، هي الكتب الستة، وسنن الدارمي، وموطأ مالك، ومسند أحمد، عكف المستشرقون في صناعته أكثر من نصف قرن، بدأه (ونسنك) فكرة سنة (1916م) أو قبلها، وعزم على وضعه سنة (1922م) وصدر المجلد الأول منه سنة (1936م) ووافته المنية سنة (1939م) فأكمل الفريق العمل بين انقطاع واتصال حتى صدر المجلد الثامن – وهو الأخير – سنة (1987م).

وقد كلف هذا العمل الجهد والوقت مع توفر الإمكانات المادية والعلمية له، إذ ساعدت على عمله ودعمه المجامع العلمية البريطانية، والدنمركية، والسويدية، والهولندية، واليونسكو وغيرها.

ومع هذا كله، ومع توفر كل ما يلزم لم يستطع الفريق النجاح في هذا المشروع أكثر من ستين بالمئة، فقد وقع في عملهم سقط وخلل واضطراب...

وبقي هذا المعجم إلى الآن يحتل الصدارة لم ينازعه في عالم المطبوعات عمل أفضل منه أو مثله، فكيف يكون الحال مع وجود هذا المحلل الصرفي؟ فإنه سيوفر موسوعة حديثية ضخمة لكل ما ورد، وما طبع في علم الحديث... وفي وقت قصير جداً.

إذن يمكن لنا أن نستفيد من "برنامج التحليل الصرفي" في موسوعات كثيرة على هذا النمط، وبذا نوفر على أنفسنا أوقاتاً كثيرة من العمل والجهد المتواصلين.

3- مجال الفهرسة في غير الموسوعات

من المعلوم أن كثيراً من الكتب والدراسات لا يصلح أكثر ما يكون إلا بخدمته بفهارس فنية تغنيه من حيث القيمة العلمية، ويتضح هذا في كتب الشعر، والحديث، والكتب العلمية التي يذكر فيها كثير من المصطلحات، ومتون النثر التراثية من خطب ونحوها...

4- مجال استرجاع المعلومات المترابطة المعاني من كتب التراث

هناك كثير من الكتب الضخمة في الخزانة العربية الإسلامية، وهي مليئة

بالمعلومات التي قد لا تكون مفهرسة، ولم يتم فهرستها، ويدخل في هذا كتب الأدب، والتفسير، وشروح الحديث، والفقه، والتاريخ، والتراجم، والطب، والفلك، والعلوم الأخرى...

فإذا كان شيء منها مخزناً في (الحاسوب)، استطعنا أن نستدعي المعلومة التي نريد بذكر ما يمكن أو يحتمل أن يكون فيها من الكلمات، ولا شك أن الكلمات كثيرة الاشتقاق، لذا قد تتغير بتغير الجمل ، فهذا البرنامج (المحلل الصرفي)، سيساعدنا في الوصول إلى الكلمة وما تحتمل من اشتقاقات، ومن ثم نصل إلى المراد في ذاك الكتاب.. ويقاس على هذا الأمر الجزئي ما يتصل من أمور كلية وعظيمة.

5- مجال التعليم

نحن نرى أن (الحاسوب)، قد غزا دور التعليم العليا والدنيا، فدخل المدارس بشتى أنواعها، ودخل الجامعات والكليات والمؤسسات بشتى اتجاهاتها، وصار جزءاً مهماً في عملية التعلم، ومن ذلك أنه يعرض توضيح المشتقات والمزيدات على البرنامج للوصول إلى الأصول، وأوزان الكلمات المعروضة... فكثرة تردد الأمثلة وتوضيحها وبيانها عن طريق (الحاسوب) يزيد الطالب فهماً للمادة، لا سيما أنها المادة التي ينبني عليها علم الصرف كله.. كما يساعد الطالب أيضاً على التمرين على بابي الإعلال والإبدال، اللذين يجد الطلاب فيهما صعوبة كبيرة، بل لا يكاد يتقنهما كثير من الأساتذة.. فهذا البرنامج يساعد الطالب على معرفة أصل الأحرف المنقلبة وأوزانها، ومن ثم يتبين الأسباب التي من أجلها حولت هذه الأحرف عن طبيعتها.

6- مجال مشاريع وبرامج الحاسوب التي تقوم على المحلل الصرفي

يعد برنامج (المحلل الصرفي)، مفتاحاً لكثير من المشاريع التي لا يمكن أن تحيا من دون التحليل، لأنه أصل في بنائها، أو لأن له دوراً جزئياً فيها ، واستقصاء ذلك يطول، نقتصر فيه على ما يلي:

أ- برنامج ضبط النص وتشكيله:

فلا يمكن الوصول إلى كثير من قواعد ضبط النص، لا سيما ما يتعلق منها

بعلم الصرف المتصل مباشرة بالنحو، دون التحليل الصرفي الذي سيزودنا بكيفية معالجة بعض الأفعال، وعمل بعض المشتقات، والزيادات العاملة وغير العاملة المتعلقة بالكلمة... مما يعين في ضبط الكلمة أو ما يحيط بها من كلمات في السياق.

ب- برنامج ترجمة النصوص:

برأينا لا بد من وجود ثلاثة أركان لقيام مترجم آلي هي: التحليل الصرفي، والتحليل النحوي، والتحليل الدلالي، ووجود هذه الأركان بهذا الترتيب لاعتماد كل منها على ما سبقه، وتحديد المعنى المراد من الكلمة المجردة.

فبما سبق يتبين لنا أن الترجمة الآلية تقترب خطوة إذا تيسر (برنامج التحليل الصرفي)، بشكل جيد، لأن الركن الأول، يكون قد وجد.

جـ- برنامج المصحح الإملائي

إذ يخطئ كثيرون في صياغة بعض الأفعال إذا زيد عليها بعض الزيادات المعروفة، خاصة إذا كان في الكلمة حرف علة أو همزة، فإن صياغتها بإثبات الزوائد في أولها أو آخرها أو وسطها قد يوقع بالخطأ، لا سيما إذا كانت الكلمة غير مألوفة.

كما يخطئ آخرون في تحديد عين الفعل المضارع ولامه إذا كانتا حرفي علة، أو إحداهما، فقد يقلبون حرفاً بآخر.

كما قد يخطئ جمع في كلمات غير قليلة من بابي الإبدال والإعلال.

ففي هذا وغيره يمكن لبرنامج التحليل الصرفي أن يكشف عن الخطأ، ويساعد في نجاح برنامج المصحح الإملائي، ويكون رافداً له.

د- مجال المسح الضوئي للحروف والتعرف الآلي عليها:

نحن نعلم منذ فترة من الزمن أن بعض العاملين في مجال الحاسوب حاولوا أن ينشئوا برنامجاً لإدخال المعلومات إلى جهاز الحاسوب عن طريق المسح الضوئي، لا الإدخال عن طريق لوحة المفاتيح (الصف)... إلا أن المساعي إلى الآن لم تنجح إلى درجة مرضية، وحددت أسباب ذلك في نقاط عدة.

وكان من ذلك أن الحاسوب قد لا يتعرف على بعض أحرف الكلمة تماماً،

وإنما يتعرف إليها على التشابه (التوهم) ، فقد تتحرف وقد تتصحف لتشابه أشكال الحروف، من حرف إلى آخر ، ومن ثم فقد تختلف الصيغ عنده وتتحول إلى صيغ أخرى.

ومن هنا نجد أهمية لبرنامج التحليل الصرفي، إذ يساعد الحاسوب على القراءة الصحيحة للكلمة من خلال التأكد بأن هذه الكلمة التي قرأها تخضع لقواعد التحليل الصرفي، أو لا تخضع، فإن خضعت فالكلمة أقرب إلى الصحة، وإلا تحول إلى احتمال آخر في تكوين حروف الكلمة يكون أصوب في القراءة.

منهجنا في بناء المحلل الصرفي

إن أول خطوة في هذا العمل هي تحليل بنية الكلمة بتفصيلها آلياً إلى مقاطع، ثم إمرار هذه المقاطع على قواعد حاسوبية تحولها إلى معادلات رياضية من خلالها يمكن للحاسوب أن يتعرف على جميع الزوائد في أول الكلمة ووسطها ونهايتها، ويمكن له أن يتعرف على الأوزان، ويمكن له أن يحلل الأحرف مبيناً ما انقلب من حرف آخر ضمن علل معينة.

وقد واجهتنا المشكلات المتوقعة التي قد تقف أمام هذا البرنامج، كوضع الهمزة، وعلامات الإعراب، وشذوذ التصريف، وإعلال الحروف، والزيادات الطارئة، وتداخل الأوزان... وغيرها ، لكن تم تجاوز هذه المشكلات بوساطة مجموعة القواعد التي تم بناؤها حاسوبياً لإتمام هذا العمل وهي كما يلي:

1- قواعد أوزان اللغة العربية

وقد تم بناؤها من خلال دراسة كل وزن على حدة، والكلمات التي تنتمي إليه، وما يدخل عليها من تأثيرات خارجية، وأثر كل مؤثر خارجي ، ثم بناء قاعدة (Rule) أو مجموعة قواعد لهذا الوزن، وبعد ذلك تحويل هذه القاعدة إلى صيغة رياضية يمكن للحاسوب التعامل معها.

2- قواعد تحليل السياق

وهذا جزء من التحليل الدلالي للنص، وقد برزت الحاجة لهذه القواعد نتيجة اللبس الذي قد يعرض للتمييز بين الاسم والفعل مثلاً.

3- قوائم الشذوذ

وقد جرى حصر ما شذ لجميع الأوزان العربية بالاستعانة بالمعاجم المتوفرة.

5 – قواعد معرفة مواضع الإعلال والإبدال في الكلمات

وجاءت هذه كنتيجة لبناء قواعد الأوزان لكننا فصلناها في بند مستقل لأهميتها.

6 – أما بالنسبة لاختلاف النحويين واللغويين في الأوزان وما سبق من ذكر لهذه الإشكالات، فقد وافقت قواعدنا آراء بعضهم، وهي في اعتقادنا أنظمها وأدقها لموافقتها النسق القاعدة في المحلل الصرفي.

وقد قمنا ببناء نظام خبير يحتوي هذه القواعد كلها، إضافة إلى قوائم تشمل الأدوات والحروف وقوائم الكلمات الشاذة باستخدام لغة **C++**، وذلك لتجاوز بطء استخدام لغات الذكاء الاصطناعي مثل LISP.

تطوير هذا البرنامج

يمكن تطوير هذا البرنامج للتيسير على الباحثين، ولا سيما العاملين في مجال كتب التراث، أو الرسائل العلمية، أو الكتب التي يعتني أصحابها ببناء الفهارس وتقديم المعلومات وذلك من خلال ما يلي:

أ. الفهرسة التلقائية للأعلام المذكورين في الكتاب، دون أي تدخل من المدخل بالترميز لها، بل يكفي أن يطلب الأعلام من خلال هذا البرنامج، فيتعرف عليها ويخرجها من الكتاب، ويفهرسها تلقائياً حسب ورودها، دون أن يرجعها إلى الأصل الصرفي.

ب. وكذا يمكن فهرسة البلدان فهرسة تلقائية، مع إمكانية التفرقة بين البلد وما قد يستخدم في معنى آخر، معتمداً في هذا على السياق في نسبة كبيرة، أي: التفريق بين مدلولات الكلمات من خلال عدد من القواعد.

جـ. الفهرسة التلقائية لجميع الشعر الوارد في الكتاب من بداية البيت أو قافيته، وفرزها مرتبة على الأحرف الهجائية من أي الجهتين أردت.

د. استخلاص الآيات المذكورة في الكتاب، وفهرستها تلقائياً على حسب السور والآيات من السورة، دون أن يستدل عليها بأن تبحث له عنها، أو تعرفه بها.

هـ. وكذا وبنسبة كبيرة جداً من النجاح فهرسة الأحاديث الواردة في الكتاب مستدلاً عليها من السياق، وذلك تلقائيا.