

نحو معالجة الدلالة في اللغة العربية عبر قواعد البيانات دراسة أولية لنص القرآن الكريم

أ. د. محمد زكي خضر

الجامعة الأردنية

المستخلص تحتاج المعالجة الدلالية الآلية للغة العربية مقداراً كبيراً من المعلومات عن مختلف جوانب اللغة. وأفضل ترتيب لهذه المعلومات هو قواعد البيانات. تشمل هذه المعلومات قواعد البيانات المتعلقة بنسق الكتابة وقاعدة بيانات الصرف وما يتعلق بها من فصل لنواة الكلمة عن لواصقها السابقة واللاحقة وربط النواة مع الأوزان الصرفية المعروفة. يلي ذلك قواعد بيانات النحو التي تحتاج إلى تقسيم اللواصق إلى مكوناتها والأخذ بعين الاعتبار الضمائر المستترة وتقديرها ، وذلك من خلال شبكة موسعة لقواعد البيانات الارتباطية. كما أن قواعد بيانات اللفظ والنطق السليم ضرورية للغة العربية لاستكمال المعالجات المناسبة لها. كل تلك المعالجات وقواعد البيانات تصب في قواعد البيانات المتعلقة بالمعاني وترتبط معها بشكل تشابكي محكم. يوضح البحث كيف يمكن أن يتم ذلك بالاستناد إلى نص القرآن الكريم ومن ثم التوطئة نحو الاقتراب من الدلالة الدقيقة في اللغة العربية.

١ - المقدمة

تحتل الدلالة ضمن المعالجة الآلية للغة العربية أهمية كبيرة، فلا يمكن أن تتم معالجات عميقة للنصوص العربية بدون معلومات كافية عن دلالة الألفاظ المكونة لتلك النصوص. تحتاج المعالجة الدلالية الآلية للغة العربية مقداراً كبيراً من المعلومات عن مختلف جوانب اللغة. وهذه المعلومات يجب ان تكون مرتبة ومبوبة بنسق معين. وليس هناك أفضل من ترتيب هذه المعلومات من قواعد البيانات. وهذه المعلومات تشمل المعلومات عن الألفاظ وما تحوي من دلالات مختلفة وصرف وما يجوز أن تستعمل معه الكلمة وما لا يجوز والكلمات المقاربة والمضادة وكل ما يوصل إلى دقة دلالة الكلمة من وسائل.

إن أية بداية في هذا الصدد لا بد وأن تستند إلى نصوص رصينة ولذلك فإن دراسة مفردات القرآن الكريم هي خير بداية لمثل هذا الجهد ، وسناقش في هذا البحث قواعد بيانات القرآن الكريم وكيف يمكن أن تساعد في المعالجة الدلالية للغة العربية.

سنعرض أولاً إلى قواعد البيانات المتعلقة بنسق الكتابة بالرسم العثماني والرسم الإملائي وما يحتاج ذلك من معالجات مسبقة للنص القرآني. ومن ثم يتم تقسيم النص القرآني إلى جمل مفيدة بصرف النظر عن اتفاق ذلك مع الفواصل بين الآيات من عدمه.

بعد ذلك نتعرض لقاعدة بيانات الصرف وما يتعلق بها من فصل لنواة الكلمة عن لواصقها السابقة واللاحقة وربط النواة مع الأوزان الصرفية المعروفة. يلي ذلك قواعد بيانات النحو الذي يحتاج إلى تقسيم اللواصق إلى مكوناتها والأخذ بعين الاعتبار الضمائر المستترة وتقديرها ، وذلك من خلال شبكة موسعة لقواعد البيانات الارتباطية. كما أن قواعد بيانات اللفظ والنطق السليم ضرورية للغة العربية (بالنسبة للقرآن التجويد والترتيل) لاستكمال المعالجات المناسبة للغة العربية.

كل تلك المعالجات وقواعد البيانات تصب في قواعد البيانات المتعلقة بالمعاني وترتبط معها بشكل تشابكي محكم.

إن معالجة اللغة العربية حاسوبياً اليوم بحاجة إلى مكنز (lexicon) متكامل لتحديد المعاني المستعملة في اللغة الحديثة وفي الأزمنة المتعاقبة التي مرت بها اللغة العربية وتبويب ذلك بشكل منظم. لقد كان التطور هائلاً في أبحاث معالجة اللغات الطبيعية وخاصة الانكليزية خلال العقود الماضية ، وكان لاستخدام قواعد البيانات أثراً كبيراً في هذا التقدم[١]. وقد تكونت مكانز هائلة للغات حديثة متعددة. أنظر مثلاً المواقع على الشبكة المعلوماتية العالمية[٢]. ولذلك هناك حاجة لجهد هائل للغة العربية.

٢ - الرسم الإملائي للكتابة العربية

يختلف التعامل مع النصوص باللغة العربية والتعامل مع نص القرآن في عدد من الأمور يمكن إيجازها على النحو الآتي :

أ- لا يورد النص القرآني غالباً إلا مشكولاً . والتشكيل يسهل التعامل مع النص من نواحي متعددة حيث يذهب الالتباس والخطأ في المعالجات الآلية الصرفية والنحوية والدلالية . هذا مع العلم أن الكتابة العربية كانت تكتب قبل الإسلام وفي صدر الإسلام غير منقوطة وغير مشكولة ولم يضاف التنقيط إلا في وقت متأخر عن ذلك.

ب- يختلف النص القرآني عن النصوص العربية في حسن الالتزام بالرسم العثماني فجمهور الفقهاء يحبون (بل بعضهم يعتبر ذلك ملزماً) كتابة المصحف بالرسم العثماني وقد أجاز بعض العلماء كتابة المصحف برسم آخر غير الرسم العثماني إن كان ذلك يحقق دقة في قراءة ولفظ القرآن في

عصر من العصور أفضل مما يحقق الرسم العثماني بسبب أن الناس أصبحوا معتادين على رسم خاص لبعض الحروف.

ج- لغرض دقة قراءة القرآن الكريم ولفظه بشكل صحيح ، تحتوي المصاحف على علامات للوقف والتجويد وهي خاصة بالقرآن الكريم ولا تستعمل في أية كتابة عربية أخرى .

د- يحتوي النص القرآني على فواصل بين آية وأخرى . أي أن بداية ونهاية الآيات محددة (رغم وجود خلاف بين بعض المتخصصين برسم المصاحف حول مواقع بعض هذه الفواصل) . إن هذه الفواصل لا تقع في نهاية جمل مفيدة كاملة دائماً بل تقع أحياناً في وسط جملة أي أن الآية قد تكون جزءاً من جملة أو قد تكون جملة كاملة أو قد تحوي جملة وبعض جملة أو أنها قد تحتوي على عدد من الجمل .

هـ- إن الخطأ في التعامل مع النصوص غير القرآنية قد لا ينطوي عليه اعتراض يذكر أما الخطأ في التعامل مع القرآن الكريم فيلحق اعتراضاً دينياً واضحاً ، لذلك فإن مقدار الحرص على دقة التعامل مع النص القرآني يجب أن يكون فائقاً .

٣- إعداد النص القرآني

قد يكون من المقبول في التعامل مع النصوص فيما عدا القرآن الكريم وجود نسبة بسيطة من الخطأ الإملائي أو الخطأ في التشكيل (إن كان النص مشكولاً) وقد تبلغ النسبة المقبولة من الدقة ٩٩,٩% ولكن ذلك غير مقبول على الإطلاق في نص القرآن الكريم . فهذه النسبة من الخطأ تعني وجود خطأ مقداره واحداً بالألف . وحيث إن النص القرآني المشكول يقرب حجمه من ٧٠٠ ألف بايت فوجود خطأ مقداره واحد بالألف يعني ٧٠٠ خطأ في القرآن كله وهذا ما لا يقبله أحد من المسلمين . لذلك كانت المهمة الأولى هي التأكد من النص القرآني بنسبة تقرب من ١٠٠% تماماً. وقد اعتمد الرسم القرآني برواية حفص عن عاصم المشهورة في المشرق العربي والإسلامي. كما اعتمد في قواعد البيانات هذه اعتبار البسملة أول آية من القرآن الكريم (أي أول آية من سورة الفاتحة) دون بقية السور [٣] و[٤] . وقد اعتمد النص القرآني آية آية . أن كل آية تعتبر قيداً مستقلاً (record) أو سطرًا كاملاً . ولا يفصل بين كلمتين من كلمات الآية سوى فراغ واحد . أما بين آية وأخرى فلا داعي لأية فاصلة لأن الانتقال من سطر لآخر إشارة إلى انتهاء الآية والبداية بآية جديدة وقد أدى ذلك إلى انتقاء الحاجة إلى ترقيم الآيات الواحدة تلو الأخرى أو وضع الترقيم بين آيتين . فالمعالجة الحاسوبية عند البرمجة يمكن أن تقوم بالعد والترقيم دون الحاجة إلى إثبات ذلك في النص . وعلى هذا فقد احتوى النص على (٦٢٣٦) سطر أو قيد .

أما الفواصل بين سورة وأخرى فكان لا بد من اختيار رمز يشير إلى ابتداء سورة جديدة . وقد اختيرت العلامة # كرمز لبداية سورة جديدة .

أما علامات الوقف والتجويد فقد اختير لكل علامة رمز من الرموز المستعملة في الحاسوب مثل \$ & % + - وغيرها . كما اختيرت علامة خاصة لمواقع سجود التلاوة . وقد لوحظ أن بعض علامات الوقف والتجويد تقع بين كلمتين من كلمات المصحف وبعضها يقع في آخر الكلمة بينما وقع البعض الآخر في وسط الكلمة . وهذا جعل الكلمات بعضها غريباً في شكله نظراً لأن الحروف تختلف في هيئتها إن وقعت في وسط الكلمة أو أولها أو آخرها إذا ما أدخل وسط الكلمة رمز من هذه الرموز . وعلى أية حال سوف يتبين فيما بعد أن المعالجة الأخيرة كانت بحذف كل هذه الرموز عند استخلاص الكلمات كلمة كلمة وذلك لسببين : السبب الأول صعوبة استخدام هذه العلامات في برامج معالجة النصوص الشائعة حيث يستدعي ذلك استخدام فونط (ترميز للحروف) خاص بالمصحف لا يتوفر عند الغالبية العظمى من مستخدمي الحاسوب والسبب الثاني والأهم أن الكلمة القرآنية غالباً تتطوق بمعزل عن علامات الوقف والتجويد (إلا إذا كانت العلامة وسط الكلمة وهي نادرة) والتي غالباً ما تكون بين الكلمات المتعاقبة أو نهاية الجمل أو الآيات . أما الأمر المهم الآخر في هذا المضمار فهو الألف الخنجرية التي توضع فوق بعض الحروف ولا تكتب ألقاً مثل "الرحمن" و "هذا" . وقد تم استخدام علامة التعجب ! لتتوب محل هذه الألف . وسيتبين فيما بعد أن هذه الألف كذلك قد تم الاستعاضة عنها بالألف الممدودة تارةً في مثل كلمة "العالمين" وبالفتح تارةً أخرى في مثل كلمة "الرحمن" و "هذا" وذلك تماشياً مع الخط العربي الشائع اليوم (الذي سندعوه بالرسم الإملائي) بكتابة هذه الكلمات خلافاً للقاعدة بدون ألف ممدودة . وكان هذا الحذف لضرورات الطباعة باستعمال معالج نصوص عادي . هذا مع الإشارة أن هذه المعالجة لم تتم في هذه المرحلة ولكن في مرحلة لاحقة كما سيتبين في ما بعد . كما استعملت علامة الاستفهام ؟ بدل همزة الوصل.

وعلى هذا فإن النص القرآني في هذه المرحلة كان نصاً استبعدت منه علامات الوقف والتجويد ومواقع سجود التلاوة ولكن أقيمت فيه علامات الألف الخنجرية . أي أن النص في هذه المرحلة هو نص بالرسم العثماني خال من العلامات (عدا الألف الخنجرية وهمزة الوصل) وليس نصاً مطابقاً للنطق ، حيث إن هناك عدداً من الكلمات القرآنية التي لا يمكن نطقها بشكل صحيح إلا بمساعدة العلامات المثبتة في المصاحف حالياً والتي استثنيت في هذه المرحلة من المعالجة . وقد حذفت العلامات المشار إليها باستخدام أحد برامج التحرير (edit) .

وعلى هذا فإن أول قاعدة بيانات هي قاعدة بيانات النص القرآني بالرسم العثماني المحتوي على علامات الوقف والتجويد والذي كل قيد فيه هو آية مستقلة ويذكر أن أطول آية في المصحف تبلغ حوالي ١١٠٠ بايت (حروف وتشكيل وعلامات)

٤ - قواعد بيانات رسم الكلمات

كان لا بد من البدء بتكوين قاعدة بيانات الكلمات بالرسم القرآني وفق الرسم العثماني . فالكلمات الأساس في قاعدة البيانات هذه هي حقل رسم الكلمة بالرسم العثماني ، محذوفاً منه علامات الوقف والتجويد فقط .

ويلاحظ أن هناك كلمات في الرسم العثماني مكتوبة في أماكن مختلفة في المصحف بأشكال مختلفة فقد تختلف فيما بينها في رسم التاء مرة كتاء مربوطة ومرة كتاء طويلة وكذلك في وجود ألف واو الجماعة مرة وعدم وجودها مرة أخرى وهكذا. أما الملاحظة الثانية فهي عن رسم الهمزة . فرسم الهمزة في المصحف يختلف عن الرسم الإملائي الحديث للهمزة في كثير من المواضع.

لا شك أن معظم الكلمات القرآنية متطابقة في كتابتها بين القديم والحديث . ولكن هناك عددًا لا يستهان به من الكلمات المختلفة في شكل كتابتها . ويمكن إجمال الاختلافات بما يأتي :

أ- استعمال الألف الخنجرية في كثير من الكلمات مثل العالمين (تكتب العلمين) .

ب- كتابة الهمزة بشكل مختلف عما هو متعارف عليه في الخط الحديث مثل ءالذكرين التي تكتب اليوم الذكرين باستعمال المدة فوق الألف (التي هي همزة بعدها فتحة طويلة (أي ألف) رغم أن المصاحف تستعمل المدة كعلامة من علامات التجويد).

ج - ألف واو الجماعة تكتب أحياناً بالألف وأحياناً بدون الألف. كما أن هناك أحياناً واو أصلية في الكلمة مضافاً لها ألف مثل ادعوا.

د - كتابة التاء تارة كتاء مربوطة وتارة كتاء طويلة .

هـ - كتابة بعض الكلمات بشكل خاص مثل الربا تكتب (الربوا) .

و - دمج بعض الكلمات أحياناً وعدم دمجها أحياناً أخرى مثل "أم من" (تكتب أحياناً أمَّن) و"أن لا" تكتب أحياناً (ألا) و"إن لم" (إلم) .

ز - دمج بعض الكلمات بطريقة مغايرة لما هو متعارف عليه اليوم مثل (يا ابن أُمي) كتبت (بينؤم) و (فما للذين) كتبت (فمال الذين) .

ح - الحروف في فواتح بعض السور التي تسمى بالنورانية غير متعارف عليها في الخط الحديث. ولكنها ربما تشبه استعمال الرموز في المعادلات الرياضية أو في الأشكال الهندسية أو تبويب الفقرات هجائياً.

وعلى هذا فإن أول خطوة في تكوين قاعدة بيانات رسم الكلمات هو إعادة كتابة الكلمات المختلف عليها بالرسم الإملائي فتتكون قاعدة البيانات هذه من حقلين الأول هو الرسم العثماني الدقيق والثاني هو الرسم الإملائي.

احتوت قاعدة بيانات القرآن الكريم الرئيسة على ٨٣٨٢٩ قيداً (Record) تحوي بينها ١١٤ علامة # لبداية السور و٦٢٣٦ علامة @ لبداية الآيات حيث أن عدد كلمات المصحف بالرسم العثماني هو ٧٧٤٧٩ فيها ٤ كلمات هي كلمات البسملة في بدء سورة الفاتحة. وعند استخلاص الكلمات القرآنية ووضعها في قاعدة بيانات ثانوية خاصة بالكلمات القرآنية فإن عددها سيكون ١٨٨٤١ كلمة. أما عند حذف تشكيل الآخر والشدة على أول حرف إن وجدت بسبب التتوين في الكلمة السابقة فعددها ينزل إلى ١٥٢٦٣.

وسّعت قاعدة البيانات الثانية بخطوات متعاقبة وذلك بحذف الشدة على الحرف الأول لبعض الكلمات نتيجة وقوعها بعد كلمة محتوية على تنوين مثل تشديد اللام في (هَدَى َ لِّلْمُتَّقِينَ) وتشديد الميم في (رِيبَ مِمَّا) وقد تأتي هذه الشدة حتى في بداية آية إن كان في نهاية الآية التي قبلها تنوين مثل قوله تعالى (عدواً مبيناً - رَبِّكُمْ أَعْلَم) وتتنصر هذه الحالات إن كانت بداية الكلمة لاماً أو ميماً أو راءاً أو نوناً أو ياء أو واواً وكانت الكلمة التي تسبقها منونة كما في الأمثلة السالفة . ولهذا حذفت هذه الشدة على مرحلتين المرحلة الأولى بواسطة برنامج يدقق إن كانت هناك شدة على الحرف الأول من كل كلمة تبتدئ بهذه الحروف فيقوم بفصلها والمرحلة الثانية كانت بأن تدقق يدوياً ومن ثم تحذف . وهناك مواضع أخرى للشدة التي كان يجب أن تحذف من بداية الكلمة مثل " لهم مآ يشأؤون " نتيجة التقاء ميمين في كلمتين متعاقبتين . كما عولجت مشكلة تشكيل اللام ألف . وتجدر الإشارة هنا إلى أن معالجات النصوص الشائعة تكتب اللام ألف بشكل غريب إن كانت اللام مشكولة أو عليها شدة (تكتب "لَا" أو "لَاَ" وليس "لا") ولذلك استعيض عن تشكيل اللام بتأخير الفتحة إلى ما بعد الألف بحيث تصبح "لَا" . وهذا خطأ كان لا بدّ من القبول به في هذه المرحلة . وبالطبع فإن المشكلة تزيد سوءاً إذا كانت الألف بعد اللام هي همزة مثل "لأنتم" حيث كتبت "لأنتم" أو "لإلى الله" حيث كتبت "لإلى الله" وقد حذفت الحركة على اللام في مثل هذه الحالات . وقد يلتقي اللام ألف والشدة على اللام نتيجة تنوين الكلمة السابقة لها وبذلك تجتمع المسألتان أعلاه في موضع واحد .

هناك حقل ثالث مهم يجب إضافته لقاعدة البيانات هذه ، ألا وهو الرسم بالخط العادي بدون تشكيل . فالكتابة الشائعة اليوم غير مشكولة بصورة عامة ولا شك أن ذلك يحدث كثيراً من اللبس في كتابة العديد من الكلمات وهذا الحقل ضروري لكي تؤخذ إمكانية التعامل مع هذا اللبس في أي معالجة آلية للغة العربية ذات معنى .

تكوين قاعدة البيانات هذه قد مرّ بمرحلتين: مرحلة آلية وذلك باستحداث قاعدة بيانات للكلمات المتشابهة ثم يجري تعديلها إلى الرسم العادي يدوياً وفي مرحلة لاحقة يحذف منها التشكيل وتعاد هذه الحقول إلى قاعدة البيانات الرئيسة لتتكون قاعدة بيانات مكونة من نص بالرسم العثماني وآخر بالرسم الإملائي وثالث بدون تشكيل .

إن كتابة الهمزة بالرسم الإملائي تستند إلى قواعد مقننة من قبل مجامع اللغة العربية ويمكن فصلها في قاعدة بيانات خاصة بالهمزة والتعامل معها ببرنامج وفق هذه القواعد .

كما تجدر الإشارة إلى مسألة الشواذ في الكتابة . فكلمات مثل هذا والرحمن تكتب بالرسم العادي بلا ألف كما أن واو الجماعة لا تلفظ كما أن مسألة دمج بعض الكلمات في الرسم العثماني لها ما يقابلها من كلمات منفصلة في الرسم الإملائي . وبذلك فإن عدد كلمات المصحف بالرسم العثماني ستختلف عن عدد كلمات المصحف بالرسم العادي أي أن تقابل هذه الكلمات بعضها تجاه البعض يحتاج إلى معالجات

خاصة . انظر الأشكال (١) و (٣) و (٦) وقد تم انجاز قواعد البيانات هذه للقرآن كاملاً من قبل المؤلف.

٥ - قاعدة بيانات الجمل القرآنية

كما سبق ذكره فإن الجمل القرآنية متداخلة مع فواصل الآيات. أي أن الجمل القرآنية يمكن أن تكون بعض آية أو آية كاملة أو جزءاً من آية مع جزء من آية أخرى أو آية كاملة مع أجزاء من آية أخرى أو عددًا من الآيات.

لذلك كان من الضروري وضع إشارات ضمن النص تشير إلى نهاية جملة وبدء جملة جديدة . وذلك للقرآن كله. وتجدر الإشارة إلى أن علامات الوقف والتجويد قد يشير بعضها إلى مثل هذه المواضع حيث إن إشارة (الوقف اللازم أو الوقف أولى) تشير إلى بعض من هذه المواضع. خذ مثلاً سورة الفاتحة ، فالجمل فيها :

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ@

الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ@الرَّحْمَنِ الرَّحِيمِ@مَالِكِ يَوْمِ الدِّينِ@

إِيَّاكَ نَعْبُدُ

وَإِيَّاكَ نَسْتَعِينُ@

اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ@صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ@

وهذه التجزئة هي اجتهادية لا غير. وقد تختلف مواقع تجزئة النص إلى جمل بحسب فهم من يقوم بها من المفسرين. فمثلاً فاتحة سورة البقرة يمكن أن تجزأ كما يأتي:

@الم

ذَلِكَ الْكِتَابُ لَا رَيْبَ

فِيهِ هُدًى لِّلْمُتَّقِينَ@

ويمكن أن تجزأ كالاتي:

@الم

ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ

هُدًى لِّلْمُتَّقِينَ@

شكل (١) نموذج الكلمات القرآنية مرتبة هجائياً

ءَ الْهَيْتَا	ءَ أَسْجُدْ	ءَ أَشْكُرْ	ءَ أَمْنُتُمْ	ءَ أَنْتُمْ
ءَ أَتَّخِذْ	ءَ أَسْلَمْتُمْ	ءَ أَقْرَرْتُمْ	ءَ أَمِنْتُمْ	ءَ أَنْذَرْتَهُمْ
ءَ أَرْبَابَ	ءَ أَشْفَقْتُمْ	ءَ أَلَا ِِدْ	ءَ أَنْتَ	

شكل (٢) الجذور والكلمات التي يزيد ترددها عن ١٢٠ مرة في القرآن الكريم

كلمة(جذر)	عدد	كلمة(جذر)	عدد	كلمة(جذر)	عدد	كلمة(جذر)	عدد	كلمة(جذر)	عدد
عزز	١٢١	بصر	١٤٩	ضلل	١٩٢	حقق	٢٨٨	هو	*٤٨٢
ءمم	١٢٢	رود	١٤٩	حسن	١٩٥	ذكر	٢٩٣	رسل	٥١٤
يدي	١٢٢	وعد	١٥٢	خير	١٩٧	نزل	٢٩٤	بـ	*٥١٨
ءهل	١٢٤	وحد	١٥٤	نور	١٩٧	قبل	*٢٩٥	شيء	٥٢٠
بشر	١٢٤	غير	*١٥٥	جنن	٢٠٢	نفس	٢٩٩	بين	*٥٢٤
رزق	١٢٤	صدق	١٥٦	عند	*٢٠٢	ظلم	٣١٦	كفر	٥٢٦
خوف	١٢٥	بعض	١٥٩	ملك	٢٠٧	سمو	٣١٨	ءتي	٥٥٠
دخل	١٢٧	نصر	١٦٠	حكم	٢١١	كتب	٣٢٠	لم	*٥٧٥
بل	*١٢٨	شهد	١٦١	دعو	٢١٢	ءو	٣٢٦	قوم	٦٦١
خلف	١٢٨	ننبء	١٦١	ءنت	*٢١٨	هذي	٣٢٧	ءذ	*٦٩٤
عظم	١٢٩	كبر	١٦٢	ءم	*٢٢١	رءي	٣٢٩	ءلى	*٧٤٣
جمع	١٣٠	بنو	١٦٤	نوس	٢٣١	ءلي	٣٣٢	له	*٨٠٦
سءل	١٣٠	مع	*١٦٥	غفر	٢٣٥	رحم	٣٤٠	من	*٨٢٩
طوع	١٣٠	موت	١٦٦	بعد	٢٣٦	ثم	٣٤٣	علم	٨٥٥
لعل	١٣٠	سوء	١٦٨	ءمر	٢٤٩	جعل	٣٤٧	ءمن	٨٨٠
نظر	١٣٠	كثر	١٦٨	ءخر	٢٥١	عمل	٣٦١	ربب	٩٨٢
لكن	١٣١	شرك	١٦٩	ءيي	٢٥٣	عذب	٣٧٤	كون	١٣٥٩
نذر	١٣١	قلب	١٦٩	وقى	٢٥٩	كل	٣٧٧	على	*١٤٤٣
دنو	١٣٤	مثل	١٧٠	ولي	٢٦٠	ءيه	٣٨٣	ءلل	١٤٧١
قدر	١٣٤	قتل	١٧١	خلق	٢٦٢	قد	٤٠٧	في	*١٧٠٢
موسى	*١٣٧	تبع	١٧٥	ءخذ	٢٧٤	لك	٤٢١	قول	١٧٢٣
سلم	١٤١	سبل	١٧٧	عبد	٢٧٦	هم	٤٤٥	لا	*١٧٣١
حتى	١٤٣	صلح	١٨١	لو	*٢٧٦	عرض	٤٦٢	ما	*٢١٨١
دون	١٤٥	خرج	١٨٣	هذا	*٢٧٦	عن	٤٦٥	ءله	٢٨٥٢
نعم	١٤٥	سمع	١٨٦	جيء	٢٧٩	يوم	٤٧٦	من	*٣١٠٠
لقي	١٤٧	حيي	١٩٠	كذب	٢٨٣	ذلك	٤٨١	ءن	*٤٠٣٨

الشكل (٣) قاعدة بيانات رسم الكلمات القرآنية

k	n	oth	no_shada	no_beg	no_sh_end	word	rnum	root	verb	root2
1	7	#	#	#	#	#	0	#	#	#
2	7	@	@	@	@	@	0	@	@	@
٣	٠	بِسْمِ	بِسْمِ	اِسْمِ	بِسْمِ	بِسْمِ	38	سمى	اِسْمِ	سمى
٤	٠	اَللّٰهُ	اَللّٰهُ	اَللّٰهُ	اَللّٰهُ	اَللّٰهُ	٥٠	ءله	اَللّٰهُ	اَللّٰهُ
٥	٠	اَلرَّحْمٰنِ	اَلرَّحْمٰنِ	رَحْمٰنِ	اَلرَّحْمٰنِ	اَلرَّحْمٰنِ	٧٧	رحم	رَحْمٰنِ	رحم
٦	٠	اَلرَّحِيْمِ	اَلرَّحِيْمِ	رَحِيْمِ	اَلرَّحِيْمِ	اَلرَّحِيْمِ	٧٧	رحم	رَحِيْمِ	رحم
٧	٧	@	@	@	@	@	٠	@	@	@
٨	٠	اَلْحَمْدُ	اَلْحَمْدُ	حَمْدُ	اَلْحَمْدُ	اَلْحَمْدُ	١٦	حمد	حَمْدُ	حمد
٩	٠	اَللّٰهُ	اَللّٰهُ	اَللّٰهُ	اَللّٰهُ	اَللّٰهُ	٥٠	ءله	اَللّٰهُ	اَللّٰهُ
١٠	٠	رَبِّ	رَبِّ	رَبِّ	رَبِّ	رَبِّ	١١١	ربب	رَبِّ	ربب
١١	٠	اَلْعٰلَمِيْنَ	اَلْعٰلَمِيْنَ	عٰلَمِيْنَ	اَلْعٰلَمِيْنَ	اَلْعٰلَمِيْنَ	١٦١	علم	عٰلَمِ	عالم
١٢	٧	@	@	@	@	@	٠	@	@	@
١٣	٠	اَلرَّحْمٰنِ	اَلرَّحْمٰنِ	رَحْمٰنِ	اَلرَّحْمٰنِ	اَلرَّحْمٰنِ	٧٧	رحم	رَحْمٰنِ	رحم
١٤	٠	اَلرَّحِيْمِ	اَلرَّحِيْمِ	رَحِيْمِ	اَلرَّحِيْمِ	اَلرَّحِيْمِ	٧٧	رحم	رَحِيْمِ	رحم
١٥	٧	@	@	@	@	@	٠	@	@	@
١٦	٠	مٰلِكِ	مٰلِكِ	مٰلِكِ	مٰلِكِ	مٰلِكِ	٥٩	ملك	مٰلِكِ	ملك
١٧	٠	يَوْمِ	يَوْمِ	يَوْمِ	يَوْمِ	يَوْمِ	٣٨	يوم	يَوْمِ	يوم
١٨	٠	اَلدِّيْنِ	اَلدِّيْنِ	دِيْنِ	اَلدِّيْنِ	اَلدِّيْنِ	٢٩	دين	دِيْنِ	دين
١٩	٧	@	@	@	@	@	٠	@	@	@

وهناك بعض المواضع التي تشير إليها علامات خاصة من علامات الوقف والتجويد التي تبين إمكان الوقف على أي من الكلمتين (فيه أو هدىً) بشكل متبادل (أي أحدهما فقط) . ولكل من الصيغتين إعراب مختلف لكل جملة.

وهكذا تتكون قاعدة بيانات للجملة القرآنية أساسها هو قاعدة بيانات النحو. وهنا نود الإشارة إلى النصوص المستعملة اليوم والمتقلة بالرموز والتي تسمى corpus والتي تستعمل كأساس في اللغات الأخرى ويمكن اعتبار هذا النص المحتوي على علامات نهاية الجملة المفيدة (والتي يمكن أن يضاف له علامات أخرى حسب الحاجة) هو مشابه من ناحية الوظيفة لتلك النصوص في أساسه. وقد تم انجاز تقسيم القرآن إلى جمل كما هو مذكر أعلاه من قبل المؤلف.

٦ - قاعدة بيانات الصرف

هناك من الخطوات التي ينبغي إجراؤها قبل البدء بتكوين قاعدة بيانات الصرف لنص القرآن الكريم . أولى هذه الخطوات هي تجزئة الكلمات القرآنية إلى لواصقها الأولى ولواصقها النهائية وأدوات التشكيل ويمكن كحد أعلى اعتبار القطع ٤ لواصق أولية كحد أعلى و ٢ من اللواصق الآخريّة ونواة الكلمة وتشكيل للنواة وتشكيل آخر .

وبذلك فإن الكلمة يحجز لها ٩ حقول . خذ مثلا كلمة " أفالباطل " تحتوي ٤ لواصق أولية هي " أ - فَ - ب - ال " ونواتها هو باطل يعقبها كسرة في الأخير . أما كلمة " فسيكفيكمهم " فتحتوي على لواصق أولية عددها ٢ هما " فَ - سَ " والنواة يكفي واللواصق الآخريّة هي " ك " و " هم " والتشكيل على النواة هو السكون (غير الظاهر على الياء) وتشكيل آخر اللواصق هو السكون أيضا (الذي يظهر أو يختفي حسب الكلمة التالية للميم) ومثل ذلك في كلمة "أنلزمكموها"

وتجدر الإشارة إلى أن ألف لام التعريف عند حذفها يكون الحرف الأول من الكلمة مشدداً إن كان من الحروف الشمسية لذلك يجب حذف هذه الشدة من نواة الكلمة . أما إن كان من الحروف القمرية فإن اللام (من ألف لام) تحمل سكوناً يجب حذفه .

أما نواة الكلمة فيمكن أن تكون مشتقة من فعل ثلاثي على وزن ما . ويمكن أن يتم ذلك بواسطة برنامج خاص يقوم باستنباط الوزن الذي اشتقت منه هذه الكلمة . وهذا الاشتقاق يمكن أن يكون بسيطاً إذا لم يكن بين أحرف الجذر حرف علة . أما عندما يكون أحدها (أو أكثر من واحد) حرف علة فإن قواعد تحويل الواو إلى ياء أو ألف أو بالعكس تجعل القواعد أكثر تعقيدا كما هو معروف . أما إذا لم يكن النواة مشتقا فيعني ذلك عدم وجود جذر للكلمة . وعلى هذا فإن عدد حقول قاعدة بيانات الصرف تصبح ١٢ حقلا هي

الكلمة ، ٤ حقول للواصق الأولية ، نواة الكلمة ، تشكيلها ، لواصق آخريّة ، تشكيل الآخر ، جذر النواة ، وزن النواة . هذا وقد استعين في تحديد جذور الكلمات القرآنية بالمعجم المفهرس لألفاظ القرآن الكريم [٥] ولسان العرب [٦]. انظر الشكلين (٢) و (٤).

الشكل (٤) قاعدة بيانات الصرف

Wazin	oth	no_shada	no_beg	no_sh_end	word	root	verb	root2
فعلان	ا؟لرَّخَمَ إن	الرَّخْمَنِ	رَخْمَن	الرَّخْمَنِ	الرَّخْمَنِ	رحم	رَخْمَن	رحم
فعليل	ا؟لرَّجِمَ	الرَّجِيمِ	رَجِيم	الرَّجِيمِ	الرَّجِيمِ	رحم	رَجِيم	رحم
فاعلين	ا؟لْعَالَمِينَ	الْعَالَمِينَ	عَالَمِينَ	الْعَالَمِينَ	الْعَالَمِينَ	علم	عَالَم	عالم
فاعل	مَإلِكَ	مَالِكِ	مَالِك	مَالِكِ	مَالِكِ	ملك	مَالِك	ملك

٧- قاعدة بيانات النحو

يمكن تكوين قاعدة بيانات النحو هذه مشتقة من قاعدة بيانات الصرف وتختلف قاعدة البيانات هذه عن قاعدتي البيانات السابقتين (الكتابة والصرف) بأن عدد كلماتها أكثر من القاعدتين السابقتين حيث إن كل جزء من أجزاء اللواصق (الأولى والأخيرة) هي كلمة مستقلة، إضافة إلى نواة الكلمة. كما يلحق بكل كلمة من هذه الكلمات الجديدة تشكيلها. كما أنه من الضروري إضافة حقل خاص يحدد الكلمة القرآنية التي يعود لها جزء الكلمة هذه وموقعها التسلسلي ضمن الكلمة.

فكلمة فسيكفيكهم التي رقمها في قاعدة بيانات كتابة كلمات المصحف هو (٢٥٩٧)

اللواصق	الحركة	رقم الكلمة	تسلسل جزء الكلمة
ف	فتحة	٢٥٩٧	١
س	فتحة	٢٥٩٧	٢
يكفي	سكون	٢٥٩٧	٣
ك	فتحة	٢٥٩٧	٤
هم	سكون أو ضم عند التقاء الساكنين	٢٥٩٧	٥

وهنا بعد هذا يضاف إعراب الكلمات واحدة واحدة مع تبيان إن كانت الكلمة مبنية أو معربة وعلامة الإعراب أو أية معلومات تفصيلية عن الإعراب.

وتجدر الإشارة إلى أن بعض أجزاء الكلمة قد لا يكون له محل من الإعراب مثل ألف لام التعريف فيشار إليه بذلك.

يبقى هنا كيف يمكن التعبير عن إعراب أشباه الجمل أو الجمل، وقاعدة البيانات المذكورة غير مناسبة لذلك فيصبح من الضروري تكوين قاعدة بيانات فرعية تحدد رقم الكلمة وتسلسل جزء الكلمة الذي تبدأ به شبه الجملة ومداها ثم موقعها من الإعراب وعلامة الإعراب مع نص شبه الجملة أو الجملة.

هناك أمر آخر على جانب من الأهمية في موضوع النحو وهو عودة الضمير. فالضمائر في اللغة العربية تعود إلى كلمة سابقة أو شبه جملة. وقد تكون الضمائر ظاهرة أو مستترة. لذلك من الضروري اعتبار الضمير المستتر كلمة فارغة بين أجزاء الكلمة ضمن تسلسلها مع إضافة حقل خاص لتبيان ما يعود له الضمير.

إن تكوين قاعدة البيانات هذه ليس أمرًا واحدًا ثابتًا كنص القرآن الكريم فهناك خلافات نحوية كثيرة بين المتخصصين في النحو في إعراب الكثير من الكلمات القرآنية واستنادًا إلى تلك الخلافات يختلف المعنى والتفسير أيضًا ولغرض استيعاب مثل هذا الخلاف بالرأي يمكن أن توضع قاعدة بيانات النحو بشكل يقبل تعدد الآراء بحيث تصبح كقاعدة بيانات ثانوية تحتوي ما يأتي :

رقم الكلمة القرآنية

رقم جزء الكلمة

حركة جزء الكلمة

نوع جزء الكلمة مبنية أم معربة وهل هي اسم أو فعل أو حرف.

موقعه من الإعراب

رمز الرأي الإعرابي

عودة الضمير

ملاحظات إن وجدت.

كما أن الاصطلاحات التي يتم اعتمادها يمكن أن تكون قواعد بيانات فرعية مثل قاعدة بيانات الكلمات المبنية وقاعدة بيانات علامات الإعراب وقاعدة بيانات المواقع من الإعراب. وسيشير إلى ذلك لاحقاً بالتفصيل بالإضافة الى حقل الجنس (التذكير والتأنيث) إن وجد ، حقل الأفراد والتثنية أو الجمع إن وجد وبالنسبة للجمع نوع الجمع إن كان جمع مذكر سالم أو جمع مؤنث سالم أو جمع تكسير. أنظر الشكل [٧].

الشكل (٥) قاعدة البيانات النحوية

Word	no_sh_end	الإعراب	علامة الإعراب
#	#		
@	@		
	=	حرف جر	
بِسْمِ	اسم	اسم مجرور وهو مضاف	الكسرة
اللَّهِ	الله	مضاف إليه	الكسرة
الرَّحْمَنِ	الرَّحْمَنِ	صفة	الكسرة
الرَّحِيمِ	الرَّحِيمِ	صفة	الكسرة
	أبتدأ		
+	+	+	نهاية جملة
@	@		
الْحَمْدُ	الْحَمْدُ	مبتدأ مرفوع	الضمة
	لِ	حرف جر	
لِلَّهِ	الله	إسم مجرور - الجار والمجرور خبر	الكسرة
رَبِّ	رَبِّ	صفة وهو مضاف	الكسرة
الْعَالَمِينَ	الْعَالَمِينَ	مضاف اليه	الياء والنون
@	@		
الرَّحْمَنِ	الرَّحْمَنِ	صفة	الكسرة

Word	no_sh_end	الإعراب	علامة الإعراب
الرَّحْمَنِ	الرَّحْمَنِ	صفة	الكسرة
@	@		
مَالِكٍ	مَالِكٍ	صفة وهو مضاف	الكسرة
يَوْمٍ	يَوْمٍ	مضاف إليه وهو مضاف	الكسرة
الَّذِينَ	الَّذِينَ	مضاف إليه	مبني
@	@		
+	+	+	نهاية جملة
إِيَّاكَ	إِيَّا	مبتدأ وهو مضاف	مبني
	كَ	مضاف إليه	ضمير متصل مبني
نَعْبُدُ	نَعْبُدُ	فعل مضارع مرفوع	الضمة
	نَحْنُ	فاعل مستتر مقدر	
+	+	+	+
وَ		حرف عطف	
إِيَّاكَ	إِيَّا	مبتدأ وهو مضاف	مبني
	كَ	مضاف إليه	ضمير متصل مبني
نُسْتَعِينُ	نُسْتَعِينُ	فعل مضارع	الضمة

٨- قواعد بيانات المعاني والدلالة

قاعدة البيانات هذه ما هي إلا أفكار عامة تحتاج إلى مزيد من العناية والتوسيع فموضوع تقنين المعاني والدلالة على جانب كبير من الصعوبة بالنسبة للغات الحديثة فكيف باللغة العربية الواسعة المعاني والدلالة وكيف إن كان الموضوع يتعلق بصفوة البيان العربي ألا وهو القرآن الكريم .

إلا أن قاعدة البيانات هذه تكاد تكون صفوة قواعد البيانات. أول قاعدة بيانات للمعاني هي قاعدة بيانات الجذور . فالجذور من ناحية المعنى مترابطة مع بعضها فمثلا جذور مثل : جيء ، أتى ، قدم ، حضر ، وصل ، ولي ، دبر ، وكذلك : دخل ، ولج ، قبل ، ورد ، وصل وهذه المجموعات مترابطة مع بعضها حيث إنها تتعلق مع غيرها من المجموعات بحركة الإنسان ذهاباً أو إياباً أو خروجاً أو دخولاً أو وصفاً لكيفية هذه الحركة من سرعة أو انحدار أو ارتفاع أو تغير أو خوف أو إجبار أو توقيت أو حلول أو تكرار . وعلى هذا فالأفعال تتجمع في مجموعات يمكن وصفها بصفة جامعة عامة ثم تنقسم إلى مجموعات أدنى من ذلك تتصف

الشكل (٦) قاعدة بيانات الكلمات مع اللواحق

WORD	ROOT	K_A	PRE1	PRE2	CENTER2	SIGN1	LAST1	LAST2	OTH	K_B
#	#				#				#	1
@	@				@				@	2
	بِسْمِ	سمى ١١٦٠٢	بِ		اسْمِ	ِ			بِسْمِ	٣
	اللَّهُ	أله ١٢٠١١			اللَّهُ	ِ			اللَّهُ	٤
	الرَّحْمَنِ	رحم ١١٨٠٦	ال		رَحْمَنَ	ِ			الرَّحْمَنِ	٥
	الرَّحِيمِ	رحم ١١٦٨٢	ال		رَحِيمَ	ِ			الرَّحِيمِ	٦
@	@				@				@	٧
	الْحَمْدُ	حمد ٧٥٢٢	الْ		حَمْدَ	ُ			الْحَمْدُ	٨
	لِلَّهِ	أله ١٢٠٠٨	لِ		اللَّهُ	ِ			لِلَّهِ	٩
	رَبِّ	رب ١٠٣٧٧			رَبِّ	ِ			رَبِّ	١٠
	الْعَالَمِينَ	علم ٦٢٦١	الْ		عَالَمَ	ِينَ			الْعَالَمِينَ	١١
@	@				@				@	١٢
	الرَّحْمَنِ	رحم ١١٨٠٦	ال		رَحْمَنَ	ِ			الرَّحْمَنِ	١٣
	الرَّحِيمِ	رحم ١١٦٨٢	ال		رَحِيمَ	ِ			الرَّحِيمِ	١٤
@	@				@				@	١٥
	مَالِكِ	ملك ١١٤٦٥			مَالِكِ	ِ			مَالِكِ	١٦
	يَوْمِ	يوم ١١٦١٧			يَوْمِ	ِ			يَوْمِ	١٧
	الَّذِينَ	دين ١١٩٦٣	ال		دِينَ	ِ			الَّذِينَ	١٨
@	@				@				@	١٩
	إِيَّاكَ	أبي ٤٢٥٣			إِيَّا		ك		إِيَّاكَ	٢٠
	نَعْبُدُ	عبد ٧٥٤٧			نَعْبُدُ	ُ			نَعْبُدُ	٢١
	وَأِيَّاكَ	أبي ٤٢٥٤	وَ		إِيَّا		ك		وَأِيَّاكَ	٢٢
	نَسْتَعِينُ	عون ٨٧٥٤			نَسْتَعِينُ	ُ			نَسْتَعِينُ	٢٣
@	@				@				@	٢٤

بجامع أخص من الأول وهكذا . لكن المجموعات النهائية لا تتفق في المعنى فهي مؤتلفة أحياناً ومختلفة أحياناً أخرى . وهناك الكثير من المؤلفات في التراث العربي التي تذكر مثل هذه الألفاظ [٧] . كما هناك قواميس حديثة تعطي مثل هذه المجموعات وما يقابلها من مفردات أجنبية [٨] .

إن الجذور وحدها تجمع في غالبيتها أصولاً لمعاني عديدة . ولذلك فإن الاستناد إليها يعطي وجهة عامة غير دقيقة للدلالة . لذلك يجب الانتقال من مستوى المجموعات العامة للجذور إلى المجموعات المتخصصة إلى مجموعات الجذور المتقاربة إلى الأوزان المختلفة التي تدخل على الجذر نفسه . فللجذر ذهب صيغ أفعال مثل ذهب ، أذهب ، ذهب وللجذر قدم : صيغ مثل : أ قدم (أ فعل) وقدم (فَعَل) واقتدم (افعل) تقادم (تفاعل) تقدّم (تفعل) واستقدم (استفعل) [٩] .

ويلاحظ أن بعضاً من هذه الصيغ ذات علامة بالسير . أما بعضها مثل تقادم فليس لها علامة بذلك بل تتعلق بمعنى الزمن ، كما أن قدّم قد لا تشير إلى حركة الإنسان وحده بل إلى حركة الأشياء أيضاً . وهكذا فإن سلسلة الأفعال المتقاربة (المزيدة) بالمعنى يجب أن ينظر إليها وليس إلى الجذور المتقاربة فقط .

هناك حاجة أيضاً إلى الإشارة إلى الفروق بين الأوزان المختلفة . فمثلاً قدّم واستقدم تشيران إلى فعل القدوم أو التقديم على الغير وللغة العربية قواعد راسخة في مثل هذه المعاني تحتاج إلى إدخال إلى قواعد البيانات بطرق مناسبة ذات ترميز دقيق .

الأمر الآخر الذي ينبغي أن تشمله قواعد البيانات هو تعدد المعاني للفظ الواحد .
الأفعال باللغة العربية من ناحية المعاني يمكن تقسيمها إلى :

أفعال لازمة	:	اركعوا واسجدوا
أفعال متعدية لمفعول واحد	:	ورد ماء مدين
أفعال متعدية بحرف الجر	:	واسجدوا لله
أفعال متعدية لمفعول واحد وبحرف الجر	:	آوى إليه أخاه
أفعال متعدية لمفعولين	:	فسيكفيكم الله
أفعال متعدية لمفعولين وبحرف الجر	:	وجد عليه أمة من الناس يسقون

وقد يكون الفعل تارة من أحد الأفعال المتعدية بحروف الجر بحيث يعطيها معانٍ مختلفة عند اختلاف حرف الجر ، فالفعل ذهب مثلاً ورد مع إلى مثل " اذهب إلى فرعون " . وقد تأتي الجمل أو أشباه الجمل في موقع المفعول به في أفعال أخرى، ومع الباء : " تذهبوا به " ومع عن : " يذهب عنكم الرجس " وقد يأتي لازماً بلا حرف جر مثل " اذهب أنت وربك " والفعل أتى أكثر من ذلك تشعباً ، مثل " يأت آباءهم " ، " فأتوا بآياتنا " و ، " وأتوا البيوت من أبوابها " ويشترك منه الفعل الرباعي المزيد آتى : " آتاهم الله من فضله " و " وأوتيت من كل شيء " وكما إن تعدي الفعل لمفعول معين يشير إلى معنى مثل " يبغونها عوجاً " . كذلك يشير تضاد المعاني بين أفعال (أو كلمات) وهذا التضاد قد يكون من جهة معينة . فقد يكون التضاد يتعلق بالاتجاه مثل ذهب وأتى أو قام وقعد أو مثل فوق وتحت . وقد يكون التضاد من جهة الزمن مثل الماضي والمستقبل وقد يكون من جهة التبادل بين شيئين مثل أخذ وأعطى وقد يكون من جهة صفة معينة مثل أمات وأحى وهكذا وهذا التضاد يعطي المعنى الدقيق أحياناً إذا عرف معنى الضد .

وفي القرآن كما في اللغة العربية أصلاً عبارات كاملة تشير إلى معنى محدد كان يمكن أن يشار له بفعل واحد (لكن ربما تحمل معنى إضافياً)

قدرناها من الغابرين = أهلكناها

خذوا ما آتيناكم بقوة = اتبعوا ما آتيناكم بدقة

واضرب لهم مثلاً = اقصص عليهم مثلاً

مما تقدم يتبين أن قواعد البيانات التي تخدم المعنى والدلالة واسعة جداً ومعقدة ، نظراً لصعوبة الموضوع أصلاً ولكنها ضرورية جداً للتقدم في مجالات معالجة بيانات اللغة العربية آلياً والترجمة والدلالة .

٩ - قاعدة بيانات اللفظ

كانت العناية بقراءة وترتيل وتجويد القرآن فائقة على مر العصور . وقد عني المسلمون بقواعد التجويد ووضع الرموز المساعدة للوقف واللفظ الصحيح أثناء القراءة . وتحتوي المصاحف الشائعة اليوم على الكثير من هذه الرموز وفق القراءة القرآنية الشائعة (هناك ثلاث قراءات تشيع المصاحف المكتوبة وفقها وهي بروايات حفص وورش وقالون) .

ويمكن لقاعدة بيانات تجويد القرآن الكريم أن تحتوي على ما يأتي :

(أ) قاعدة بيانات القراءات القرآنية

فالمعروف أنه على الرغم من التزام كل القراء بالاعتماد على الرسم العثماني إلا أن هناك اختلافات طفيفة بينها ويمكن جمع هذه الاختلافات في قاعدة بيانات تبين ذلك .

يضاف إلى ذلك الرموز المختلفة التي تشير إلى طريقة نطق بعض الحروف كالهزمة والتنوين .

ولا يكفي أن يوضح نطق كل كلمة منفردة لوحدها بل يجب تبيان نطق ارتباط الكلمات المتتالية مع بعضها حيثما كان ذلك ضرورياً للإيضاح.

(ب) قاعدة بيانات الألفاظ الشاذة وهي كلمات خاصة تلفظ بصيغة معينة في بعض القراءات دون أخرى .

(ج) قاعدة بيانات الأصوات .

فالمعروف أن وحدة الكلام المنطوقة هي الأصوات phonemes وليس الحروف وتحتوي اللغة العربية على أصوات لا توجد في غيرها كما أن أصوات التجويد وخاصة المدود ولفظ النون الساكنة والميم الساكنة والإدغام والإمالة والتخفيف تنطق بشكل معين . وجمع رموز هذه البيانات يجعلها مفيدة للاستعمال في دراسات وأبحاث عن اللغة العربية .

إن نطق بعض الحروف أو الأصوات يخضع في القراءات القرآنية لكثير من القواعد غير الظاهرة في المصاحف. فهناك أصوات مختلفة لبعض الحروف مثل الراء المخففة والمفخمة والنون والنون مع الغنة المخففة والنون مع الغنة المدغمة والواو والواو الاختكاكية والواو الصائتة والياء والياء الاحتكاكية والياء

الصائتة وصوت الإمالة وصوت التفخيم واللام المرققة واللام المفخمة. هذا بالإضافة إلى الحركات الثلاث: الفتحة والضمة والكسرة [١٠].

(د) قواعد البيانات الصوتية .

لقد أصبح بالإمكان تجزئة الصوت بالحاسوب وتقطيعه وإعادة تركيبه والتعرف عليه. فإذا ما خزن صوت نموذجي لقارئ ما وكان هذا النموذج وافيًا وقطّع إلى أجزاء مناسبة فإن بالإمكان إعادة تركيب هذه الأجزاء لقراءة نص آخر مغاير للنص الأصلي لكنه يحتوي على تلك القطع والأجزاء وبصوت القارئ نفسه .

١٠ - قواعد البيانات الارتباطية الموسعة

يبين الشكل (٨) مخططاً لقواعد البيانات الارتباطية التي تخدم معظم الأغراض المشار إليها أعلاه. فهي تبدأ بالرسم العثماني للمصحف والذي يحوي عددًا من العلامات المبوبة في ملف خاص يشير إلى معنى كل منها. ومن الرسم العثماني تتكون قاعدة بيانات الرسم الإملائي مشكولاً وغير مشكول . أما الجمل القرآنية والفواصل بينها فتوضع في قاعدة بيانات خاصة بها ومنها يمكن تكوين ملف أشباه الجمل والتراكيب أو العبارات التي ينبغي التعامل معها مجتمعة.

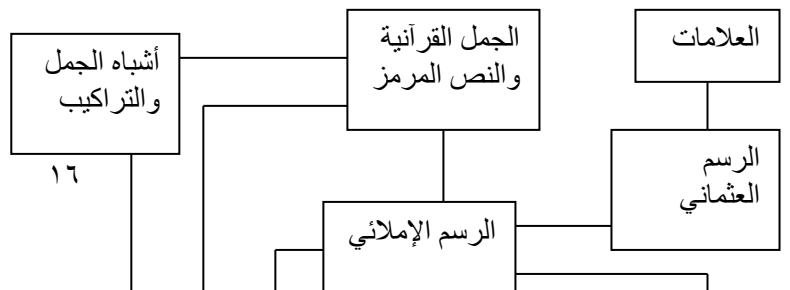
من قاعدة بيانات الرسم الإملائي يمكن استنباط الكلمات الكاملة ومنها قاعدة الكلمات بعد تجزئتها إلى لواصقها ونواتها. هذه الكلمات هي المصدر الذي تتكون منه مواد المعجم والذي يستقي أصناف المعاني فيه من قاعدة خاصة تبين صفات المعاني المختلفة للألفاظ.

الكلمات المجزأة تتكون من لواصق ونواة. وهذه النواة إن كانت مشتقة فهي منصرفة وفق الأوزان الصرفية المعروفة وهي إما أن تكون أسماء أو أفعال. الأفعال يمكن أن تكون مجردة أو مزيدة وفق أوزان معينة كما يمكن أن تكون الأفعال أنواعًا مختلفة من ناحية التعدي واللزوم أو التعدي لأكثر من مفعول أو تعدي بحروف الجر. وهذه الأفعال والأسماء ترتبط بقاعدة بيانات الجذور.

الجذور يمكن وضعها بمجموعات مثل مجموعة الجذور المتقاربة أو المتدرجة والجذور المتضادة المعاني. أما النحو فيحتاج إلى تصنيف الكلمات إلى أسماء وأفعال وحروف والأسماء إلى مبنية ومعربة والمبنية منها الضمائر وأسماء الإشارة وأسماء الموصول وغيرها وتحتاج إلى تخصيص من نواحي التذكير والتأنيث ومن نواحي الأفراد والتنثنية والجمع. أما الأفعال فتحتاج إلى تخصيصات أخرى مثل الزمن من ماضي وحاضر ومستقبل أو من نواحي كونها تامة أو ناقصة.

هذه القواعد الارتباطية تقوم بخدمة المعجم بشكل عام وتجعل الوصول إلى البيانات المختلفة سهلة وممكنة.

الشكل (٨) قواعد البيانات الارتباطية



١١ - المعاني اللغوية العميقة وبلاغة اللغة

إن أية لغة تحتوي إيماءات ودلالات تؤخذ من سياق الجملة ومن نسق لفظ المتكلم ومن موضوع النص ولا تدل عليها الكلمات إن عولجت كلمة كلمة ، بشكل منفصل وبمعزل عن السياق العام. وتعتبر اللغة العربية من أغنى لغات العالم اليوم في ذلك ويأتي القرآن الكريم في قمة تلك البلاغة. إن المعالجات الآلية البلاغية للغة والمعاني اللغوية العميقة تعتبر خطوة متقدمة جدًا في هذا الحقل وتحتاج إلى أبحاث معمقة كثيرة. إلا أن من المؤكد أن قواعد البيانات إذا ما صممت بشكل معين فإنه يمكنها أن تساعد في هذه المعالجة مستقبلاً.

١٢ - ماذا يمكن أن تقدم قواعد بيانات القرآن الكريم للغة العربية

إن صغر حجم بيانات القرآن الكريم بالمقارنة مع اللغة العربية جعل بالإمكان إكمال بعض قواعد البيانات بجهد معقول . أما البيانات الكاملة للغة العربية فتحتاج إلى جهود كبيرة . لكن الأمثلة التي عرضت في هذا البحث تعطي إشارة واضحة إلى سهولة الكثير من هذه الجهود وإمكان برمجتها . ويمكن تلخيص ما يمكن أن تقدم بيانات القرآن الكريم للغة العربية على النحو الآتي :

١- تعطي قواعد بيانات القرآن الكريم أمثلة واضحة للكثير من المعالجات التي يمكن أن تتم على اللغة العربية . وما تحتاجه اللغة العربية سوى توسيع بعض هذه القواعد لكي تشمل كل الكلمات والمعاني المستعملة في اللغة .

٢- إن البرامج التي كتبت وفق قواعد معينة لمعالجة كلمات القرآن الكريم يمكن استخدامها نفسها على اللغة العربية بشكل عام.

٣- إن نصوص القرآن الكريم التي تمت معالجتها كانت دقيقة لكنها تحتوي على الكثير من أشكال الكتابة غير القياسية . أما اللغة العربية الحديثة فهي تستعمل أدوات قياسية أكثر ، لكن احتمال الخطأ في الكتابة يمكن أن يستفيد من أشكال الكتابة غير القياسية في المصاحف . وعندها يمكن تصحيح أخطاء الكتابة مثل الهمزة أو التنوين وغيرها .

٤- إن نصوص القرآن الكريم التي تمت معالجتها كانت مشكولة واشتق منها نص غير مشكول . وحيث إن الكتابة العادية في غالبيتها غير مشكولة لذلك فإن بالإمكان الاستفادة من قواعد بيانات القرآن الكريم لإعادة بعض التشكيل بصورة آلية إلى الكتابة العربية .

٥- إن الكثير من الأبحاث على قواعد بيانات القرآن الكريم لم تؤت ثمارها بعد لحاجتها إلى الكثير من المعلومات اللغوية . لكن تصميم هذه القواعد يعطي فوائد للغة وفي الوقت نفسه يعود بالفائدة على أبحاث القرآن الكريم وخاصة في علم التفسير . ويمكن أن نذكر هنا بعض قواعد البيانات التي تخدم المعاني والدلالة في القرآن الكريم .

أ- قاعدة بيانات مجموعات الجذور العامة .

ب- قاعدة بيانات مجموعات الجذور المتخصصة . أنظر الشكل (٧)

الشكل (٧) الجذور الواردة في القرآن ذات العلاقة بحركة الإنسان وغيره من المخلوقات

عيق عتي آخر عزف عوب بعث بعد بغت بقي بلغ بيت ترك ثبت جرر جلب جيء حضر خطو خفض خلد خلص خلف
خنس دبب دبر دخل درج دفع دنو دهم دور ذهب رجع رحل ردد رسو رقد رفع رقي ركب ركض ركن زحف زفف سبق
سحب سحل سرح سرع سري سعي سفر سكن سلك سير شنت شحن شرد شطء شيع سحب صدد صدف صرط صعد
صلي ضلل طرد طفق طلب طلق طوف طوق طوي طير ظهر عبر عدو عرض عرو عزب عزل علو عمق عود
غدو غرب غيب فرر فوت فوج فيء قبل قحم قدم قرب قصد قصو قعد قفل قفو قوم كبكب كرر لحق لوذ لوي مخر مدد
مرد مرر مضي نزل نفر نفى هرب هلع هلم هيم وجه ودع ورد وري وصل وطم وطن وقر وقع وقف ولج يمم

ج - قاعدة بيانات مجموعات الجذور المتسلسلة المعنى .

د - قاعدة بيانات الأفعال المزيدة .

هـ - قاعدة بيانات أنواع الأفعال من ناحية اللزوم والتعدي .

و - قاعدة بيانات تعلق حرف الجر بالأفعال .

ز - قاعدة بيانات الكلمات المتضادة .

ح - قاعدة بيانات العبارات التي تشير إلى معنى محدد .

ط - قاعدة بيانات المعاني المعقدة .

٦- إن دراسات الصوت التي يمكن أن تتم على القرآن الكريم يمكن أن تؤدي ثمارها في

الترجمة الآلية وآلات الإملاء الآلي وتركيب الأصوات .

١٣ - المصادر:

١- Naphtali D. Rishe, Database Design: The Semantic Modeling Approach, McGraw-Hill, 1992

٢- http://www.larflast.bas.bg/balric/eng_files/dictionary_eng1.php

<http://www.hpdr.c.fiu.edu/library/papers/SDBMS.RDBMS.OODBMS.html>

<http://www.cogsci.princeton.edu/~wn>

٣- تفسير القرآن العظيم للحافظ ابن كثير

٤- الجامع لأحكام القرآن العظيم للقرطبي

٥- المعجم المفهرس لألفاظ القرآن الكريم لمحمد فؤاد عبد الباقي

٦- لسان العرب لابن منظور

٧- الألفاظ المختلفة في المعاني المؤتلفة - جمال الدين الجباني دار عمار ١٩٩١

٨- قاموس المفردات المتضادة - انكليزي - عربي د. كاظم عادل ناصر دار البشير ١٩٨٩

٩- إحصاء الأفعال العربية في المعجم الحاسوبي : مروان البواب وغيره، مكتبة لبنان ١٩٩٦

١٠- القراءات القرآنية بين الدرس الصوتي القديم والحديث للدكتورة مي فاضل الجبوري - دار الشؤون الثقافية العامة - بغداد ٢٠٠٠

TOWARDS SEMANTICS PROCESSING OF ARABIC LANGUAGE VIA DATABASES:

A PRELIMINARY STUDY OF THE TEXT OF THE HOLLY QURAN

Professor Mohammed Zeki Khedher

Jordan University

Amman - Jordan

Email:khedher@ju.edu.jo

Semantic processing of Arabic language needs a huge amount of information about various aspects of the language. The best method of tabulation of this information is via databases. This information starts from the databases related to the mode of script and the database of morphology. This includes separation of the nucleus of the word from its prefix and suffix as well as relating the nucleus with the morphological rules. The syntactical database needs separation of the components of the prefixes and suffixes taking into accounts the hidden pronouns so as to use extendable relational databases. The database of the pronunciation is also needed for comprehensive processing of Arabic language. All this processing and databases serve for the information related to semantics in a comprehensive relational system. The paper explains how all that is possible based on the text of the holly Quran, hence approaching the exact contextual understanding of the Arabic language.