

اللسانيات الحاسوبية واللغة العربية
إشكالات وحلول
كنوز المعرفة - ٢٠١٨

الذخيرة اللغوية لتراكيب القرآن الكريم

أ.د. محمد زكي خضر

لقد خُدم نص القرآن الكريم بصيغ عديدة. وقد كان آخرها الذخيرة اللغوية لكلمات القرآن الكريم، لكنه لم تصدر أية محاولة لعرض تراكيب القرآن الكريم بشكل ذخيرة لغوية. هذا البحث يستند إلى التراكيب المتشابهة لفظاً في القرآن الكريم المرتبة وفق جذور الكلمات المكونة للتركيب وليس للكلمات نفسها. أي أن التراكيب التي تشترك كلماتها بالجذور نفسها تعتبر متشابهة ومجمعة مع بعضها. لذلك فقد استثنيت التراكيب التي تظهر في القرآن الكريم مرة واحدة لعدم وجود تكرار لتلك التراكيب. إن الأسماء والأفعال في اللغة العربية غالباً تُشتق من جذور الكلمات الثلاثية وأحياناً الرباعية والخماسية. لذلك فإن هذه الذخيرة يمكن أن تكون أساساً لكثير من الأبحاث المستقبلية المتعلقة بعلاقة كلمات اللغة العربية مع بعضها البعض، خاصة في الأبحاث ذات العلاقة بالمعاني والدلالة. ويؤمل أن تخدم هذه الذخيرة ربط المفاهيم في القرآن الكريم مع كتب الحديث النبوي والفقه ومختلف العلوم الشرعية بالإضافة إلى اللغة العربية (١)

الكلمات المفتاحية: الذخيرة اللغوية، ذخيرة تراكيب القرآن الكريم، ذخيرة تراكيب اللغة العربية

١- تقديم

تعتبر الذخيرة اللغوية ذات أهمية بالغة في دراسة اللسانيات الحديثة. ففي مختلف عمليات التحليل والمعالجة للذخيرة اللغوية خدمة للمعالجة الحاسوبية للغات الطبيعية وفي التعرف على الأصوات وفي الترجمة الآلية، حيث يستخدم نموذج ماركوف المخفي في عمليات التأشير الداخلي للأصوات وتطبيقات عديدة أخرى (٢).

تتوفر ذخيرة لغوية للقرآن الكريم (٣) نتيجة تعاون من جهات عديدة بمبادرة من جامعة ليدز في بريطانيا. هذه الذخيرة مفتوحة المصدر وجُهزت بواسطة تجزئة الكلمات القرآنية وترميزها ووضعت بشكل منظم ومرتبب بمخططات بيانية للنحو. وقد نشرت الطبعة ٠,٤ في مايس عام ٢٠١١. وسنأتي على ذكر لها في ما بعد.

إن معنى كلمة لا يمكن أن يحدد في كثير من الأحيان بفصل الكلمة عن السياق الذي وردت فيه. توصف الكلمة بما ي صاحبها من أفعال وليس بكلمة منفردة. فمثلاً كلمة الليل تقترن بمفهوم الظلام. فوصف الاقتران يعني أن كثرة تكرار الليل مع الظلام هو الذي أعطى صفة الظلام لليل، وهذا الاقتران قد وجد بعدد كافٍ من التكرارات التي جعلت من غير الممكن تصور أنه لا توجد علاقة بين الليل والظلام(٤).

يقصد بالكلمات المتعددة (هنا سنستخدم مصطلح "تراكيب متعددة الكلمات") ارتباط كلمتين أو أكثر من الكلمات المستقلة إملانياً بحيث تردا مرتبطين ببعضهما إصطلاحياً. ونعني بالاصطلاح أنها بالأساس لا يمكن التنبؤ بالعلاقة بينهما في الصيغة والدلالة والتردد(٥).

لقد حظي موضوع التراكيب المتعددة الكلمات بنظرة واهتمام خاص. فالتعرف على التراكيب متعددة الكلمات يؤثر على نوعية النتائج للمهام المختلفة بكثرة في اللغات الطبيعية، مثل عمليات الترميز والإعراب وتوليد الكلام. وقد استخدمت طرائق مختلفة لتحديد التراكيب متعددة الكلمات كالوسائل الاحصائية والتي استخدمت لإيجاد نماذج التكرار في اللغة(٦).

في إحدى الدراسات (٧) استخدمت الذخيرة اللغوية للتعرف على النهج الذي تؤثر الكلمات المتعددة فيه على المحاورات والكتابة الأكاديمية، كما استخدمت النهج الذي تسلكه في صيغتين، فهناك طريقتان للتعامل مع الكلمات المتعددة: الأولى تعنى بالمفهوم الناجم عن دمج الكلمات والثانية تعنى بأداة كل كلمة بالاضافة إلى المفهوم.

في أحد الأبحاث نُشرت هيكلية واسعة الشمول وعميقة لتمثيل الذخيرة متعددة الكلمات مع المعجم(٨). ولم تحدد الطريقة في مضمونها نوعاً خاصاً من المعاجم أو المخزون النحوي. فقد احتوت على طيف عريض من الذخائر متعددة الكلمات وأصنافها، من بينها أسماء الأشياء و التركيبات العادية ووسائل الارتباط والضمائر وغيرها، كما شملت العبارات التي تحوي أفعالاً مع ما يدخل عليها من تحويرات داخلية أو تحويرات بتسلسل الكلمات ووسائل الاتصال بينها. فمثلاً عبارة " موقعه يجلب الانتباه" يمكن أن ترتبط مع عبارة " الموقع يسترعي الانتباه" أو "ليس في الموقع ما يسترعي الانتباه". ففي هذه الحالات هناك فجوات بين الكلمات المشتركة في كل عبارة من العبارات الأخيرة. كذلك فإن حروف الجر والعطف وغيرها لا تعتبر جزءاً أساسياً من الذخيرة اللغوية، لذلك فإن البحث اعتبر أن مهمة حروف الجر والعطف ليست جزءاً كاملاً من الذخيرة إذا ما أجريت عليها عملية تحليلاً نحويًا.

تُنشأ الأنظمة المحتوية على ذخيرة متعددة الكلمات لغرض عمل جدول فهارس للعبارات باستعمال علامات مع أنظمة التشكيل والتجزئة الخفيفة(٩). وقد أضاف هذا البحث استعمال أدوات مع مرشحات لتفادي المشاكل التي تصاحب الانحراف الدلالي الناتج عن الاقتباس. إن التطور في استعمال المصطلحات والحاجة لدمج الكلمات للحصول على تعابير جديدة قد وصل إلى أن تحتل هذه التعابير إلى حد يصل إلى ٣٠% بدقة قد تصل إلى ٩٠% وذلك لتمييزها عن التعابير الأخرى في هذا البحث، ويمكن أن يكون ذلك منصة تصلح للاستعمال في الترجمة الآلية، وقد جُرب ذلك فعلاً بين اللغتين الإنكليزية والفرنسية.

هناك محاولة لعملية إنتاج تعابير مكونة من كلمات متعددة موجودة في الكتب الإنكليزية الشائعة المؤلفة من أناس ليست لغتهم الأم هي الإنكليزية(١٠). وكان الهدف من تلك الدراسة هو البحث عن تأثير اللغة الأم على إنتاج تعابير إنكليزية صحيحة أو خاطئة وخاصة في عملية ربط الكلمات مع بعضها.

إن عملية استرجاع المعلومات باللغات الطبيعية وفق قوالب معينة للاستفسار الحواري قد عولجت بشكل أو بآخر(١١). هناك إمكانية لزيادة القابلية التحوارية باستعمال العبارات متعددة الكلمات المختارة من أعلى عدد من الوثائق. إن العبارات المتعددة الكلمات هي وحدات لنصوص تمثل حيزاً لمفاهيم وأشياء. فهي بذلك أكثر مناسبة لتمثيل المواضيع من العبارات العادية.

هناك نظام آخر(١٢) يحتوي على منحيين جديدين: الأول هما: استعمال مخرجات أدوات معالجة اللغات الطبيعية مثل عملية استخراج عبارات متعددة الكلمات ثنائية اللغة لتحسين عملية الترجمة الآلية، والثاني هو تقديم طريقة ترشيح جديدة مستندة إلى استعمال أدوات في مقابلة الجمل في اللغتين. لقد أظهر هذا النظام زخماً في تحسين مستوى الترجمة الآلية بإدخال عناصر جديدة في معالجة اللغات الطبيعية وبالأخص الترجمة الآلية.

كما استحدثت طريقة تعتمد على الذخيرة اللغوية لدراسة العبارات متعددة الكلمات التي تحتوي على كمية محسوسة من محتويات اللغة. وقد اختصت هذه الدراسة(١٣) بذخيرة اللغة الألمانية التي تحوي على بليون كلمة والمحفوظة في معهد الدراسات الألمانية. وقد استخدمت هذه الطريقة كثرة الاستخدام كأساس لها.

يمكن الحصول على عبارات متعددة الكلمات من بيانات الذخيرة المتوفرة. استعملت إحدى الدراسات الهنغارية تطبيق الطرائق الإحصائية لاستخلاص العبارات متعددة الكلمات(١٤). ونظراً

لتعقيد النحو الهنغاري، فقد استحدثت سلسلة وسائل معقدة للوصول لتلك الغاية. وقد استعملت هذه السلسلة من الوسائل بحيث يمكن استعمالها في لغات أخرى أيضاً.

أما في اللغة العربية فقد جمعت كمية من العبارات متعددة الكلمات من معاجم مختلفة (١٥). وقد صنفنا هذه العبارات وفق حالاتها الإعرابية. وقد أجري ترميز كل الكلمات المكونة للعبارات يدوياً مع تفاصيل المفهوم والوضع النحوي. وقد احتوت بعض التعبيرات على متغيرات لكي يمكن استبدال الكلمات في مواقعها بحيث تؤدي إلى الصيغة النحوية نفسها.

ولغرض استخلاص الكلمات المتعددة من الذخائر اللغوية (١٦)، فقد وضعت مواصفات للعبارات متعددة الكلمات. بعد ذلك وضع برنامج للاستخلاص ووضعت عدة محددات إحصائية لترشيح وحدات العبارات المشابهة لكي تكون الذخيرة على أفضل تمثيل ضمن المواصفات المطلوبة.

كما أجريت دراسة لمعرفة مدى الفائدة من المعاجم اللغوية من الانكليزية إلى العربية باستعمال كلمات متعددة مثل العبارات الشائعة والأفعال المرتبطة بحروف الجر والتركيبات المتعارف عليها (١٧). يمكن التعرف على الكلمات المتعددة من معانيها الصرفية وتركيباتها التي ربما تسبب ارتباكاً للمتدربين أو المتعلمين الأجانب للغة. لقد تم فحص الفائدة من معاجم الترجمة بالاستناد إلى ظاهرتين: الأولى بإيجاد فائدة هذه المعاجم بتقديمها أثناء تدريس طلبة الدراسات الأولية في الترجمة واللسانيات. الثانية بتقييم التحليل التركيبي لوثائق هذه المعاجم. وقد استنتج أن النسبة المئوية لاستعمال هذه المعاجم في الصفوف المتقدمة من أقسام الترجمة واللسانيات كانت أعلى، مما يشير إلى أن تبيان أهمية هذه المعاجم كانت أعلى. هذا بالإضافة إلى أن فصول الترجمة من العربية إلى الانكليزية كانت أقل استعمالاً للمعاجم من الترجمة من الانكليزية إلى العربية، وذلك لأن الطلبة كانوا يتعاملون مع لغتهم الأم، وأن العبارات متعددة الكلمات بالانكليزية أكثر صعوبة من العربية. كما وجد بأن فقرات الكلمات المتعددة بالعربية قد عوملت بشكل أفضل من قرينتها في المعاجم الانكليزية، هذا على الرغم من أن المعاجم من الانكليزية إلى العربية أكثر توفراً من المعاجم من العربية إلى الانكليزية.

١- ذخيرة القرآن الكريم

أنشأت ذخيرة القرآن الكريم (٣) المذكورة أعلاه باللغة الانكليزية نتيجة عمل تعاوني بمبادرة من جامعة ليدز ببريطانيا، مع ترميز متعدد الطبقات بضمنها ترميز أجزاء الكلام والتجزئة النحوية بالاعتماد على قواعد النحو العربي. إن الدافع وراء هذا العمل هو إنتاج مرجع يمكن الاعتماد عليه

للتمكن من تحليل إضافي لنص القرآن الكريم. وهذا المشروع يتوافق مع الشبكات الشجرية الأخرى المتوفرة في اللغة العربية لكي تكون أرضية مناسبة للوسائل اللغوية التراثية والمعروفة بالإعراب. تمت هذه الطريقة في ترميز الذخيرة اللغوية بطريقة تعاونية، وذلك بوضعها على الانترنت بعدة مراحل. كل مرحلة تضمنت ترميزاً آلياً ثم تدقيقاً يدوياً أولاً ثم التدقيق على الانترنت المراقب والتأكد بطريقة تعاونية. وقد اجتذب الموقع " الذخيرة العربية للقرآن " آلاف الزوار بينهم حوالي ١٠٠ متطوع قاموا بالترميز واقتراح تصحيحات على الترميز الأصلي. ولغرض الوصول إلى جودة عالية في العمل فقد أوكل التدقيق لمجموعة صغيرة من المتخصصين لتدقيق عمل المتطوعين. لقد استفيد من وجود خدمة كبيرة لإعراب القرآن الكريم عبر القرون والتي ساعدت في التدقيق. ويمكن الرجوع إلى تفاصيل الخطوات أعلاه في البحث(١٨).

٢- المعجم المفهرس للتراكيب المتشابهة في القرآن الكريم

كان المعجم المفهرس لكلمات القرآن الكريم هو تأليف محمد فؤاد عبد الباقي(١٩) والذي جمع الكلمات وفق جذورها أبجدياً، ولم يحتوي هذا المعجم على الحروف بل شمل الأسماء والأفعال فقط. أما المعجم الذي قام بوضعه مؤلف هذا البحث(٢٠) فينكون من جزأين: الجزء الأول مرتب حسب الجذور أبجدياً ويعطي كلّ الكلمات التي تقع تحت ذلك الجذر مع عدد تكراراتها. أما الجزء الثاني فهو للكلمة بترتيب هجائي ويعطي مواضعها في القرآن الكريم (السورة ورقم الآية). يبين الشكل(١) صورة بعض صفحة من الجزء الأول والشكل (٢) من الجزء الثاني.

يرق
إِسْتَرَقَ (1) التَّرَقَّ (2) التَّرَقَّ (1) تَرَقَّ (1) تَرَقَّه (1) وَأَتَرَقَ (1) وَاسْتَرَقَّ (1)
وَاسْتَرَقَّ (2) وَتَرَقَّ (1)
عدد الكلمات المختلفة = 9
عدد الكلمات الكلي لهذا الجذر = 11
يرك
السَّارِكَةُ (1) تَارَكْنَا (5) تَرَكَتُ (1) بُورِكَ (1) تَتَارَكَ (6) فَتَتَارَكَ (2) مِتَارَكًا (4)
مِتَارَكٌ (4) مِتَارَكَةٌ (1) مِتَارَكَةٌ (2) وَتَارَكَ (1) وَتَارَكْنَا (1) وَتَرَكَتُ (1) وَتَرَكَتُهُ (1)
وَتَتَارَكَ (1)
عدد الكلمات المختلفة = 15
عدد الكلمات الكلي لهذا الجذر = 32
يرع
أَتَرَعُوا (1) مِتَرَعُونَ (1)
عدد الكلمات المختلفة = 2
عدد الكلمات الكلي لهذا الجذر = 2

الشكل (١) صورة من صفحة من الجزء الأول من معجم كلمات القرآن

حرف اللام

- لؤلؤ
لؤلؤ مَكْنُونٌ: الطور/24 (1)
اللؤلؤ المَكْنُون: الواقعة/23 (1)
- لا
ولا ءاباؤكُم: الأنعام/91 (1)
ولا ءاباؤنا: الأنعام/148 النحل/35 (2)
ولا لآبائهم: الكهف/5 (1)
ولا ءاباؤنا ولا حرَمنا من: الأنعام/148 النحل/35 (2)
ولا يآب: البقرة/282/282 (2)
ولا تُؤثوا: النساء/5 (1)
لا تُؤثونهن: النساء/127 (1)
لا تاتيكُم: الأعراف/187 (1)
لا تاتينا: سبأ/3 (1)
لا تاتيهن: الأعراف/163 (1)

الشكل (٢) صورة من صفحة من الجزء الثاني من معجم كلمات القرآن

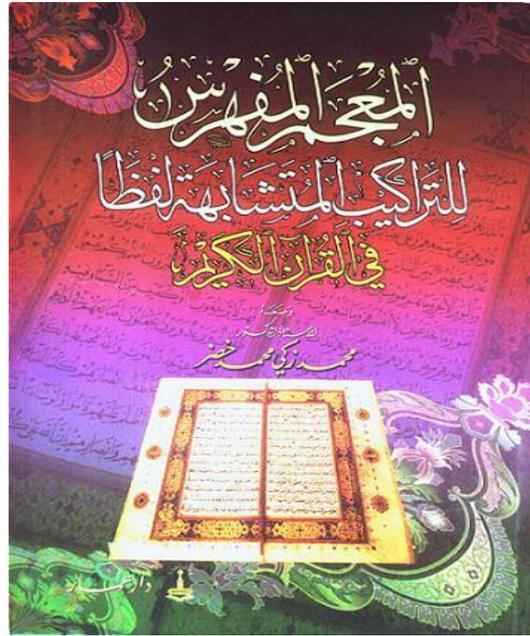
في عام ٢٠٠٢ صدر المعجم المفهرس للتراكيب المتشابهة لفظاً في القرآن الكريم (٢١). وحسب علم المؤلف لم يتوفر أي معجم لتراكيب القرآن المكونة من أكثر من كلمة قبله. استعملت كلمة "تراكيب" وليست "عبارات" نظراً لتجنب اعتبار مفهوم العبارات الذي له دلالة لغوية معينة. صنف هذا المعجم وفق الترتيب الهجائي لجذور الكلمات المكونة للتركيب. ويمكن الرجوع إلى تفاصيل خطوات عمل هذا المعجم في المرجع (٢٢). **مُلِكِ يَوْمِ الدِّينِ**

المعجم مرتب وفق جذر أول كلمة ثم جذر الكلمة الثانية وهكذا. أما إذا كانت الكلمة الأولى هي نفسها بين التراكيب فقد وضع الترتيب حسب ورودها في التسلسل في المصحف. أما إذا كان التشكيل مختلفاً فقد اعتبر تسلسل الكلمات حسب التشكيل: الفتحة ثم الضمة ثم الكسرة. وقد استعمل نص المصحف بما يقرب من الرسم العثماني قدر الإمكان، وهو يشابه خط مصحف المدينة فيما عدا استعمال الألف بدل الألف الخنجرية حيث لم تكن متوفرة في الخط الحاسوبي إلا مؤخرًا في ترميز (اليونيكود). فمثلا كلمات مثل " **مُلِكِ الْعُلَمِينَ الْكُتُبُ** " استعملت "مالك العالمين الكتاب" بينما أبقيت كلمات شائعة مثل: "ذلك هذا الرحمن" كما هي لشيوع استعمالها بهذه الصيغة. هناك كلمات في المصحف كتبت بأشكال مختلفة في مواضع مختلفة مثل: "رحمت رحمة" فقد أبقيت كما هي في كل موضع. كما اعتبرت أول آية من سورة الفاتحة هي البسملة: "بسم الله الرحمن الرحيم".

إن عدد آيات القرآن الكريم وفق مصحف المدينة المنورة هو ٦٢٣٦ آية. وبلغ عدد التراكيب التي تضمنها المعجم ٧٧٤٧٩ تركيباً. بلغت كلمات القرآن المختلفة ١٨٨٤١ كلمة. لكنه عند حذف الشدة على بداية الكلمات التي يسبقها تنوين على الكلمة السابقة لها انخفض العدد إلى ١٧٨٨٤ كلمة. هذا وقد استعملت جذور الكلمات من المصدرين ١٩ و ٢٠. مع العلم أن معظم كلمات القرآن من الأسماء والأفعال جذورها ثلاثية والقليل منها رباعية وليس في القرآن كلمة جذرها خماسي. أما الكلمات التي لا جذر لها مثل أسماء الأعلام مثل: "جبريل وميكايل وعيسى وموسى ونوح وآدم و مكة

ومصر ويثرب" فقد اعتبرت الكلمة نفسها محل الجذر. وكذلك الحال بالنسبة للحروف المقطعة في بدايات بعض السور. أما الحروف فقد اعتبرت نفسها جذورًا لكن مع جمع المتقارب منها مع بعض. فمثلاً جمعت: "أن" و"إن" وكذلك "لم" و"الن" و"لا". كما جمعت الكلمات التي وردت بإملاء مختلف مع بعضها مثل: "إبراهيم" و"إبراهيم" وبذلك أصبح عدد جذور الكلمات بعد أخذ هذه الملاحظات ١٧٦٨ جذرًا. وحيث أن المعجم هو للتركيب المكررة مرتين فأكثر، لذلك فإن كل كلمة وردت مرة واحدة في القرآن اعتبرت كلمة فاصلة بين عبارتين. وقد أخذ بعين الاعتبار التراكيب الممتدة لأكثر من آية، أي أن عبارة تحوي جذورًا مشابهة لعبارة أخرى يمكن أن تكون الأولى في آية واحدة بينما تكون الثانية في آيتين، وقد استعملت الإشارة # للتعبير عن الفاصلة بين آيتين.

كانت أطول عبارتين متطابقتين بالجذور مكونة من ٢٦ كلمة. وهما الآيات ٥-٨ من سورة المؤمنون والآيات ٢٩-٣٢ من سورة المعارج. ولم تؤخذ أي عبارة مشتركة عبر سور متعاقبة، بل اعتبرت كل سورة مستقلة في المعالجة. وهكذا فإن التراكيب المأخوذة بعين الاعتبار هي تلك التي تحوي كلمتين فما فوق وتتشرك بالجذور نفسها.



الشكل (٣) كتاب المعجم المفهرس للتركيب المتشابهة في القرآن الكريم

٣- خطة جمع التراكيب المشتركة في جذورها

اتباع الاسلوب الآتي في الحصول على التراكيب المشتركة في جذورها:

- ١- البدء من نص القرآن الكريم المحتوي على كلمات القرآن وما يوازيها من جذور مع رموز فواصل الآيات وفواصل السور.
- ٢- جُزئ النص إلى سلاسل طول الواحدة منها ٣٠ كلمة أي من الكلمة رقم ١ في السورة إلى ٣٠ ثم ٢ إلى ٣١ ثم ٣ إلى ٣٢ وهكذا. نقلت هذه السلاسل إلى ملف خاص مع سلسلة جذورها وأرقام الآيات والسور التي تعود لها. أما إذا كانت السورة تحوي كلمات أقل من ٣٠ فلم يستخلص منها أي سلسلة.
- ٣- بعد ترتيب السلاسل هجائياً ومقارنتها مع بعضها، لم يعثر على أي تطابق بين سلسلتين فأهمل الجميع.
- ٤- أنقص الرقم بواحد أي أعيدت العملية بتجزئة النص إلى ٢٩ وأعيدت العملية ثم ٢٨ ثم ٢٧ ولم يعثر على أي تطابق سوى عند الوصول إلى ٢٦ حيث عثر على الآيات: ٥-٨ من سورة المؤمنون: " وَالَّذِينَ هُمْ لِفُرُوجِهِمْ حَافِظُونَ (٥) إِلَّا عَلَىٰ أَرْوَاجِهِمْ أَوْ مَا مَلَكَتْ أَيْمَانُهُمْ فَإِنَّهُمْ غَيْرُ مَلُومِينَ (٦) فَمَنْ ابْتَغَىٰ وَرَاءَ ذَلِكَ فَأُولَٰئِكَ هُمُ الْعَادُونَ (٧) وَالَّذِينَ هُمْ لِأَمَانَاتِهِمْ وَعَهْدِهِمْ رَاعُونَ (٨) " مع الآيات ٢٩-٣٢ من سورة المعارج: " وَالَّذِينَ هُمْ لِفُرُوجِهِمْ حَافِظُونَ (٢٩) إِلَّا عَلَىٰ أَرْوَاجِهِمْ أَوْ مَا مَلَكَتْ أَيْمَانُهُمْ فَإِنَّهُمْ غَيْرُ مَلُومِينَ (٣٠) فَمَنْ ابْتَغَىٰ وَرَاءَ ذَلِكَ فَأُولَٰئِكَ هُمُ الْعَادُونَ (٣١) وَالَّذِينَ هُمْ لِأَمَانَاتِهِمْ وَعَهْدِهِمْ رَاعُونَ (٣٢).
- ٥- استمر إنقاص طول السلسلة واحداً واحداً إلى أن وصل طول السلسلة إلى كلمتين فقط. يبين الشكل (٤) المخطط الانسيابي لهذه العملية.
- ٦- جُمعت سلاسل الكلمات مع سلاسل الجذور من مختلف الأطوال ومواقعها في السور وأرقام الآيات ووضعت في قاعدة بيانات.
- ٧- أجريت عمليات ترتيب التراكيب وفق تسلسل جذورها هجائياً وجمعت التراكيب ذات الجذور المتشابهة كمجموعة واحدة ثم أعيد ترتيبها وفق تسلسل السور والآيات التي تقع أول واحدة منها فيها.
- ٨- إذا وجد أن عدد الكلمات في مجموعة التراكيب العائدة إلى الجذور نفسها مساوٍ لعدد مجموعة التراكيب لمجموعة أخرى لكن عدد كلمات الأولى أكبر من الثانية، حذفنا المجموعة ذات الكلمات الأقل وأبقي على المجموعة ذات عدد الكلمات الأعلى.
- ٩- أما إذا وجدت عبارة قصيرة تحوي عدداً أكبر من التراكيب المشابهة في جذورها من العبارة الأطول، اعتبرت المجموعات مختلفة أي أعطيت التراكيب القصيرة مع كل ما يتفق معها بالجذور ثم أعطيت التراكيب الطويلة مع ما يتفق معها بالجذور.

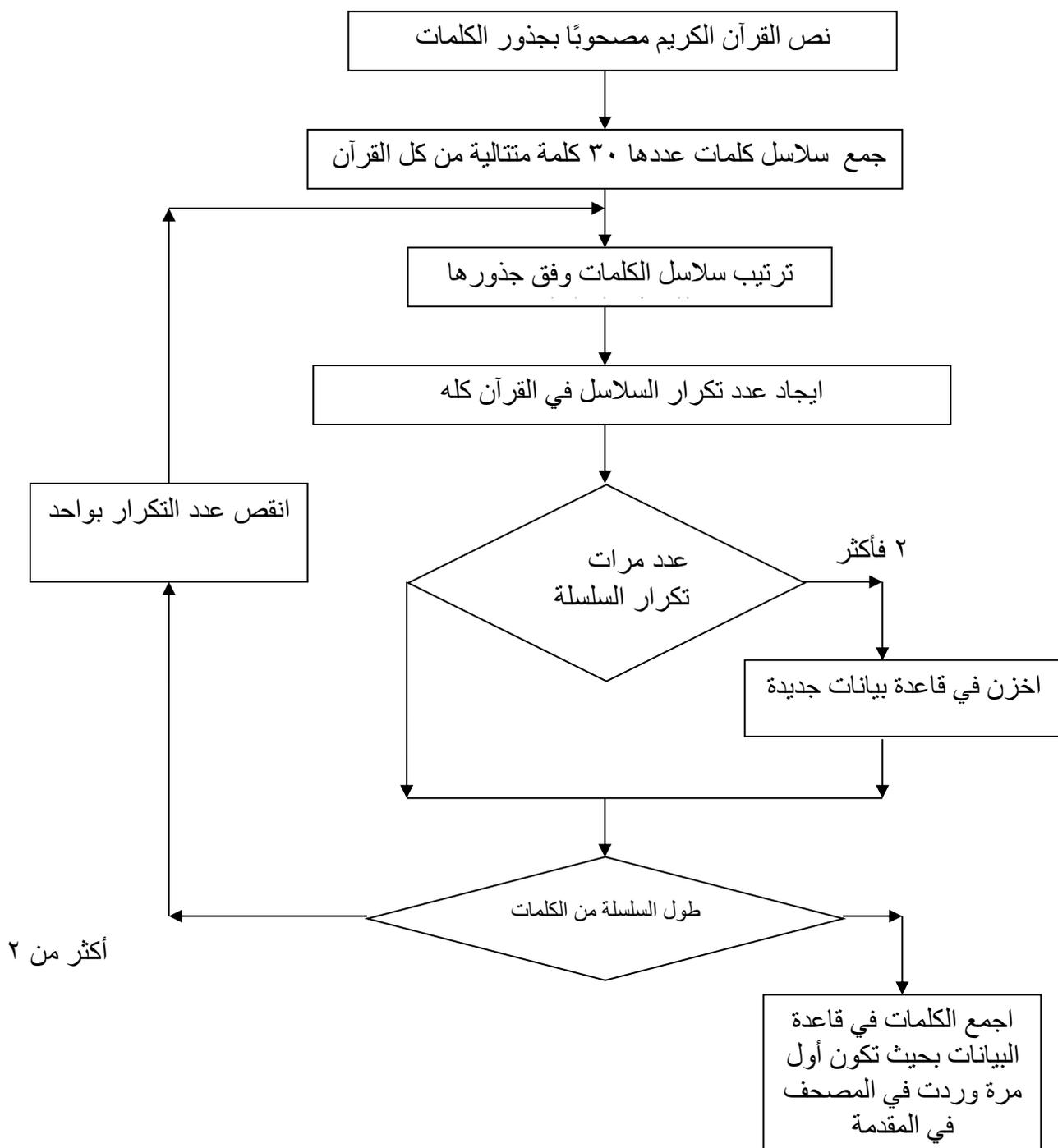
١٠- روعي في تسلسل التراكيب أن تكون التراكيب المتفقة بالكلمات مع الاتفاق بالتشكيل أولاً حسب تسلسل ورودها في المصحف ثم يليها التراكيب المتفقة بالكلمات لكنها مختلفة بتشكيل الآخر ثم التراكيب المتفقة بالجزور والمختلفة بالكلمات.

١١- أضيف في نهاية كل عبارة بعد ذكر محل ورودها في السور والآيات، عدد مرات تكرار التراكيب المتفقة في الجذر نفسه والكلمات نفسها والتشكيل نفسه.

١٢- صدر الكتاب بحيث جمعت التراكيب حسب تسلسل الجذور هجائياً كما يتبين من الشكل (٥)

١٣- إن البيانات التي جُمعت صُنفت بتسلسل آخر حسب ورود أول عبارة من التراكيب في المصحف. أي أن أول التراكيب هي التراكيب التي وردت في أول سورة الفاتحة ومع كل منها ما يتفق معها في الجذور. لكن العبارة حينما ترد في محلها في المصحف تذكر وحدها ويشار عندها إلى موضع أول مرة وردت، لكي يمكن الرجوع إلى ذلك الموضع ومن ثم العثور على كل التراكيب المصاحبة.

يبين الشكل (٤) المخطط الانسيابي الذي اتبع في الحصول على التراكيب.



الشكل (٤) المخطط الانسيابي لعمل برنامج الحصول على التراكيب

كما يبين الشكل (٥) نموذجاً هو أول صفحة من جداول الكتاب (٢٠)

حرف الهمزة	
أق: الألف: ١٧/	أق: من: الألف: ١٢٢/ الزخرف: ١٨/
أق: الواو: ٤٨/ (١)	أق: العين: (٢)
أق: الياء: ٧٧/ الألف: ١٢٦/ (٢)	أق: الألف: ٩١/ ٥١/ (٢)
أق: تم: الألف: ٢٦٠/ الألف: ١٠٠/	أق: الألف: ١٧/ (١)
أق: الألف: ١٨٥/ الألف: ٤١/ الحجر: ٧٠/	أق: الألف: ١١/ (١)
أق: الألف: ٤٨/ الألف: ٩٩/ طه: ١٣٣/	أق: الألف: ٢٤-٢٣/ (١)
أق: الألف: ١٩٧/ الألف: ٤٨/ (١)	أق: الألف: ٤/ (١)
أق: الألف: ٧٨/ الألف: ١٩/ (١)	أق: الألف: ٩/ (٢)
أق: الألف: ٣٧/ الألف: ٩٧/ (١)	أق: الألف: ٢١-٢٠/ (١)
أق: الألف: ٧٧/ الألف: ٧١/ (١)	أق: الألف: ٨/ (١)
أق: الألف: ٥٢/ الألف: ٢١/ (١)	أق: الألف: ٦٥/ (١)
أق: الألف: ٣٣/ الألف: ١٥/ (١)	أق: الألف: ٩٥/ الألف: ٧/ (٢)
أق: الألف: ١٩/ (٣٣)	أق: الألف: ٢/ (١)
أق: الألف: ٤١/ الألف: ٤٨/ (١)	أق: الألف: ٤٢/ (٢)
أق: الألف: ٩٩/ الألف: ٧/ (١)	أق: الألف: ٧٠/ (١)
أق: الألف: ١٩/ الألف: ٦٧/ (١)	أق: الألف: ١٠/ (١)
أق: الألف: ٧١/ الألف: ١٥/ (١)	أق: الألف: ٩٣/ (١)
أق: الألف: ٣٣/ الألف: ١٩/ (١)	أق: الألف: ٢٣/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ٦٢/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ٦٧/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ٦١/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ٩٤/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ٨٦/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ٨١/ (١)
أق: تم: الألف: ٧٧/ (١)	أق: الألف: ١٧/ (١)

الشكل (٥) نموذج من كتاب المعجم المفهرس للتراكيب المتشابهة لفظاً في القرآن الكريم

٤- الذخيرة اللغوية لتراكيب القرآن الكريم

أعيد تصميم البيانات التي احتواها الكتاب وفق الترتيب المذكور في الفقرة ١٣ أعلاه، لكي تكون مناسبة لوضعها كذخيرة لغوية تحوي مداخل كل تركيب منها على سطر مع مراعاة ما يأتي:

١- وُضعت التراكيب وفق ورود أول تركيب منها في المصحف. عند ورود تركيب جديد يدرج معه كل التراكيب التي تشترك معه في سلسلة الجذور بدءاً من التراكيب التي تتفق معه في الكلمات والتشكيل ثم التي تتفق في الكلمات وتختلف بالتشكيل ثم التي تختلف في الكلمات وتتفق بالجذور.

٢- كل مدخل يحوي معلومات عن محل وروده في المصحف RF والإشارة إلى أول تركيب ورد في المصحف متفق معه بالجذور RPH.

٣- يعاد التركيب عند وروده في محله من آيات وسور المصحف بحيث يشير إلى وجود مجموعة التراكيب المتفقة بالجذور عند أول ورودها لكي يكون بالامكان الرجوع إليها.

وقد أُشير إلى هذه المداخل بالرمز RF: REF

٤- كما أدرج نص التركيب بالرمز PH وجذوره RT

٥- وعلى هذا فإن كل مدخل يحوي ستة أرقام تحت رمز RF وستة أرقام تحت الرمز RPH ونص التركيب تحت الرمز PH ونص الجذور تحت الرمز RT. أي بالصيغة:

RF[A : B : C : D : E : F] RPH[A1 : B1 : C1 : D1 : E1 : F1]

[الجذور العائدة للتركيب] RT [نص-التركيب] PH

حيث اعتبرت التراكيب المتفقة بالجذور والكلمات والتشكيل من النوع الأول والمتفقة بالجذور والكلمات من النوع الثاني والمتفقة بالجذور فقط من النوع الثالث.

A: رقم السورة في المصحف بدءًا من سورة الفاتحة = ١ إلى سورة الناس = ١١٤

B: رقم الآية

C: رقم المجموعة من النوع الأول والتي تضم كل التراكيب المتفقة بالجذور في القرآن كله

D: رقم المجموعة من النوع الثاني والتي تضم كل التراكيب التي تتفق بالجذور والكلمات

E: رقم تسلسل المجموعتين من النوعين الثاني والثالث التي تقع ضمن المجموعة من النوع الأول

F: رقم تسلسل المدخل ضمن النوع نفسه

أما الرموز [A1 : B1 : C1 : D1 : E1 : F1] فهي نفس الرموز

[A : B : C : D : E : F] لكنها تعود إلى أول ورود للتركيب في المصحف.

وفيما يأتي مثال لأول مداخل في الذخيرة والتي تبدأ ببداية سورة الفاتحة:

[سمى ءله RT] بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ PH [1 : 1 : 1 : 1 : 1 : 1] RPH [1 : 1 : 1 : 1 : 1 : 1] RF [1 : 1 : 1 : 1 : 1 : 1]

[رحم رحم]

[سمى ءله RT] بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ PH [1 : 1 : 1 : 1 : 1 : 1] RPH [1 : 1 : 1 : 1 : 1 : 1] RF [27 : 30 : 1 : 1 : 1 : 2]

[ءله رحم رحم]

[سمى ءله RT] بِسْمِ اللّٰهِ PH [1 : 1 : 2 : 2 : 1 : 1] RPH [1 : 1 : 2 : 2 : 1 : 1] RF [1 : 1 : 2 : 2 : 1 : 1]

[سمى ءله RT] بِسْمِ اللّٰهِ PH [1 : 1 : 2 : 2 : 1 : 1] RPH [11 : 41 : 2 : 2 : 1 : 2] RF [11 : 41 : 2 : 2 : 1 : 2]

RF[27 : 30 : 2 : 2 : 1 : 3] RPH[1 : 1 : 2 : 2 : 1 : 1] PH [بِسْمِ اللَّهِ]RT[سمى ءله]

RF[5 : 4 : 2 : 3 : 2 : 1] RPH[1 : 1 : 2 : 3 : 2 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[6 : 138 : 2 : 3 : 2 : 2] RPH[1 : 1 : 2 : 3 : 2 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[22 : 28 : 2 : 3 : 2 : 3] RPH[1 : 1 : 2 : 3 : 2 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[22 : 34 : 2 : 3 : 2 : 4] RPH[1 : 1 : 2 : 3 : 2 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[22 : 36 : 2 : 3 : 2 : 5] RPH[1 : 1 : 2 : 3 : 2 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[6 : 118 : 2 : 4 : 3 : 1] RPH[1 : 1 : 2 : 4 : 3 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[6 : 119 : 2 : 4 : 3 : 2] RPH[1 : 1 : 2 : 4 : 3 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[6 : 121 : 2 : 4 : 3 : 3] RPH[1 : 1 : 2 : 4 : 3 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

RF[22 : 40 : 2 : 4 : 3 : 4] RPH[1 : 1 : 2 : 4 : 3 : 1] PH [اسْمُ اللَّهِ]RT[سمى ءله]

وفيما يأتي مثال ثانٍ لتبيان الرجوع إلى التركيب الأول:

RF[1 : 1 : 4 : 10 : 1 : 1] RPH[1 : 1 : 4 : 10 : 1 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

RF[1 : 3 : 4 : 10 : 1 : 2] RPH[1 : 1 : 4 : 10 : 1 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

RF[27 : 30 : 4 : 10 : 1 : 3] RPH[1 : 1 : 4 : 10 : 1 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

RF[41 : 2 : 4 : 10 : 1 : 4] RPH[1 : 1 : 4 : 10 : 1 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

RF[2 : 163 : 4 : 11 : 2 : 1] RPH[1 : 1 : 4 : 11 : 2 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

RF[59 : 22 : 4 : 11 : 2 : 2] RPH[1 : 1 : 4 : 11 : 2 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

RF[7 : 151 : 4 : 12 : 3 : 1] RPH[1 : 1 : 4 : 12 : 3 : 1] PH[الرَّحِيمُ الرَّحِيمُ] RT[رحم رحم]

RF[12 : 64 : 4 : 12 : 3 : 2] RPH[1 : 1 : 4 : 12 : 3 : 1] PH[الرَّحِيمُ الرَّحِيمُ] RT[رحم رحم]

RF[12 : 92 : 4 : 12 : 3 : 3] RPH[1 : 1 : 4 : 12 : 3 : 1] PH[الرَّحِيمُ الرَّحِيمُ] RT[رحم رحم]

RF[21 : 83 : 4 : 12 : 3 : 4] RPH[1 : 1 : 4 : 12 : 3 : 1] PH[الرَّحِيمُ الرَّحِيمُ] RT[رحم رحم]

REF: RF[1 : 3 : 4 : 10 : 1 : 2] RPH[1 : 1 : 4 : 10 : 1 : 1] PH[الرَّحْمَنُ الرَّحِيمُ] RT[رحم رحم]

الشكل (٦) مثال يبين الإشارة في محل لاحق لمدخل سبق وأن وردت مجموعته

فالسطر الأول يشير إلى المعلومات عن تركيب [الرَّحْمَنِ الرَّحِيمِ] والمعلومات عنه في الرقمين الأولين الذين يشيران إلى سورة الفاتحة=١ والآية=١ والرقم الثالث يشير إلى أن هذا هو التركيب الرابع في المصحف (الأول كان [بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ] والثاني [بِسْمِ اللَّهِ] والثالث [اللَّهُ الرَّحْمَنُ]) لذلك كان الرقم الثالث هو ٤ . أما الرقم الرابع أي ١٠ فيشير إلى رقم التركيب من أول المصحف إلى هذا التركيب بضمناها التركيبات المشتقة من التراكيب الثلاثة السابقة بتشكيل مختلف أو كلمات مختلفة ([اسْمُ اللَّهِ] و [اسْمُ اللَّهِ] و [اللَّهُ رَحْمَتٌ] و [اللَّهُ بِرَحْمَةٍ] و [اللَّهُ وَبِرَحْمَتِهِ] و [اللَّهُ وَرَحْمَةً]). أما الرقم الخامس فهو ١ لمجموعة [الرَّحْمَنِ الرَّحِيمِ] وهو ٢ لمجموعة [الرَّحْمَنُ الرَّحِيمُ] وهو ٣ لمجموعة [أَرْحَمُ الرَّاحِمِينَ]. أما الرقم الأخير فيشير إلى تسلسل التركيب ضمن كل مجموعة فهو ١ لتركيب [الرَّحْمَنِ الرَّحِيمِ] في الآية الأولى من الفاتحة وهو ٢ في الآية الثالثة وهو ٣ في السورة ٢٧ الآية ٣٠ وهو ٤ في السورة ٤١ الآية ٢. ويلاحظ أن الجزء الثاني:

RPH[1 : 1 : 4 : 10 : 1 : 1]

مشترك بين التراكيب الأربعة.

أما التركيب

REF: RF[1 : 3 : 4 : 10 : 1 : 2] RPH[1 : 1 : 4 : 10 : 1 : 1] PH[الرَّحْمَنِ

[رحم رحم] RT[الرَّحِيمِ

والذي يرد في الآية الثالثة من سورة الفاتحة (يقع في السطر ١٧٩ من الذخيرة) فيشير إلى الموقع نفسه المشار إليه (يقع في السطر ٢١) أي إلى:

RPH[1 : 1 : 4 : 10 : 1 : 1]

بلغ عدد المداخل أو التراكيب في الذخيرة كلها ١١٠٨٠٣ بينما بلغ العدد بدون تكرار ٤٥٦٣٧.

وبالطبع نشير مرة أخرى إلى أن كل كلمة أو تركيب ورد مرة واحدة في القرآن غير مشمول بهذه الذخيرة

٥- الاستنتاجات

إن ذخيرة التراكيب المكونة من أكثر من كلمة مهمة في معالجة اللغات الطبيعية، وبالأخص في الترجمة الآلية. هذا البحث يركز فقط على التراكيب من أكثر من كلمة في القرآن الكريم بالاستناد

إلى جذور الكلمات المكونة للتراكيب. التراكيب في البحث ثلاث أنواع: النوع الأول هو التراكيب المتشابهة بالكلمات مع تشكيلها والنوع الثاني الكلمات المتشابهة لكن تشكيلها مختلف والنوع الثالث المتشابهة في جذورها لكنها مختلفة الكلمات والتشكيل. فالأنواع الثلاثة تشكل مجموعة واحدة مشتركة بالجذور نفسها. الذخيرة متوفرة في الموقع:

mw.quran.corpus@al-mishkat.com

٥- المراجع:

[1] Mohammed Zeki Khedher, Multiword corpus of the Holy Quran, International Journal on Islamic Applications in Computer Science And Technology, Vol. 5, Issue 1, March 2017, 01-11

[2] Yoon, H., & Hirvela, A. (2004). ESL Student Attitudes toward Corpus Use in L2 Writing. Journal Of Second Language Writing, 13(4), 257–283. Retrieved 21 March 2012.

[3] <http://corpus.quran.com>

[4] Chris Greaves , Martin Warren, What can a corpus tell us about multi-word units? The Routledge Handbook of Corpus Linguistics, March 2010, ISBN: 9780415464895,

<https://www.routledgehandbooks.com/doi/10.4324/9780203856949.ch16>

[5] <http://people.cs.georgetown.edu/nschneid/p/sst-slides.pdf>

[6] Samah Meghawry, Abeer Elkorany , Akram Salah , and Tarek Elghazaly, SEMANTIC EXTRACTION OF ARABIC MULTIWORD EXPRESSIONS

[7] Douglas Biber, A corpus-driven approach to formulaic language in english: Multi-word patterns in speech and writing,

<https://nau.pure.elsevier.com/en/publications/a-corpus-driven-approach-to-formulaic-language-in-english-multi-w>

[8] Nathan Schneider, Lexical Semantic Analysis in Natural Language Text EXTENDED ABSTRACT

<http://www.cs.cmu.edu/~nschneid/thesis/longabstract.pdf>

<http://www.cs.cmu.edu/~nschneid/thesis/thesis-print.pdf>

[9] Judith Klavans and Christian Jacquemin. 1997. A natural language approach to multi-word term conflation. In Proceedings, DELOS Workshop on Cross-Language Information Retrieval, ETHZ, Zurich. ERCIM: European Consortium for Informatics and Mathematics.

[10] Juan Pedro Rica Peromingo, Corpus analysis and phraseology: Transfer of multi-word units, LINGUISTICS AND THE HUMAN SCIENCES, VOL 6, NO 1-3 (2010)

[11] Eric SanJuan1 and Fidelia Ibekwe-SanJuan, Multi Word Term queries for focused Information Retrieval ,

<http://dev.termwatch.es/cv/pub/ibesan10.pdf>

[12] Manawi: Using Multi-Word Expressions and Named Entities to Improve Machine Translation Liling Tan and Santanu Pal, <http://statmt.org/wmt14/pdf/W14-3323.pdf>

[13] Annelen Brunner and Dr Kathrin Steyer, Corpus-driven study of multi-word expressions based on collocations from a very large corpus , <http://www1.ids-mannheim.de/fileadmin/lexik/uwv/dateien/BrunnerSteyerBirmingham2007.pdf>

[14] Lexemes Kis, Balázs; Villada, Begoña; Bouma, Gosse; Ugray, Gábor; Bíró, Tamás; Pohl, Gábor; Nerbonne, John Rijksuniversiteit Groningen, A New Approach to the Corpus-based Statistical Investigation of Hungarian

Multi-word, Proceeding of the 4th International Conference on Language resources and Evaluation, Lisbon, Portugal, 2004

[15] Abdelati Hawwari, Kfir Bar, Mona Diab, Building an Arabic Multiword Expressions Repository

<https://aclweb.org/anthology/W/W12/W12-3403.pdf>

[16] Siham Boulaknadel, Beatrice Daille, Driss Aboutajdine, A multi-word term extraction program for Arabic language, LINA FRE CNRS 2729 University of Nantes, GSCM LRIT University Med V. 2 rue la Houssiniere BP 92208 44322 Nantes, B.P. .

[17] ALNASER, MOHAMMAD, Multi-word Items in Dictionaries from a Translator's Perspective, Ph.D Thesis Durham University,2010

[18] Kais Dukes1, Eric Atwell, Nizar Habash, Supervised Collaboration for Syntactic Annotation of Quranic Arabic, Language Resources and Evaluation Journal, 2011

[19] محمد فؤاد عبد الباقي: المعجم المفهرس لألفاظ القرآن الكريم- دار إحياء التراث العربي

[20] محمد زكي خضر ، معجم كلمات القرآن – www.al-mishkat.com/words

[21] محمد زكي خضر ، المعجم المفهرس للتراكيب المتشابهة لفظاً في القرآن الكريم – دار

عمار – الأردن ٢٠٠٢

[22] محمد زكي محمد خضر ، الجوانب البرمجية في إعداد المعجم المفهرس للتراكيب

المتشابهة لفظاً في القرآن الكريم – دراسات – الجامعة الأردنية – المجلد ٣١ العدد ١, ٢٠٠٤