

العنوان:	استخراج المتلازمات اللفظية من النصوص العربية باستعمال الأداة "غايث": تطبيق على النص القرآني
المصدر:	أعمال اليوم الدراسي : المحتوى الرقمي بالعربية في نظام الإدارة الإلكترونية - المجلس الأعلى للغة العربية - الجزائر
المؤلف الرئيسي:	زايدي، صورية
مؤلفين آخرين:	العسكري، محمد الطيب، عبدعلي، أحمد(م. مشارك)
محكمة:	نعم
التاريخ الميلادي:	2011
مكان انعقاد المؤتمر:	الجزائر
الهيئة المسؤولة:	المجلس الأعلى للغة العربية
الشهر:	مارس
الصفحات:	1 - 18
رقم MD:	806102
نوع المحتوى:	بحوث المؤتمرات
قواعد المعلومات:	HumanIndex
مواضيع:	القرآن الكريم، اللغة العربية، المعالجة الآلية، المتلازمات اللفظية
رابط:	http://search.mandumah.com/Record/806102

استخراج المتلازمات اللفظية من النصوص

العربية باستعمال الأداة "غايت"

(تطبيق على النص القرآني الكريم)

صورية زايدي¹, محمد الطيب العسكري¹, أحمد عبد العلي²

soraya.zaidi@univ-annaba.org, laskri@univ-annaba.org,
aabelal@nmsu.edu

¹ قسم الإعلام الآلي جامعة باجي مختار عثابة- الجزائر

² قسم الاعلام الآلي جامعة نيو مكسيكو الولايات المتحدة الأمريكية

الملخص

إنّ كمّية المعلومات التي ما فتئت تتزايد في شتى الميادين، ولدت لدى الباحثين حاجة ماسة إلى هيكلة وتنظيم المحتوى الرقمي للوثائق المختلفة. فبينما تتوفر أدوات عديدة مجانية الاستعمال والمستخدم في استخراج المصطلحات من النصوص اللاتينية كالفرنسية والانجليزية، نلاحظ أن اللغة العربية تفتقر لمثل هذه الأنظمة، الشيء الذي اضطرنا إلى محاولة تكيف البعض منها بغية تطبيقها على النصوص العربية، رغم أنه في أغلب الحالات يتطلب منا القيام بكثير من التغييرات المعقدة للتوصل إلى نتائج مقبولة.

في هذا البحث، نقدم أداة أنشئت في الأصل لمعالجة نصوص اللغة الانجليزية، لكننا قمنا بتكييفها بغية استعمالها في معالجة النصوص العربية وبالضبط في الاستخراج الآلي للمتلازمات اللفظية الخاصة بميدان محدد كالطب، الاقتصاد، التاريخ، الدين، الرياضة الخ...

تستعمل هذه المصطلحات في بناء انطولوجيا يمكن توظيفها في تطبيقات مختلفة أهمها تحسين البحث على الشبكة العنكبوتية. وإن كنا قد بدأنا بتطبيق هذه الطريقة على النص القرآني الكريم، لقدسيته، فإنه يمكن استعمالها على أي ميدان آخر.

الكلمات الجوهرية: معالجة المدونات، المصطلحات النصية، الألفاظ المتلازمة، اللغة العربية، استخلاص المصطلحات، استخلاص المعلومات.

1. مقدمة

تعد العربية اللغة الأم لأكثر من 300 مليون نسمة وهي تنتشر في أكثر من 22 دولة عبر العالم. وحتى ظهور الإسلام في شبه الجزيرة العربية في القرن السابع ميلادي، كانت اللغة العربية شفاهية (أي لغة كلام أساسا). وجاء تنزيل القرآن الكريم على النبي محمد صلى الله عليه وسلم بالعربية ليفضي عليها صبغة دينية خاصة وكان هذا الدافع الأساسي الذي حمل هذه اللغة خارج مجالها الجغرافي وانتشارها في جميع أنحاء العالم والاهتمام بها في العديد من المجالات وخاصة منها ميدان المعلوماتية. وتعتبر اللغة العربية لغة معقدة في ميدان المعالجة الآلية وذلك بسبب خصائصها الصرفية والنحوية.

مع انتشار استعمال شبكة الإنترنت بشكل عام والواب الدلالي على وجه الخصوص، أصبحت بعض مهام الحوسبة كالبحث عن الوثائق الإلكترونية أو فهرستها، أو إمكانية إعادة استعمال الخبرات والمعارف تستوجب تطوير العديد من البرامج القادرة على إدارة المعرفة المتوفرة باللغة الطبيعية، و لبلوغ ذلك نحتاج إلى موارد لسانية مثل قواعد مصطلحات، ذخائر، قواميس مرادفات أو انطولوجيات والتي يتم إنشاؤها انطلاقا من

مدونات تحتاج بدورها إلى معالجة نصوصها قبل عملية الإنشاء في حد ذاتها.

نقدم في هذا البحث طريقة لسانية لاستخراج المتلازمات اللفظية والتي تتميز بكونها ظاهرة لغوية عالية الإنتاجية تجعل لفظتين أو أكثر، في توارد ما، متلاصقتين ومتضامتين ومجتمعتين بشكل دائم لا يتغير وذلك راجع إلى الاستعمال التداولي الخاص لتلك المتلازمات مثلا تقوم مقولة الصفة في اللغة العربية وفق اختيارها التوزيعي بتحديد عناصر معينة مع عدد قليل لتتألف مع الأسماء الخاصة التي تناسبها، نفس الشأن يحصل مع المقولات الأخرى كالأسماء والأفعال [18] وذلك آليا من المدونات العربية قصد استعمالها كمصطلحات لبناء انطولوجيا في ميدان معين.

هيكلا هذا البحث على النحو التالي: القسم الثاني خصص للتعريف باستخراج المعلومات بصفة عامة والقسم الثالث يشمل أعمالا ذات صلة نتطرق فيه إلى أهم الأدوات المتوفرة والمستعملة في استخراج المصطلحات من النصوص بطريقة آلية، أما الجزء الرابع فخصصناه لتكييف الأداة غايت لاستخراج الألفاظ المتلازمة من النص القرآني الكريم، يتبع بعملية تقييميه للنتائج المحصل عليها وفي الأخير نعطي خاتمة مع بعض الآفاق.

2. استخراج المعلومات

1.2. استخراج المعلومات

هو نوع من استرجاع المعلومات هدفه استخلاص معطيات من وثائق غير مهيكلة مقروءة آليا، عن طريق معالجة نصوص باللغة الطبيعية. ونظرا لصعوبة العملية، فإن النهج الحالي لاستخراج المعلومات يتركز على مجالات محدودة ضيقة، كالاعتماد على تقارير وكالات الأنباء والهدف

غالبا ما يكون استعمال هذه المعطيات لتأهيل قواعد المعطيات أو الأنطولوجيات [19].

من بين المهام الفرعية لاستخراج المعلومات :

* استخراج المسميات: كأسماء أشخاص، منظمات، أماكن، عملات، تواريخ، أعداد إلخ.

* استخراج المصطلحات: التعرف على المصطلحات الخاصة بميدان معين كالطب، الرياضة، الدين إلخ.

* استخراج العلاقات : تحديد العلاقات بين المسميات أو المصطلحات.

* التعرف على الضمائر التي تعوض أو تدل على مسميات.

2.2. تصنيف أنواع المتلازمات اللفظية :

أ- التعبيرات المتلازمة المفتوحة *Open collocations*

وهو تلازم لفظتين أو أكثر لتظهرها في تعبير واحد، من دون أية خصوصية علائقية بين هذين اللفظتين مثل: انتهت الحرب.

ب- التعبيرات المتلازمة المقيدة *Restricted collocations*

وهو تلازم مفردتين أو أكثر لتستعمل في سياق مطرد، وليس في استعمال مسكوك، بحيث تتبع نموذجها البنيوي وتتقيد بتأليفها الاستبدالي نحوا ودلالة وتداولاً نحو: جريمة نكراء.

ج . التعبيرات المتلازمة المتصلة *Bound collocations*

هذا الصنف من التعبيرات المتلازمة هو جسر بين المتلازمات والمسكوكات، ويعتبر العنصر القوي في درجة المسكوكية هو الذي يتم اختياره عن الآخر كما ذكر إميري في مثال: أطرق الرأس

د- التعبيرات المتلازمة المسكوكة *Idioms*

يتميز هذا الصنف بعتمته وصممه وثباته القوي أمام أي استبدال وهو غالبا ما يستعمل في السياقات والمعاني الخاصة بحيث تشكل جميعها وحدة دلالية خاصة بها.

إن معرفة المتلازمات ضرورة ملحة ما دامت ظاهرة مشتركة بين جميع اللغات الطبيعية، فبالنسبة للغة الإنجليزية مثلا هي موضوعة في قواميس خاصة لمساعدة المتعلمين والقراء في تصحيح أخطائهم[18].

3.2 أنواع المتلازمات المقيدة:

لاحظ /يميري [18] بخصوص هذا الصنف من المتلازمات ظهورها في بنيات تركيبية صورية مثل :

- فعل /اسم
- فعل /اسم / فضلة
- اسم /صفة

وعليه تم تصنيفها في قاموس بي بي إي كما يلي وفق 7 تصنيفات صورية أساسية:

- (1) فعل + اسم / فضلة
- (2) فعل + اسم
- (3) صفة + اسم
- (4) اسم + فعل
- (5) وحدة مجمعة مع الاسم
- (6) ظرف + صفة
- (7) فعل + ظرف

يمكننا اعتبار التراكيب التالية و التي سنعتمدها في بحثنا هذا:

■ اسم + صفة

■ اسم + اسم

■ فعل + اسم

■ اسم + فعل

■ اسم + أداة + اسم

استعملنا هذا النوع من الأنماط لأنها تعتمد على ترميزات نحوية وبما أن مدونة القرآن الكريم مرمّزة نحويًا فإنها تمكننا من التعرف على المتلازمات اللفظية.

مثال على ذلك:

اسم_اسم مثل (صوت_النبي)،

اسم_صفة مثل (خير_كثير)،

اسم - أداة - اسم مثل (سلالة_من_طين)، إلخ.

نهتم في هذا البحث باستخراج هذا النوع من المتلازمات اللفظية الثنائية أو الثلاثية.

3. أعمال ذات صلة

نقدم في ما يلي بعض الأدوات المستخدمة على نطاق واسع لاستخراج المصطلحات من نصوص لاتينية، والتي يمكن تكييفها مع اللغة العربية، لأنه ولحد كتابة هذا البحث لا توجد أدوات متخصصة في استخراج المصطلحات من النصوص العربية [9].

1.3. أكابيت (ACABIT)

هو أداة استخراج المصطلحات من مدونات فرنسية مرمّزة (أي ألحقت بكل كلمة منها فئتها النحوية) وذلك باستخدام مقارنة لسانية ، طورتها بياتريس داي لشركة آي بي إم [12]. وهناك عمل قامت به سهام بولقنادل

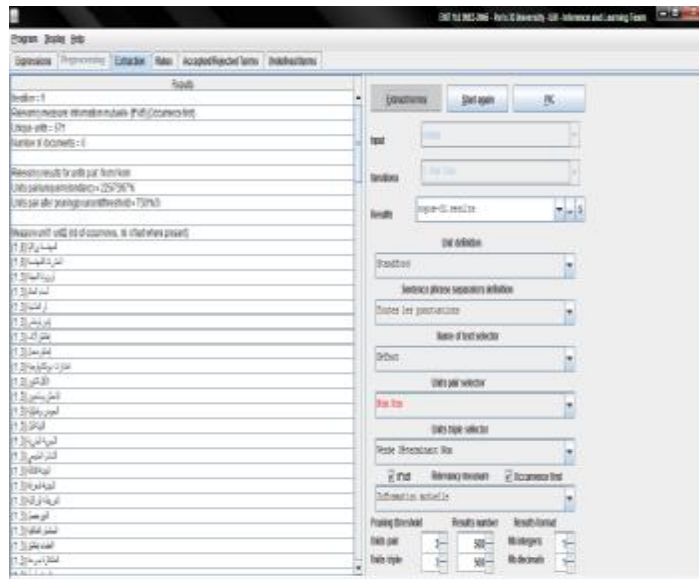
لتكليفه مع اللغة العربية و لكن لحد الساعة لا يمكن تحميله أو استعماله لأغراض البحث.

2.3. أنا (ANA)

هو أداة استخراج مصطلحات_مرشحة بدون أي تحليل لغوي. ولكن تتم عملية التعرف عليها بالاعتماد على مشاهدة تكرار الكلمات في النصوص المستعملة أي بمقاربة إحصائية ولذلك يسهل استخدامها مع أي لغة كانت[13].

3.3. إكسيت (EXIT)

هو نظام نصف_آلي لاستخراج المصطلحات من مدونات متخصصة. وينبغي أن يتم استخدام إكسيت من قبل مستعملين ذوي خبرة في المجال المحدد، إذ يجب أن تكون لديهم معرفة كبيرة بالمفاهيم المستخدمة في تحليل النصوص من أجل التعرف على الألفاظ المتلازمة، وثيقة الصلة بالموضوع من بين تلك المستخرجة من النصوص وتعتمد الأداة على مقاربة إحصائية-لغوية [16]، قمنا بتجريبها على نصوص عربية بعد تعديلها واستعمال رمز خاص (يعتمد نصوص عربية عوضت حروفها بحروف لاتينية)



الشكل 1: استخراج المتلازمات اللفظية بواسطة إكسيت

4.3. نوج و يونيتاكس (NOOJ, UNITEX)

كما يمكن ذكر بعض الأدوات التي يمكن استخدامها في مراحل المعالجة أو استخراج المعلومات وكذا تأهيل الانطولوجيات بالعربية نخص بالذكر منها نوج ويونيتاكس اللذان يتوفران على فهرس (باحث كلمات معينة وعرضها داخل الجملة التي وقعت فيها) والتي تسمح باستخراج مصطلحات على شكل متلازمات لفظية [16].

4. استعمال " غايت " لاستخراج المتلازمات اللفظية من النص القرآني الكريم

بعد عرض بعض الأدوات التي يمكن استعمالها في استخراج المعلومات بالعربية، نقدم الآن الأداة غايت التي تم اعتمادها في بحثنا هذا وذلك لقوتها واستعمالاتها المتعددة من طرف عدد كبير من المؤسسات والمنظمات العالمية إلى جانب مخابر الأبحاث العلمية عبر العالم وكذا سهولة تكييفها مع النصوص العربية.

1.4. الأداة غايت¹ (GATE)

هي أداة لتطوير ونشر وحدات ومكونات برمجية مختصة في المعالجة الآلية للغة الطبيعية. يمكن لـ غايت مساعدة الباحثين والمطورين بعدة طرق منها:

- تحديد البنية أو الهيكل التنظيمي لبرامج المعالجة الآلية للغة
- توفير إطار عمل مع مخازن بيانات، يسمح باستخدام ودمج النظام في تطبيقات متنوعة

¹ يمكن تحميل غايت على الصفحة: <http://gate.ac.uk>

- توفير بيئة تطوير مكونة من أدوات مناسبة لتطوير عناصر جديدة [10].

بنية غايت تعتمد على مكونات على شكل وحدات برمجية قابلة لإعادة الاستخدام يمكن استعمالها ونشرها في سياقات مختلفة ومتنوعة [11]، من أهم هذه المكونات:

1.1.4 كريول (CREOLE): هي مجموعة من المكونات يمكن إعادة

استخدامها في معالجة اللغة الطبيعية، و يمكن تقسيمها إلى ثلاثة أنواع:

1. الموارد اللغوية: كالمستندات، المدونات، الأنطولوجيات إلخ.
2. الموارد البرمجية : كالمحلات والمقطعات إلخ.
3. الموارد البيانية: وهي مكونات تسمح بعرض ومعالجة أنواع أخرى من الموارد.

2.1.4 آني (ANNIE): وهو نظام طُوّر أساسا لاستخراج المعلومات،

وهو العنصر الرئيسي في غايت. مكونات آني تشكل سلسلة من البرمجيات بما في ذلك محلّ ، معجم جغرافي ، مقطعّ ، مرّمز إلخ [2].

الأنظمة التي تعتمد غايت واسعة الاستعمال قد نجدها في الشركات التي تهتم بتحليل آراء ورد فعل مستخدمي العديد من شركات النقل في المملكة المتحدة وتحليل مشاعر الزبائن في بعض المؤسسات بنيويورك ، كما يتم استخدامه في العديد من المشاريع كالبحوث الخاصة بالسرطان، التنقيب عن المعلومات الشخصية على الشبكة العنكبوتية والإعلانات عن عمليات التوظيف وتحليل السير الذاتية. ويستعمل النظام Text2Onto غايت في عملية تحليل النصوص لبناء الانطولوجيات آليا.

3.1.4 لغة جايب (JAPE language)

هي لغة تعتمد جافا وتقدم محولات ذات حالات منتهية، تسمح بعملية وضع الهوامش أو التوسيمات على النصوص. القواعد النحوية لجايب تتكون من مراحل كل واحدة منها عبارة عن مجموعة من (نمط /أثرالقاعدة) يتم تشغيل هذه المراحل بطريقة متسلسلة. الجهة اليسرى للقاعدة تشمل وصف النمط بينما تشمل الجهة اليمنى تسمية الهوامش أو التوسيمات[4].

2.4. استخراج المسميات العربية باستعمال غايت

استعملنا في بداية الأمر غايت على مدونتين لاستخراج المسميات العربية: المدونة الأولى هي مدونة اللغة العربية المعاصرة المطورة من طرف لطيفة السليطي¹ في جامعة ليدز [5] و التي تتكون من أكثر من 843 000 كلمة موزعة على 416 وثيقة والتي تغطي مجموعة واسعة من الفئات منها الدين، الرياضة، العلوم، الاقتصاد، قصص الأطفال الخ. والثانية مدونة الهلال للقرآن الكريم² المطورة من طرف قيس ديوكس بنفس الجامعة تحت إشراف إريك أتوال والذي يتألف من النص القرآني المرمز حسب الكلمات، الآيات والسور[6].

1.2.4. موارد معالجة اللغة العربية المستخدمة في غايت

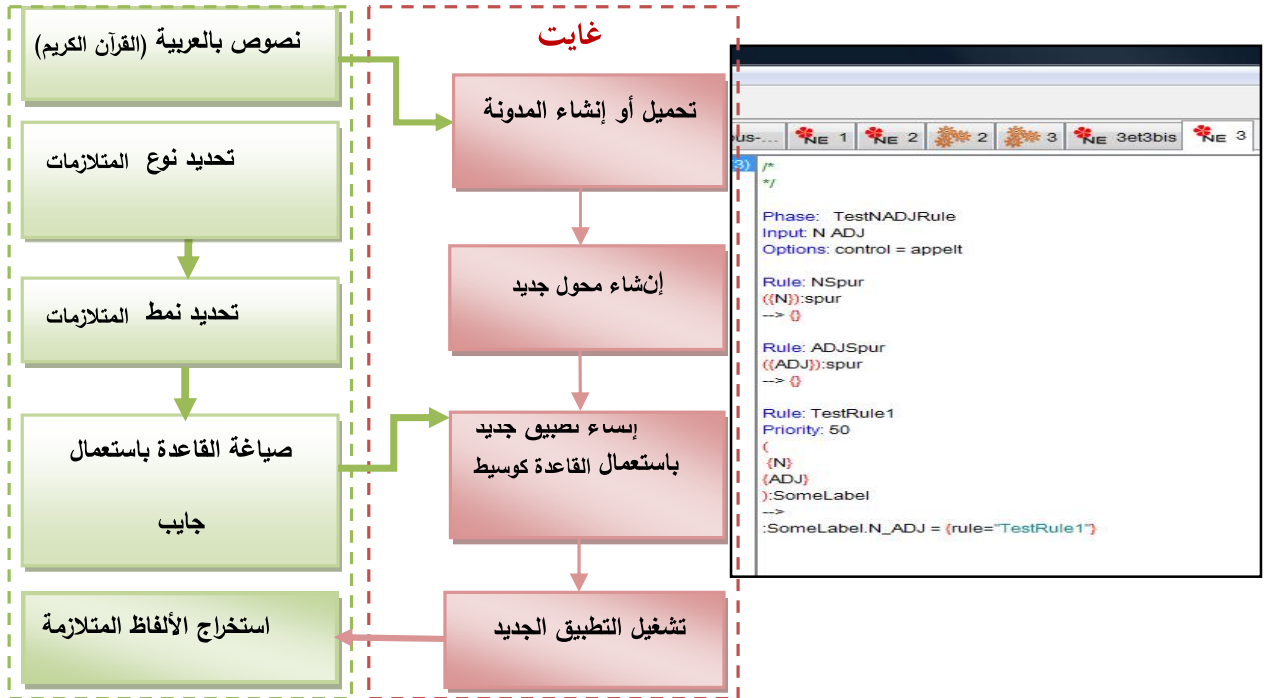
- المعجم الجغرافي: هي قوائم على شكل ملفات نصوص عادية، كل قائمة تمثل مجموعة من الأسماء مثل أسماء أشخاص، مدن، منظمات، أماكن، عملات الخ.
- القواعد الأساسية للعربية: تسمح باستخدام الملفات التي تحتوي على قواعد مختلفة.

تحميل المدونة¹ على الصفحة <http://www.comp.leeds.ac.uk/eric/latifa/research.htm>

² يمكن تحميل المدونة² على الصفحة: <http://quran.uk.net>

بعد اختيار نوع المتلازمات نقوم بوضع النمط الذي نريد استخدامه ، مثلا اسم- صفة، نكتب القاعدة بلغة جايب والتي تسمح بالتعرف على هذا النمط ثم نقوم بإنشاء محول جديد ذو حالات منتهية، حيث يتم تمرير القاعدة الجديدة كوسيط أو معلمة للمحول[2] ، بعد استخراج المتلازمات اللفظية يمكن استعمالها فوراً أو الاحتفاظ بها في مخزن معطيات لاستغلالها في تطبيقات أخرى.

قمنا بتطبيق هذه الطريقة على النص القرآني الكريم لاستخراج المتلازمات اللفظية والهدف من ذلك هو إعداد قائمة ألفاظ متلازمة قصد استعمالها كمصطلحات في البناء الآلي لانطولوجيا القرآن الكريم. يمكن استعمال هذه الانطولوجيا في تطبيقات عديدة مثل اعتمادها في تحسين البحث في معاني القرآن الكريم على الشبكة العنكبوتية، الترجمة الآلية لمعاني القرآن أو الفهرسة الآلية.



الشكل 3: قاعدة بسيطة لاستخراج اسم-صفة

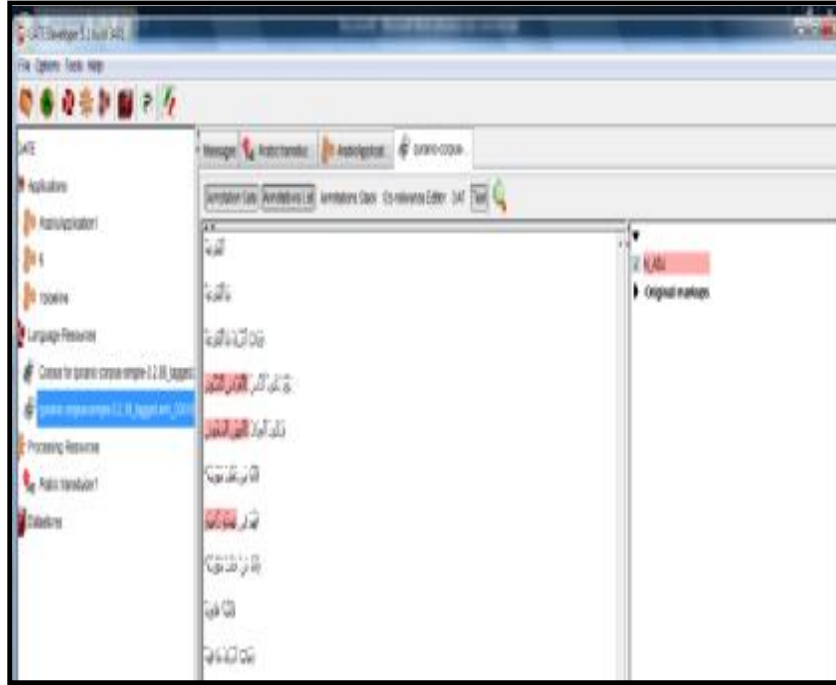
الشكل 4: مراحل استخراج المتلازمات اللفظية

بنفس الطريقة نستطيع كتابة قواعد إضافية بغية التعرف على أنواع أخرى من الألفاظ المتلازمة. يمكن الاستعانة بخبراء في الميدان للتحقق من النتائج المحصل عليها لقبول أو رفض الألفاظ المستخرجة، ذلك لأنه من الممكن الحصول على تتابع ألفاظ لا تمثل في الحقيقة متلازمات لفظية.

5. نتائج وتقييم

قمنا بإتباع المراحل المبينة في الشكل 4 لاستخراج الألفاظ المتلازمة من القرآن الكريم.

فمن بين ما استخرج مثلاً في سورة القارعة، الشائيات التالية:
الفراش_المبثوث، العهن_المنفوش، عيشة_راضية، وهي على شكل:
اسم_صفة، كما هو موضح على اليسار في الشكل 5.



الشكل 5: استخراج و عرض الألفاظ المتلازمة

عملية تقييميه بسيطة تمثلت في ترميز يدوي لجزء من القرآن و مقارنته مع النتائج المحصل عليها بواسطة الأداة
أنوطايشن_ديف [17]، [8] ، أعطت دقة تقارب 0.6، نطمح في تحسين هذه النتيجة وذلك بأخذ الخصائص الصرفية بعين الاعتبار عند صياغة القواعد المختلفة.

6. خاتمة و آفاق

تطرقنا في هذا البحث إلى إشكالية استخراج المتلازمات اللفظية من النصوص العربية، حيث قمنا أولاً باستعمال النظام غايت بموارده الخاصة كآني و كريول لاستخراج المسميات التي عامة ما تستعمل في تأهيل الانطولوجيات. ثم فكرنا في إثراء النظام غايت لاستخراج المتلازمات اللفظية وتم ذلك بصياغة قواعد جديدة تمرر كوسائط أة معلومات لمحاولات ذات حالات منتهية يتم إنشاءها بواسطة غايت وقد قمنا بتشغيله وتطبيقه

على النص القرآني الكريم. ونعمل حالياً من جهة على تحسين النتائج المحصل عليها، ومن جهة أخرى على استخراج العلاقات الدلالية لربط هذه المصطلحات على شكل أنطولوجيا القرآن الكريم.

7. المراجع

- [1] كونيغهام، ح. 1999 آني : استخراج الكينونات الاسمية العربية، استخراج المعلومات -- دليل المستخدم (الطبعة الثانية).
- [2] كونيغهام، ح. ، مينارد ، د. ، بونتشيفا، ك. ، طابلان ، ف. ، ديميتروف ، م. ، الأسواني ، ن. ، روبرتس ، إ. ، 2006 ، تطوير مكونات معالجة اللغة مع غايت الإصدار 1.3 (أ دليل المستخدم).
- [3] بلامونون، ل.، 2004 هندسة اللغة بوتسطة غايت، رالى / ديرو ، جامعة مونتريال.
- [4] ثاكر، د.، السمان، ط.، لاكن، ب.، 2009 ، البرنامج التعليمي لقواعد جايب في غايت، الإصدار 0.1 ، ب أ. فوطو، والمملكة المتحدة.
- [5] السليطي ، ل.، أتويل، إ.، 2004. تصميم و تطوير مدونة اللغة العربية المعاصرة. في وقائع المؤتمر السادس للمعالجة الآلية للغة. غرناطة ، اسبانيا ، ص. 92.
- [6] ديوكس، ق.، أتويل ، إ. ، شرف ع.ب.، 2010 ، مبادئ نحو الترميز النحوي لمدونة القرآن ، في المؤتمر الدولي السابع للموارد اللغة والتقييم فاليوتا ، مالطا.

- [7] عماري، س.، 2009 ، استخراج الكينونات الاسمية من نصوص عربية باستعمال غايت ، مذكرة ماستر جامعة باجي مختار عنابة.
- [8] مينارد، د.، أسواني، ن.، 2009 ، توسيم وتقييم ، والبرنامج التعليمي للمدرسة الصيفية، شفيد، المملكة المتحدة 2009.
- [9] زايدي، ص. ،العسكري، م-ط.، 2009، استعراض أدوات استخراج المصطلحات النصية العربية لبناء الانطولوجيات، المؤتمر الدولي (MIC) سوسة ، تونس.
- [10] غايت: دليل المستخدم على الصفحة <http://gate.ac.uk/sale/tao>
- [11] كونيغهام وآخرون. (LREC2000) تصنيف مكونات هندسة اللغة ومناقشة التحليل اللازم ل غايت لإصدار 2.
- [12] داي، ب.، سرياح، ف.، هيكلية الخدمات من أجل تخصيص عملية اكتساب المصطلحات. المعالجة الآلية للغات: 47(3) 39-61 2006
- [13] أنغهارد، ك.، الاكتساب الطبيعي الآلي لشبكة دلالية، رسالة في مراقبة الانظمة، جامعة التكنولوجيا بكمبياني، 1992.(ANA)
- [14] روش، م.، هايترز، ط.، ماتي-طايب، أ.، كودراطوف، إ.، إكسيت: نظام تكراري لاستخراج المصطلحات من مدونات متخصصة، 2004.
- [15] مونداري، و الاخرون، بناء أنطولوجيات من النصوص: مرحلة التصور المفاهيمي، الايام الفرنكوفونية الـ 19 لهندسة المعارف، نانسي، فرنسا 2008.
- [16] زايدي، ص.، العسكري، م-ط.، عبد العالي، أ، دراسة مدى تكيف بعض أدوات المصطلحات النصية مع العربية، الندوة السابعة حول تحسين الانظمة المعلوماتية، ورقلة 18- 20 أبريل 2010.

- [17] كنفهم، ح. والآخر غايت: 1 هيكل لتطوير تطبيقات HLT قوية.
- [18] عز الدين غازي المتلازمات اللفظية في العربية المعاصرة ومعالجتها في القواميس، 2007 rhazi@maktoob.com
- [19] كوي، ج.، ويلكس، ي.، استخراج المعلومات، (في مطبوعات ويلكس، ي.) عدد خاص ب المعالجة الآلية للغة (ACM) 1996.

المصطلحات المترجمة المستعملة

انطولوجيا: ontologie

Collocations	المتلازمات اللفظية، المتلازمة
Peuplement	تأهيل
Extraction d'information	استخراج المعلومات
Recherche d'information	استرجاع المعلومات
Entités nommées	المسميات
Datastore	مخازن بيانات
Transducteur à états finis	محولات ذات حالات منتهية
Annotations	هوامش توسيمات
GATE (General Architecture for Text Engineering)	غايت
Patron	نمط
Règle	قاعدة

المحتوى الرقمي بالعربية في نظام الإدارة الالكترونية

Candidats-termes	مصطلحات مرشحة
Paramètres	وسائط أو معلمات
Concordancier	مفهرس