البرامج الحاسوبية وتطبيقاتها على التحليل المعجمي للمتون اللغوية

د. عبد الرحيم ناجح
جامعة شعيب الدكالي، المغرب

ملخص:

تكمن أهمية البحث في التعريف ببعض البرامج الحاسوبية وطرق استخدامها في التحليل المعجمي للمتون اللغوية بالعربية. وتكمن أهمية التعريف بالبرامج الحاسوبية التي يمكن توظيفها في التحليل المعجمي للمتون باللغة العربية في إمداد الباحثين في تخطيط متن اللغة العربية بأحدث البرامج التي يمكن أن تسعف في إجراء حوسبات على معجمها. كما تكمن أهميتها في التحسيس بالثغرة التي تسم تخطيط متن اللغة العربية المتمثلة في عدم وجود قواميس تخص تردد المفردات في العربية، وفي عدم وجود قواميس المتلازمات، مما ينعكس سلبا على تخطيط متن العربية.

Abstract:

This research aims to explore the nature of the vocabulary in the textbook (Mufid) devoted to the 1st level of learning the Arabic language through its inventory, its classification, And to examine the nature of the relationship that links it to the mother tongue, and learning the Arabic language, and to study the possibility of psychological, social and cultural needs of the learner, as well as his ability to express it in his relationship with himself and those around him. We proceed in this research from the following hypothesis: Learning the Arabic language is linked to the ability to meet the needs of the learner, and the ability to respond to his needs. And if the textbook does not pay attention to the need to choose the vocabulary to be promoted for the beginner in learning the Arabic language on the basis of clear criteria, it will negatively affect his learning and help to create a negative impression. The research concluded that the vocabulary promoted in the txtbook is not based on its planning according to exact and precise criteria.

مقدمة

بتطور البحث في هندسة البرامج الحاسوبية قصد تحليل المتون اللغوية، أضحى البحث اللساني يعرف ثورة على مستوى توظيف هذه البرامج من أجل تحليل المتون على جميع المستويات اللسانية المعجمية والصرفية والدلالية والتركيبية والذريعية. وقد باتت استفادة البحث اللساني في اللغة العربية من هذه ملحة؛ قصد مواكبة تطور الأبحاث اللسانية. نهدف في هذا البحث إلى رصد التطورات التي شهدتها هندسة البرامج الحاسوبية التي تعنى بالتحليل المعجمي. ونستقصي الإمكانيات التي تتيحها هذه البرامج للبحث اللساني.

أسئلة الدراسة:

ينطلق البحث من الأسئلة التالية:

- ما أهم البرامج الحاسوبية الموظفة للتحليل المعجمي للمتون اللغوية؟
- ما الإمكانيات التي تتيحها هذه البرامج للبحث اللساني؟
- ما أهم البرامج الحاسوبية المصممة للتحليل المعجمي لمتون اللغة العربية؟ وما إمكاناتها وحدودها؟

البحث إذن منظم كالتالي: نقوم بتعريف البرنامج الحاسوبي (The Stanford Tokenizer) التي يقوم بمجموعة من العمليات الحاسوبية على المتن اللغوي باللغة العربية. ونستعرض الإمكانيات التي يتيحها البرنامج ورد سميث سيفن (7 word smith) المتمثلة في إنشاء لوائح التردد ولوائح التوافقات ولوائح المتلازمات اللفظية. ثم نختم بالبرنامج الحاسوبي (lexical complexity analyzer) الذي يهتم بإنشاء لوائح الثراء المعجمي (Lexical Richness) المتمثلة في الأصالة المعجمية (Lexical Sophistication)، والكثافة المعجمية (Lexical Density)، والمغايرة المعجمية (Lexical Variation).

1. محلل الوريدات

يعتبر النص الرقمي عبارة عن سلسلة من الرموز. ويتعين تقطيع النص الرقمي إلى مجموعة من الوحدات قبل أية معالجة حاسوبية. تمثل هذه العملية التي يتم بموجبها تحويل النص الرقمي إلى وحدات «التوريد» (tokenization). ويسمى البرنامج الحاسوبي الذي يضطلع بهذا الدور ب (toknezir). وتمثل الوحدات الخَرْج وريدات (Tokens)[1].

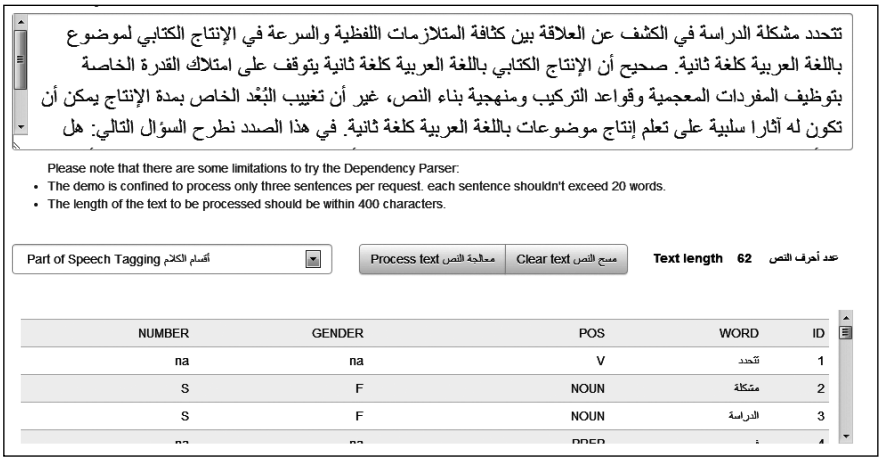
يقصد بمحلل الوريدات ستاندفور (The Stanford Tokenizer) البرنامج الحاسوبي الذي يعمل على عد الوريدات في المتن اللغوي. ويعتمد قياس حجم المفردات في المتن اللغوي على التمييز بين الكلمات النمط /النميطات types[2] والوريدات (tokens). ويقصد بالكلمات النمط عدد الكلمات المختلفة في المتن. فالجملة «يلعب القط الكبير مع القط الصغير» تتضمن خمس كلمات نمط (يلعب، القط، الكبير، مع، الصغير). وتحسب كلمة «القط» مرة واحدة. وتُحسب الوريدات بوصفها جميع الكلمات بصرف النظر عن تكرارها. ففي المثال السابق، يمثل عدد الوريدات ست كلمات.

يقدم البرنامج الحاسوبي ستاندفور توكنايزر مجموعة من العمليات إلى جانب تحليل الوريدات. ويمثل تحديد أقسام الكلام أحد أهم هذه العمليات. وتسمى (part speech tagging). وتوجد برامج حاسوبية تقوم بتحليل أقسام الكلام في لغات مختلفة (الإنجليزية والفرنسية والإسبانية والعربية والصينية). ويتاح تحليل أقسام الكلام على الرابط التالي: http://qatsdemo.cloudapp.net/farasa/demo.html. يتم تحميل النص المراد تحليله في الخانة المعدة لهذا الغرض. يتم الضغط على **معالجة النص**. نقدم فيما يلي الكيفية التي يمكن انتهاجها من أجل معالجة النص على الشبكة العنكبوتية، حيث يتم اتباع الخطوات التالية:

1. فتح الرابط المشار إليه أعلاه
2. إدخال النص المراد معالجته
3. انتقاء العملية

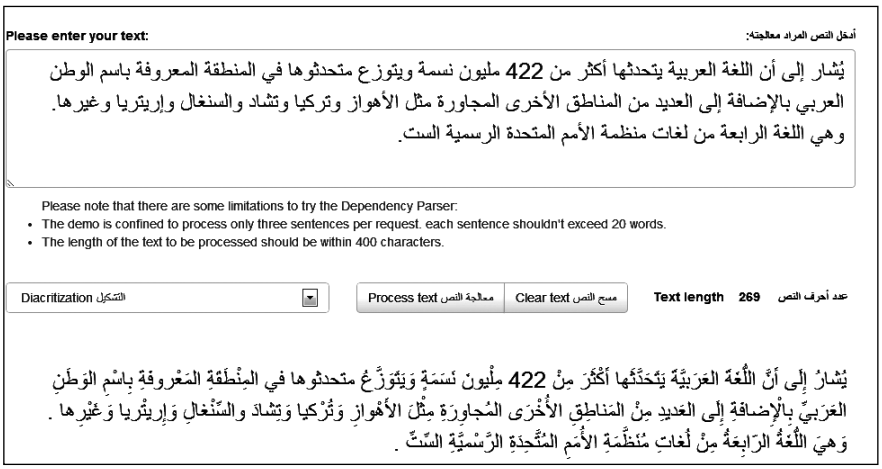
(1) راجع ترجمة الفاسي الفهري (1997)، ص338.
(2) راجع ترجمة الفاسي الفهري (1997) ص343.

4. الضغط على معالجة النص



يتم الحصول على الخرج أعلاه الذي يحدد رتبة الكلمات المتضمنة في النص، والمقولة المعجمية لكل كلمة على حدة، ومقولتين وظيفيتين متمثلتين في العدد والجنس. ويعرف هذا البرنامج الذي يحدد المقولة النحوية لكل كلمة (Part speech tagging). بالإضافة إلى هذا، يقوم البرنامج الحاسوبي بمجموعة من العمليا ت. تتمثل فيما يلي:

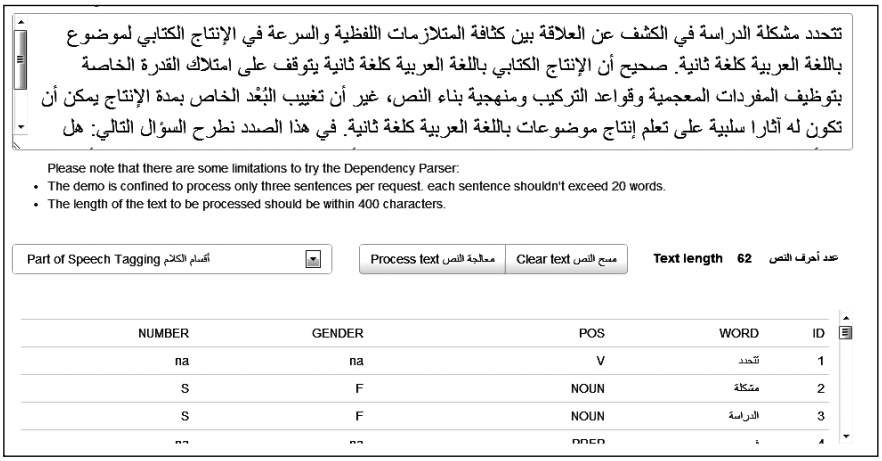
- **الشكلdiacritization -**



- التقطيع الصرفي للكلمات، حيث يقوم البرنامج الحاسوبي بتقطيع الكلمات العربية إلى أجزائها الصرفية كما في المثال التالي: (الكلمات.....ال+كلم+ت).

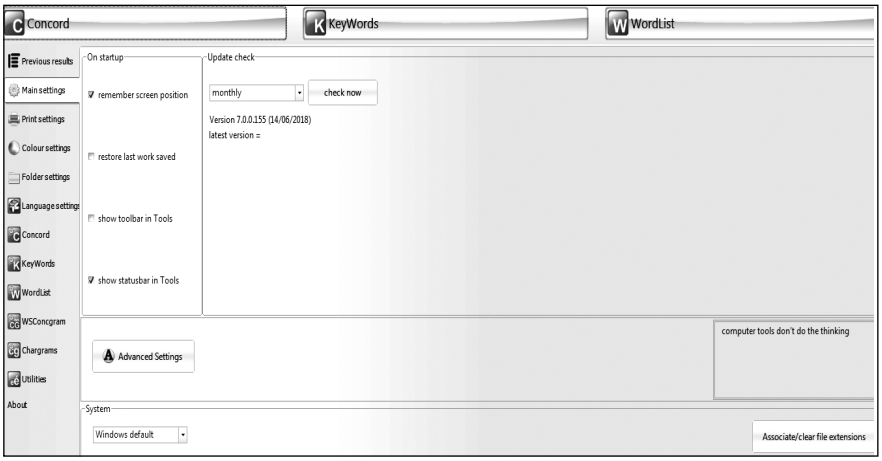
يلاحظ على التقطيع الصرفي للكلمات المتضمنة في المتن مجموعة من المشاكل؛ حيث لا يهتم بتقطيع الصرفيات الخاصة بالجهة أو بالزمن(تشير). كما يمكن أن تقوم بتقطيع شبه-صرفيات مثل (لغ+ة).

لاحظ التقطيع الصرفي للنص التالي:



2.برنامج ورد سميث (صيغة 7)

يمثل برنامج ورد سميث أحد أهم البرامج الحاسوبية التي تعالج متن اللغات. وقد تم تصميمه لإنشاء لوائح التردد ولوائح التوافق الآلي ولوائح المتلازمات كما هو موضح في الصورة التالية:



1.2.لوائح التردد

تعتبر لوائح التردد لوائح يتم تنظيمها ألفبائيا أو ترددیا لجميع الكلمات المتضمنة في المتن اللغوي المراد تحليله. وتتضمن معلومات عن عدد ورود الكلمة في المتن اللغوي بصرف النظر عن مقولتها التركيبية وعن معناها (Leech et al., 2001)). وتعتبر المعلومات الخاصة بتردد الكلمة في المتن مقارِنة؛ أي لها علاقة بتردد الكلمات الواردة في المتن موضوع التحليل. وبترددها في مختلف المتون المراد تحليلها.

وتعد لوائح التردد من مخرجات العديد من البرامج الحاسوبية. ويقصد بها تردد مختلف الكلمات المتضمنة في النص. ويمكن توظيفها لأغراض مختلفة. من أهم هذه الأغراض معرفة تردد الكلمة في المتون اللغوية قصد توظيفها في اختيار الكلمات التي يمكن أن تشكل موضوع التعلم بالنسبة لمتعلم اللغة كلغة ثانية أو أجنبية. كما يمكن تعرف تردد الكلمة (تردد عال أو تردد منخفض)، وتتبعه عبر زمنيا في المتون. في هذا الصدد، نفتقد إلى معجم خاص بتردد الكلمات في العربية في المتون الشفهية والكتابية.

يعترض إنشاء لوائح التردد خلال تحليل المتون جملة من الصعوبات. أهمها أولا، بدون معرفة معلومات عن الليمة، يمكن حساب جميع الصيغ المختلفة الليمة الواحدة كلمات مختلفة. ثانيا، بدون معرفة معلومات عن المقولة من خلال توظيف (part-of-speech)، يمكن حساب كلمات تختلف مقولتها باعتبارها كلمة واحدة.

يمكن استخلاص لائحة التردد بصيغ مختلفة، وذلك بالنظر إلى غايات البحث وأغراضه. نوضح فيما يلي كيفية استخلاص لائحة التردد باستعمال ورد سميث تولز (wordsmith tools) الصيغة السابعة.

- أولا نقوم بتحويل النص إلى صيغة (.txt)
- ثانيا، نحول صيغة النص الخام (texte brute) إلى صيغة أحادي القن (unicode)؛
- ثالثا، نفتح البرنامج الحاسوبي، ثم نقوم بتحميله النص أحادي القن؛
- رابعا، نفتح لائحة التردد، ثم نستخلص لائحة التردد.

تشير المعطيات المتحصل عليها إلى تردد كل كلمة على حدة، والنسبة المئوية لكل منها، وعدد النصوص التي تكررت فيها والانحراف الطرازي (ecart type) لكل كلمة في

علاقتها بالنصوص التي تم تحليلها. ويقصد بالانحراف الطرازي معدل تشتت تردد كلمة من الكلمات بالنسبة إلى معدل تردد الكلمات. ويتيح البرنامج الحاسوبي إمكانية ترتيب الكلمات ترتيبا ألفبائيا. كما يسعف في ضبط المواقع التي ترد فيها كل كلمة على حدة.

Word list (unsaved)

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	%	Dispersion	Lem
1	كاتب	1	0,26	1	100,00		
2	قواعد	1	0,26	1	100,00		
3	قاعدة	1	0,26	1	100,00		
4	كالتالي	1	0,26	1	100,00		
5	ككل	1	0,26	1	100,00		
6	كتب	2	0,52	1	100,00	0,35	
7	كان	1	0,26	1	100,00		
8	فإن	2	0,52	1	100,00	0,35	
9	فأكثر	1	0,26	1	100,00		
10	فاعل	1	0,26	1	100,00		
11	فرضيات	1	0,26	1	100,00		
12	فيها	1	0,26	1	100,00		
13	في	18	4,71	1	100,00	0,72	
14	فعل	1	0,26	1	100,00		
15	كلمة	1	0,26	1	100,00		
16	متغير	1	0,26	1	100,00		
17	ليبر	1	0,26	1	100,00		
18	لها	1	0,26	1	100,00		
19	مثلا	2	0,52	1	100,00	0,35	
20	مدخل	3	0,79	1	100,00	0,51	

ويتيح البرنامج ترتيب الكلمات الواردة في المتن قيد التحليل ترتيبا ألفبائيا كما هو موضح في الصورة الموالية:

Word list (unsaved)

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	%	D...e...s ...
45	الصرفية	1	0,26	1	0...	
46	الصَّرفية	1	0,26	1	0...	
47	الصَّرفية	2	0,52	1	0...	0,35
48	الصيغة	1	0,26	1	0...	
49	الضوء	1	0,26	1	0...	
50	العربية	5	1,31	1	0...	0,71
51	الفاعل	1	0,26	1	0...	
52	الفطري	2	0,52	1	0...	0,35
53	الفقرة	3	0,79	1	0...	0,51
54	القاعدة	2	0,52	1	0...	0,35
55	القائمة	1	0,26	1	0...	
56	القواعد	1	0,26	1	0...	
57	الكلمات	8	2,09	1	0...	0,62

frequency alphabetical statistics filenames notes

34 entries Row 1 0% T S 1

2.2. لوائح التوافق

تسمى هذه اللوائح التي تعين مواقع الكلمات الواردة في المتون بلوائح التوافق (concordance lists). (لاحظ التطبيق الملحق). وترتبط لوائح التوافق بالتحليل المعجمي للمتون اللغوية. وتفيد في تحديد المواقع التي ترد فيها الكلمات الهدف. ويسعف إنشاء لوائح التوافق في تحديد الكلمات التي ترد إلى يمينها وشمالها في المتن موضوع التحليل. وقد تكون هذه الكلمات قليلة، لكنها تسهم في تحديد أنواع التراكيب التي يمكن أن تتضمن الكلمة الهدف. تسهم لوائح التوافق في تحديد السياق الجملي الذي يمكن أن ترد فيه الكلمة.

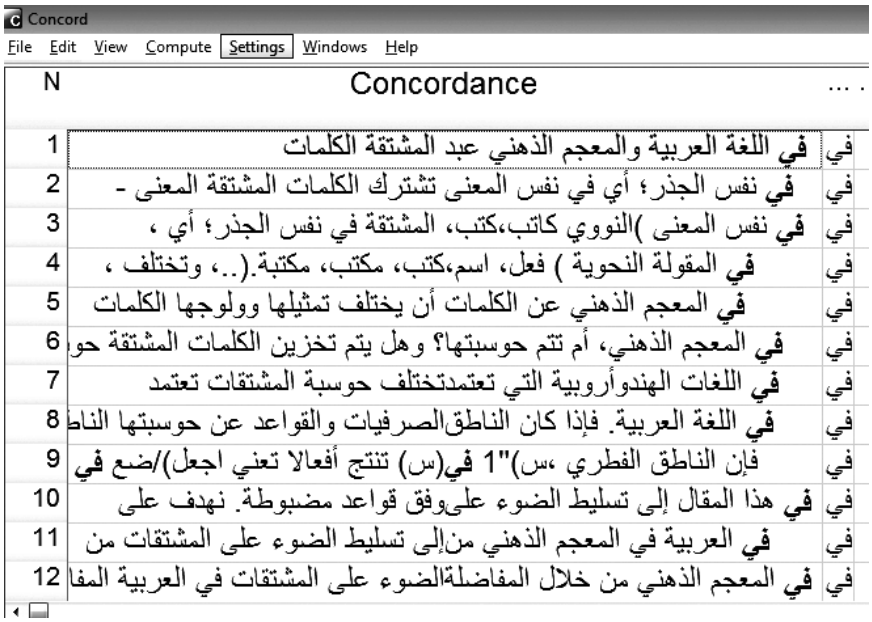
يمكن استخلاص لائحة التوافق بصيغ مختلفة، وذلك بالنظر إلى غايات البحث وأغراضه. نوضح فيما يلي كيفية استخلاصها باستعمال ورد سميث تولز الصيغة السابعة.

- أولا نقوم بتحويل النص إلى صيغة (.txt)

- ثانيا، نحول صيغة النص الخام (texte brute) إلى صيغة أحادي القن (unicode)؛

- ثالثا، نفتح البرنامج الحاسوبي، ثم نقوم بتحميله النص أحادي القن إليه؛

- رابعا، نفتح لائحة التوافق، ثم نستخلص لائحة التوافق بالضغط على (ok).



Concord

File Edit View Compute Settings Windows Help

N	Concordance	...
1	**في** اللغة العربية والمعجم الذهني عبد المشتقة الكلمات	في
2	**في** نفس الجذر؛ أي في نفس المعنى تشترك الكلمات المشتقة المعنى -	في
3	**في** نفس المعنى)النووي كاتب،كتب، المشتقة في نفس الجذر؛ أي ،	في
4	**في** المقولة النحوية) فعل، اسم،كتب، مكتب، مكتبة.(..، وتختلف ،	في
5	**في** المعجم الذهني عن الكلمات أن يختلف تمثيلها وولوجها الكلمات	في
6	**في** المعجم الذهني، أم تتم حوسبتها؟ وهل يتم تخزين الكلمات المشتقة حو	في
7	**في** اللغات الهندوأروبية التي تعتمدتختلف حوسبة المشتقات تعتمد	في
8	**في** اللغة العربية. فإذا كان الناطقالصرفيات والقواعد عن حوسبتها الناط	في
9	فإن الناطق الفطري ،س)"1 **في**(س) تنتج أفعالا تعني اجعل)/ضع **في**	في
10	**في** هذا المقال إلى تسليط الضوء علىوفق قواعد مضبوطة. نهدف على	في
11	**في** العربية في المعجم الذهني منإلى تسليط الضوء على المشتقات من	في
12	**في** المعجم الذهني من خلال المفاضلةالضوء على المشتقات في العربية المفا	في

3.2. لوائح المتلازمات

1.3.2. مفهوم المتلازمة اللفظية

تستعمل المتلازمات بمعان مختلفة في اللسانيات. ويتمثل القاسم المشترك بين هذه التوظيفات المختلفة في كونها تحيل على بعض أنواع العلاقات المركبية (syntagmatic) بين الكلمات. ويعرف سينكلير (sinclair 170: 1991)، المتلازمات بوصفها « توارد كلمتين أو أكثر في حيز قريب من بعضهما البعض في النص». وقد تم تحديد الحيز في أربع كلمات على يمين الكلمة المعنية أو على يسارها على أبعد تقدير. وتسمى هذه الكلمة نواة المتلازمة. ويربط سينكلير (173: 1999) المتلازمة بمبدأ المسكوكة (idiom-principle) عندما يؤكد أن «المتلازمات توضح مبدأ المسكوكة طالما أن اختيار كلمة يؤثر على اختيار الكلمات المجاورة لها»

وحيث إن لسانيات المتن (corpus linguistics) تقوم على الفكرة التي مؤداه أن كل ما هو متردد في اللغة له دلالة، يعرف ستوبس (Stubbs, 29 :2002) التلازم بوصفه

« تردد دال». لذا، يعد التلازم في نظره « علاقة بين الكلمات في سلسلة خطية؛ حيث تسمح النواة بتوقع الكلمة التي ترد بعدها». الشيء الذي حذا بسمادجا (Smadja, 143: 1993) إلى وصفها بكونها «توليفات متواترة لكلمات تتوارد أكثر مما يتوقع من خلال الحظ».

ويرى كروز (Cruse, 40 :1986) أنه «يمكن أن يستعمل لفظ التلازم للدلالة على مقاطع من عناصر معجمية تتوارد مع بعضها البعض، شريطة أن تكون شفافة؛ بمعنى أن يكون كل مكون معجمي منها مكون دلالي» . ويعتبرها مانين وسكوتز (Manning& Schütze, 59 :1999) «تعبير يتكون من كلمتين أو أكثر يناسب بعض الطرق الاصطلاحية لقول الأشياء».

ومن جهته، يستند مالوك (Mel'ˇcuk, 53: 2002) إلى معيار الاصطلاحية لتعريف المتلازمة، حيث يقول:» لا يمكن تكوين المتلازمة من الكلمات أو المركبات البسيطة انطلاقا من قواعد عامة للغة، بل ينبغي أن تخزَّن وتستعمل بوصفها كُلاًّ».

في هذا الإطار يوضح مالوك التلازم بالطريقة التالية: لنعتبر أن الزوج (أ،ب) تعبير في اللغة (ل)، حيث (أ) و(ب) وحدتان معجميتان في (ل). ولنعتبر أن (م) هو معنى التعبير (أ،ب) وأن (أ') هو معنى(أ) و(ب') هو معنى (ب). نقول إن (أ،ب) متلازمة إذا وفقط إذا تحققت الشروط التالية:

(1) يتضمن المعنى (م) المعنى (أ')؛

(2) يتم انتقاء(أ) بطريقة منتظمة وغير مقيدة،

(3) يتم انتقاء (ب) بطريقة منتظمة وغير مقيدة لكن متعلقة باختيار (أ) وبالمعنى (م) المراد التعبير عنه.

ويعتبر فيرث (Firth, 181 :1957) أن «متلازمات كلمة ما هو التصريح بالمواقع المعتادة والمألوفة لهذه الكلمة» كما يلفت فيرث الانتباه إلى أنه «لا يمكن اعتبار متلازمات كلمة مجرد تجاور بسيط فقط، بل توقع متبادل بينها».

وتعتبر المتلازمة اللفظية ظاهرة معجمية تفيد ميل المفردات المعجمية إلى التوارد مع بعضها البعض. وقد تم تغييب المتلازمات اللفظية باعتبارها بعدا من الأبعاد

المعجمية في الدراسات المعجمية العربية. ويرجع السبب في نظرنا إلى الحاجة إلى آليات متطورة ترصد تلازم الكلمات في المتون اللغوية. يقصد بلوائح المتلازمات اللفظية اللوائح التي تحصر الكلمات التي ترد مع بعضها البعض. تسمح هذه التقنية في استخلاص لوائح المتلازمات اللفظية (collocations lists) في المتون المعجمية، خصوصا إذا تم تبني سينكلير لمفهوم المتلازمة الذي يعتمد على المسافة من الكلمة النواة (سنكلير،1991). في هذا الصدد، تجدر الإشارة إلى افتقار الخزانة العربية إلى معاجم المتلازمات اللفظية التي يتم استخلاصها من المتون اللغوية العربية اعتمادا على برامج حاسوبية. (لاحظ التطبيق في الملحق).

وتتميز المتلازمات عن لائحة التوافق بكون البرنامج الحاسوبي يميز بين ورود الكلمة الهدف بين كلمات (سياق جملي) بالنسبة لإنشاء لائحة التوافق، وبين تردد هذا الورود بالنسبة لإنشاء لائحة المتلازمات اللفظية.

ويفيد تحديد المتلازمات اللفظية في إنشاء قواميس المتلازمات في اللغة العربية على شاكلة القواميس في اللغة الإنجليزية مثل (collocation oxford dictionary)، وقاموس المتلازمات الإلكتروني في اللغة الفرنسية[3] . وتعتبر هذه القواميس أدوات مهمة تسعف في التخطيط اللغوي. كما تسهم في تعليم اللغات وتعلمها.

Collocate List (unsaved)

ile Edit View Compute Settings Windows Help

N	Word	...Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	...e...e	R1	R2	R3	R4	R5
1	الكلمات	1	7	0	0						7					
2	المشتقة	1	7	0	7							7				
3	في	1	7	1	6		1						6			
4	الذهني	1	6	1	5				1					1	4	
5	المعجم	1	5	1	4			1						4		
6	يتم	1	3	3	0				3							
7	تخزين	1	3	3	0					3						
8	والمعجم	1	2	0	2								1			1
9	أم	1	2	0	2											2
10	هل	1	2	2	0			2								

(3) . انظر الموقع الإلكتروني التالية: http://www.tonitraduction.net/

3. محلل التعقيد المعجمي

يضطلع محلل التعقيد المعجمي (**Lexical Complexity Analyzer**) برصد مختلف جميع أبعاد الغنى المعجمي (Lexical Richness) المتمثلة في المغايرة والكثافة والأناقة المعجمية. ويرتبط التعقيد المعجمي بالإنتاج الكتابي أو الشفهي. ويراد بالكثافة المعجمية (lexical density) نسبة الكلمات ذات المحتوى (content word) في النص. ويقصد بالكلمات ذات المحتوى الكلمات المملوءة معجميا؛ أي الاسم والفعل والصفة والظرف (ستوبس ((Stubbs, (2004: 122)) وجوهانسن (Johansson, (2009 : 146)). وتأتي الكلمات ذات المحتوى في مقابل الكلمات الوظيفة (function word) التي تتضمن الحروف والروابط والمصدريات وغيرها. ويندرج مفهوم الكثافة المعجمية ضمن ما يسمى بالثراء المعجمي (lexical richness) الذي يعتبر مؤشرا على غنى النص المكتوب بالمعاني؛ حيث يُفترض أنه كلما ارتفعت نسبة الكثافة المعجمية كلما كان النص غنيا بالمعاني والمعلومات والأفكار (الوحي (AL-wahy, 5 :2016)). لذلك، أُعتبرت الكثافة المعجمية دالة على كثافة معاني النص، حيث وُجد أن نسبة الكثافة المعجمية في المكتوب تكون مرتفعة مقارنة بالشفهي. وقد عُزي الأمر إلى كون اللغة المكتوبة تنزع إلى تكثيف المعاني مقارنة باللغة الشفهية.

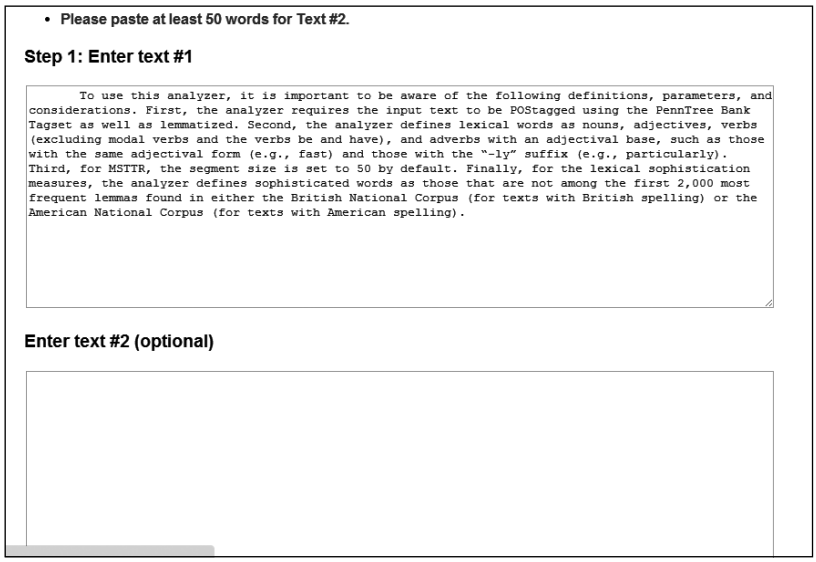
وقد وُظّف مؤشر الكثافة المعجمية في دراسات تطبيقية متعددة، حيث أُستند إليها للتمييز بين إنتاج الناطق الفطري وإنتاج الناطق غير الفطري (لينارید ((Linnarud, (1976)). كما وُظف للتمييز بين النصوص المترجمة والنصوص غير المترجمة. وقد تم اعتمادها أيضا لمقارنة السجل اللغوي عبر-لغويا (نومان ((Neumann ,(2014)).

ويقصد بالكثافة المعجمية (lexical variation) نسبة الكلمات المختلفة الموجودة في النص، وتختلف حسب طول النصوص. لذا، ينبغي استعمال عينة من النصوص من نفس الطول. ومع ذلك، قد تبقى مجموعة من المشاكل عالقة: أولا تتعلق المغايرة المعجمية بتعريف الكلمة. فإذا تم اعتبار المشتقات كلمات مختلفة، فإن المغايرة تكون أكبر من اعتبار الكلمات التي تنتمي إلى نفس الأسرة كلمة واحدة. ويتم حسابها كالآتي:

$$1.\ \text{المغايرة المعجمية} = \frac{\text{عدد الكلمات المختلفة} \times 100}{\text{إجمالي الكلمات الموظفة}} \times 100$$

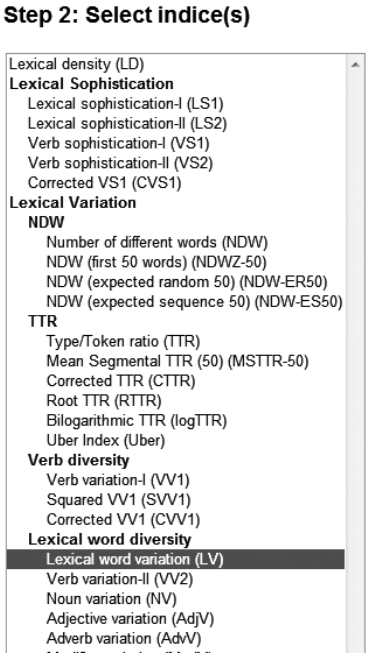
يتيح البرنامج الخاص بالتحليل المعجمي للغة الإنجليزية رصد مختلف المؤشرات الخاصة بالثراء المعجمي. ويمكن مقارنتها بمؤشرات نص ثاني على الشبكة العنكبوتية. ويتطلب البرنامج الحاسوبي ثلاث مراحل:[4]

2. إدخال النص



(4) . انظر البرنامج على الرابط: https://aihaiyang.com/software/lca/single/

3. اختيار مؤشر الثراء المعجمي



4. المعالجة والحصول على النتائج

من الملاحظات التي يمكن تسجيلها في هذا الصدد: أولا، يعمل البرنامج بسعة محدودة. ثانيا، يشتغل البرنامج على تحليل ومعالجة اللغة الإنجليزية. ولا يتعداها إلى اللغة العربية.

خاتمة

إن تصميم برامج حاسوبية تعنى بالتحليل المعجمي للمتون في اللغة العربية. تقتضي تضافر جهود الباحثين في اللسانيات والمعلوميات والرياضيات. في هذا الصدد، تم الاشتغال على مجموعة من البرامج الحاسوبية. تعمل على تحليل متن اللغات. ويمكن أن يستفاد منها لتحليل متن اللغة العربية بإجراء مجموعة من التعديلات. وقد تطرقنا إلى محلل الوريدات الذي طوره معهد قطر لبحوث الحوسبة. ورصدنا الإمكانات التي يتيحها والقضايا والإشكالات العالقة. ثم رصدنا البرنامج الحاسوبي ورد سميث (صيغة7). واستقصينا الإمكانات التي يتيحها في التحليل المعجمي لمتن اللغة العربية والمتمثلة في إنشاء لوائح التردد والتوافق والمتلازمات. ثم ختمنا برصد البرنامج الحاسوبي الذي طوره لو (Lu, 2010; 2012) والذي يعنى بتحليل مؤشرات الثراء المعجمي المتمثلة في الكثافة المعجمية والمغايرة المعجمية والأناقة المعجمية. واستنادا إلى خلاصات البحث نوصي بما يلي:

- إنشاء فريق متعدد التخصصات يعمل على تطوير برنامج حاسوبي يقوم بتحليل التعقيد المعجمي يمكن الاستفادة منه في تقويم الإنتاجات الكتابية المختلفة؛
- ضرورة إنشاء قواميس تردد الكلمات في العربية بالاستفادة من برامج حاسوبية متطورة؛
- إنشاء قواميس المتلازمات اللفظية من خلال تحليل متن العربية باستعمال برامج حاسوبية قادرة على تحليل متون كبيرة.

المراجع

- الفاسي الفهري، عبد القادر. (1997). معجم المصطلحات اللسانية، ط 1، دار الكتاب الجديد المتحدة، بيروت-لبنان.
- Al-Wahy, A.S. (2015). Towards a Methodology for Measuring Lexical Density in Arabic, *The International Journal of Arabic Linguistics*, Vol. 3 Issue 1, pp.1-33.
- **C**ruse DA. (1986). *Lexical Sémantiques*. Cambridge University Press, Cambridge.
- **F**irth, J. R. (1957). *Papers in linguistics 1934-1951*. London : Oxford University Press.
- Johansson, V. (2009). Developmental aspects of text production in writing and speech. *Travaux de l'Institut de Linguistique de Lund*, 48. Lund: Lund University.
- Leech G, Rayson P&Wilson A (2001).Word frequencies in written and spoken English. London: Longman.
- Linnarud, M. (1976). Lexical density and lexical variation: An analysis of the lexical texture of Swedish students' written work. *Studia Anglica Posnaniensia: An International Review of English Studies*, 7, pp.45 –52.
- Lu, Xiaofei(2014).*Computational Methods for Corpus Annotation and Analysis*, Springer New York Heidelberg Dordrecht London
- **M**anning CD, Schütze H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.
- **M**el'ˇcuk I. (1998). Collocations and lexical functions. In: Cowie AP (ed) *Phraseology. Theory, Analysis, and Applications*, Claredon Press, Oxford, 23–53.
- Neumann, S. (2014). *Contrastive register variation: A quantitative approach to the comparison of English and German*. Berlin: Mouton.
- **S**inclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- **S**madja F (1993). Retrieving collocations from text: Xtract. *Computational Linguistics* , 19(1): 143–177.
- **S**tubbs, michael (2002). *word and phrases corpus studies of lexical semantics*, Blackwell publishing.