تقنيات حاسوبية لبناء المعاجم اللغوية
لمتعلمي اللغة من غير الناطقين بها

د. عبدالله بن يحيى الفيفي
قسم علم اللغة التطبيقي - معهد تعليم اللغة العربية
جامعة الإمام محمد بن سعود الإسلامية

تقنيات حاسوبية لبناء المعاجم اللغوية لمتعلمي اللغة من غير الناطقين بها

د. عبدالله بن يحيى الفيفي

قسم علم اللغة التطبيقي - معهد تعليم اللغة العربية - جامعة الإمام محمد بن سعود الإسلامية

تاريخ تقديم البحث: ٢٢/ ٩/ ١٤٤٠هـ تاريخ قبول البحث: ٤/ ١/ ١٤٤١هـ

ملخص الدراسة:

تستعرض هذه الورقة مجموعة من التقنيات الحاسوبية التي يستخدمها المتخصصون في بناء المعاجم الحديثة، وبالتحديد تلك الموجهة إلى متعلمي اللغة من الناطقين بغيرها، ومن هذه التقنيات تحليل سلوك المفردة في سياقاتها اللغوية الذي يساعد على اكتشاف المتلازمات اللغوية التي يفترض إدراجها في المداخل المعجمية، إضافة إلى قياس مدى قوة التلازم الذي يساعد على ترتيب المتلازمات وفق أهمية استعمالها، ومن التقنيات كذلك قوائم شيوع المفردات وأدوات قياس شيوعها التي يستفاد منها في كتابة شرح المداخل المعجمية، ومنها أيضاً مستخرج الأصول المعجمية الذي يعتبر أساساً في المعاجم الحاسوبية وخصوصاً معاجم الطلاب، وكذلك كشاف السياقات الذي أضحى إحدى ركائز الباحثين في المدونات اللغوية، سواء أكانت مدونات عامة أم مدونات لمتعلمي اللغة. وقد سعى الباحث في هذه الدراسة إلى ربط تلك التقنيات الحاسوبية بصناعة المعاجم الحديثة خصوصاً الموجهة لمتعلمي اللغة؛ وذلك لإبراز دور التطورات التقنية الحديثة، وأهمية الاستفادة منها في صناعة مهمة كصناعة المعاجم اللغوية.

الكلمات المفتاحية: معجم، تقنيات حاسوبية، كشاف سياقات، مداخل معجمية، مدونات لغوية، شيوع المفردات.

المقدمة:

استفاد الباحثون من التطور الكبير في الوسائل والأدوات اللغوية والحاسوبية في إثراء محتويات المعاجم الحديثة، مع ما صاحب ذلك من تسهيل لعملية تأليف المعجم، وتقريب المادة المعجمية من المستخدم المستهدف بها، ولا زال البحث العلمي يظهر لنا بين الفينة والأخرى وسائل وأدوات تخدم هذا المجال، بعد أن كانت صناعة المعاجم من أشقّ المهامّ على الخبراء والمختصين لما تستغرقه من وقت وجهد كبيرين لإنجاز أي عمل معجمي، وهذا لا يعني أن صناعة المعاجم أصبحت تُنجَز آلياً بالكامل، لكن كثيراً من مراحلها تقلص إلى خطوات يسيرة يقوم بها المعجمي في ساعات أو دقائق معدودة، بعد أن كانت تستغرق منه شهوراً وربما سنوات، وتحتاج إلى أعداد كبيرة من المتخصصين، ومن ذلك على سبيل المثال استخراج قائمة بالكلمات الشائعة في مجموعة كبيرة من النصوص، مع حساب عدد التكرار ونسب الشيوع بدقة، وهذا بلا شك قد شجع البحث في مجال المعاجم، بل وأظهر مجموعة كبيرة من المعاجم المتخصصة أو الموجهة لفئات معينة مثل متعلمي اللغة الناطقين بغيرها، وهذا النوع من المعاجم بحاجة إلى الاستفادة من التقنيات الحاسوبية الحديثة التي تشرحها هذه الدراسة.

١ سؤال البحث:

ما هي الأدوات الحاسوبية المستعملة في المعالجة الآلية للغة العربية والتي يمكن توظيفها في بناء المعاجم اللغوية الحديثة خصوصاً تلك الموجهة للناطقين بغير العربية؟

٢ مشكلة البحث:

لو استعرضنا المعاجم العربية المعاصرة فسنجد أن أغلبها وإن أخذ بطرف من التقنية الحديثة إلا أنها لم تستفد بشكل كامل من الأدوات الحاسوبية التي

حملها التطور التقني في مجال الصناعة الحديثة للمعاجم، وخصوصاً أدوات المعالجة الآلية للغة العربية، وهو ما يبرز الحاجة إلى التعريف بهذه الأدوات مع ربطها بآلية استعمالها في صناعة المعاجم اللغوية الحديثة الموجهة للناطقين بغير العربية.

٣ الإطار النظري

من المعلوم أن المكتبة العربية تزخر بالكثير من المعاجم التراثية التي خدمت الباحثين على مرّ العصور – ولا زالت – بما تحويه من معلومات ثرية عن معاني المفردات العربية، لكن القليل منها فقط موجه لمتعلمي اللغة العربية من الناطقين بغيرها، ولعل "المعجم العربي الأساسي" (المنظمة العربية للتربية والثقافة والعلوم، ١٩٨٩م) يُعدّ الأبرز في هذا المجال رغم قدمه وعدم اعتماده على التقنيات الحاسوبية الحديثة في بناء المعاجم، والتي باتت ركيزة أساسية في نظرائه من المعاجم الأجنبية الموجهة للمتعلمين.

ولو نظرنا إلى حاجات مستعمل المعجم، سنجد أن الناطق باللغة لا يبحث في الغالب عن الجوانب التي يتقنها بالسليقة في لغته كالنطق والرسم والتراكيب مثلاً، وإنما يركز في بحثه على دلالات المفردات، لمعرفة حدود استعمالها أو اكتشاف الدلالات التي قد تخفى عليه، وقد يبحث في بعض المشتقات الصرفية للمفردة التي يقل أو يندر استعمالها، أما من يتعلم لغة جديدة – وخصوصاً تلك التي تختلف كثيراً عن لغته الأم – فهو بحاجة إلى معجم يوفر له معلومات أكثر سهولة عن المفردات؛ ليسهل عليه تعلمها واستعمالها بطريقة صحيحة، إذ إنه قد يجهل كثيراً من الأساسات التي يعرفها الناطق باللغة، ومن ذلك على سبيل المثال المشتقات الرئيسة للمفردة مثل: المفرد والمثنى والجمع في الأسماء، والماضي والمضارع والأمر في الأفعال، كما

أنه من المفيد في معاجم متعلمي اللغة العناية بالشيوع سواء أكان ذلك في الألفاظ أم في المعاني، فإن المتعلم يستفيد من الشائع أكثر من غير الشائع وخصوصاً في المستويات الأولى من تعلمه للغة، ويقاس على ذلك كتابة شرح المداخل المعجمية بألفاظ شائعة تُسهِّل على المتعلم فهم الشرح، كما أن متعلم اللغة يستفيد كثيراً من وجود معلومات حول أبرز الأخطاء اللغوية في المفردات التي يكثر الخطأ فيها، سواءً أكان ذلك في رسمها أو تصريفها أو تركيبها أو دلالتها. ومن هنا فإن أهم ما يميز المعاجم اللغوية الموجهة لمتعلمي اللغة من الناطقين بغيرها احتواؤها على مثل هذه المعلومات اللغوية الإضافية التي وإن كانت بدهية بالنسبة للناطق باللغة إلا أن لها فائدة كبيرة للناطق بغيرها. ولا شك أن إثراء المعجم بمثل هذه المعلومات يضاعف الجهد والوقت اللازمين لبناء المعجم مما يؤكد على أهمية الاستفادة من التقنيات الحاسوبية الحديثة في مجال بناء المعاجم.

يقول عمر (١٤١٨هـ، ص١٧٩): "كتب أحد المعجميين المشهورين عام ١٩٧٠ يقول: إننا مقدمون على عصر حينما يكون المعجم الذي لا يتم التعامل معه آلياً معجماً ناقصاً. وتنبأ Zgusta عام ١٩٧١ بأن المعاجم الأكاديمية الضخمة لن تنشر ورقياً بعد ذلك، لأن المعجم الورقي عاجز عن استيعاب ما هو مخزن، واختصار المادة المخزنة قد يخلّ بها. ويقول Landau: من غير المتصور الآن أن معجماً كبيراً يمكن أن يصنع اليوم دون تخزين المادة في الحاسوب". ولهذا بدأت صناعة المعاجم تتحول من المعاجم اليدوية إلى المعاجم الحاسوبية منذ بداية ظهور الحاسب الآلي، بل إن مجالاً علمياً جديداً بدأ في الظهور تحت اسم علم المعجم الحاسوبي Computational Lexicography الذي يعرفه هنكس (Hanks, 2003) بأنه ذلك العلم الذي

يعيد هيكلة القواميس البشرية للاستفادة منها في أغراض حاسوبية، كما يعرفه أيضاً بأنه العلم الذي يهتم باستعمال التقنيات الحاسوبية لبناء المعاجم الجديدة. ولذا ظهرت معه المعاجم المحوسبة أو المعاجم المقروءة آلياً Machine-readable dictionary، كما ظهرت معه مجموعة من الأدوات والتطبيقات الحاسوبية التي سهلت بناء هذه المعاجم.

ونظراً لحجم العمل الكبير الذي يصاحب بناء المعاجم فإن هذه الأدوات والتطبيقات لا زالت تتطور في محاولة للقيام بأكبر دور ممكن في سبيل اختصار الجهد والوقت من ناحية، وتوفير أعلى درجات الدقّة من ناحية أخرى، كما أنها تسهم في توفير مادة معجمية أكثر عملية وواقعية خصوصاً عندما يتعلق الأمر باستعمال المعجم من أجل تعلم اللغة، ويؤكد هذا عمر (١٤١٨هـ) إذ يرى أن عذر المعجميين العرب المعاصرين في عدم اللجوء إلى الجمع الميداني لمادة معاجمهم صعوبة العمل من ناحية، وضخامة حجم المادة من ناحية أخرى، مما يجعل التعامل مع ملايين الكلمات أمراً مستحيلاً، وأن هذا العذر لم يعد مقبولاً الآن بعد استخدام الحواسيب وإمكانية التعامل مع ملايين الكلمات والاقتباسات، كما يشير إلى أن أهم ما يميز المعاجم الحديثة احتواؤها على كثير من الاستعمالات التي تتردد في النصوص الحية، وهي تلك التي تُجمع من المادة الحية المأخوذة من نصوص واقعية، كما أن من أهمّ مظاهر التحول في صناعة المعجم العربي التوسع في القيام بالإحصاءات الحاسوبية والمعالجة الآلية للغة، كل هذا اختصر عشرات السنين من الوقت اللازم لبناء المعاجم، ويمثل على ذلك ببناء المعجم الكبير الذي بدأ مجمع اللغة العربية في القاهرة العمل فيه عام ١٩٤٦م، إذ وجد أنه من خلال حساب معدل الإنجاز خلال ٤٠ سنة والذي لم يتجاوز ١٠٪ من حجم المعجم، فإنه يتوقع أن

ينتهي المجمع من هذا المعجم بعد ٤٠٠ سنة، بينما انتهت معاجم تاريخية للغات شرقية وغربية استخدمت التقنيات الحديثة في جمع المادة وتحليلها ومنها: معجم أستراليا الوطني التاريخي، والمعجم التاريخي للغة العبرية، والمعجم التاريخي للغة الهولندية، والمعجم التاريخي لجامعة شيكاغو، والمعجم التاريخي لجامعة جلاسكو، ومعجم المدونة اللغوية الفرنسية.

وحيث إن استعمال الأدوات الحاسوبية الحديثة في بناء المعاجم اللغوية لا يزال حديثاً بالنسبة للمعاجم العربية فإنه لا تزال هناك بعض السلبيات التي قد تحتاج إلى النظر فيها ومعالجتها سلفاً، ومن ذلك ما يقتضيه استعمال هذه الأدوات من مهارات خاصة قد لا تتوفر لكثير من مستخدميه (العارف، ١٤٢٨هـ)، ومنها أيضاً عدم ملاءمة بعض هذه الأدوات للنص العربي مما يوجب على المشتغلين بمجال اللغويات الحاسوبية تطويعها أو بناء أدوات جديدة يمكنها معالجة النص العربي؛ ليسهل على صناع المعاجم بعد ذلك الاستفادة منها وتوظيفها في بناء معاجم حاسوبية عربية معاصرة تواكب نظيرتها الأجنبية التي سبقتها بمراحل.

٤ الدراسات السابقة

تناول المتخصصون في صناعة المعاجم اللغوية في مؤلفاتهم وأبحاثهم أهمية بناء المعاجم الحاسوبية التي تسهم في تحليل اللغة آلياً وسبر أغوارها والنهضة بالعلوم القائمة عليها وغير ذلك من الفوائد التي لا تخفى، كما أن منهم من تناول المناهج والأدوات التي تسهم في بناء هذا النوع من المعاجم الحاسوبية، ومنهم – وخصوصاً المشتغلين باللغويات الحاسوبية – من أشار إلى بعض التقنيات الحاسوبية التي يمكن الاستفادة منها وتوظيفها في بناء المعاجم العربية، وسنستعرض هنا بعضاً من هذه الدراسات لعلها تكون لنا صورة أولية عن هذا الميدان.

فمن أوائل من تحدث عن المعاجم الحاسوبية نبيل علي (١٩٨٨م) في كتابه "اللغة العربية والحاسوب: دراسة بحثية"، إذ ناقش المستوى الدلالي للغة تحت فصل بعنوان "ميكنة المعجم العربي"، وأوضح فيه الكثير من الأساسيات اللغوية اللازمة لتهيئة الطريق إلى منظومة متكاملة للمعجم العربي، مع توصيف لمكونات هذه المنظومة والعلاقات البينية التي تربط بين هذه المكونات، وذلك سعياً إلى إبراز حتمية ميكنة المعجم العربي. وهو في بحثه لم يتطرق لشرح التقنيات الحاسوبية اللازمة لمعالجة المعجم أو المستوى الدلالي، معللاً ذلك بالحاجة إلى كثير من البحوث النظرية في الذكاء الاصطناعي والمنطق والرياضيات للوصول إلى المستوى المطلوب لبناء المعاجم الآلية، والتي يمكن للحاسب معها معالجة دلالات اللغة فهماً وتوليداً.

ومثله نهاد الموسى (٢٠٠٠م) في كتابه "العربية: نحو توصيف جديد في ضوء اللسانيات الحاسوبية"، الذي ركز عند حديثه عن تمثيل المعجم على شرح التوصيف الدلالي اللازم لتمثيل مفردات المعجم وتسهيل معالجتها آلياً، وقد اعتمد في ذلك على قالب Template يمكن اعتباره معياراً لتوصيف كل مفردة شاملاً في ذلك الأسماء والأفعال، وكذلك الحروف والأدوات التي يرى ضرورة استيفاء دلالاتها ووجوه استعمالها وفق الأدلة العلائقية التي تؤدي إلى هذه المعاني والاستعمالات. ويشتمل هذا القالب على معلومات عن المفردة، وعن التغير الصرفي، والبنية، والعدد، والجنس، والاقتضاء، والزمان، والجهة، والمعنى، والأصل، والصيرورة، والوظيفة، والرتبة، والموقع، وحالة الإعراب، والتعريف والتنكير. ويرى الموسى أن المعجم إذا أصبح بهذا التوصيف فإنه يشرع لمستعملي اللغة أبواباً في ابتكار الألفاظ الجديدة، وأنه يمكننا من تلقي ألفاظ لم تعرض لنا قبلاً، إذ إنه يحاكي بذلك

آلية تشكل المعجم لدى الناس عند اكتسابهم لغاتهم، ذلك أنه يمثل نظاماً دلالياً لا مجرد قائمة من المفردات.

أما سلوى حمادة (٢٠٠٩م) فتشرح في كتابها "المعالجة الآلية للغة العربية: المشاكل والحلول" منهجية لبناء قاعدة بيانات معجمية آلية في شكل معجم كوني ذي تسلسل هرمي، تبوب فيه المعاني تبويباً هرمياً فتُذكر متدرجة من الأصل إلى الفرع ومن الحسي إلى المعنوي ومن المألوف إلى الغريب، مع أهمية العلاقات التي تربط هذه المعاني، وتسهم في حل مشاكل التحليل اللغوي. وتُجمع مادة المعجم ابتداءً عن طريق تحليل نصوص لغوية شاملة لأغلب الموضوعات والأساليب والمصادر تحليلاً إحصائياً بغرض إحصاء الكلمات والأساليب الواردة فيها وحصر معانيها المختلفة وفق نسبة ورودها، ومن ثم وضعها في المعجم الذي يجب أن يقسم إلى معاجم صغيرة ترتبط آلياً ببعضها. ورغم الميزات المشروحة في هذا المعجم إلا أن المؤلفة لم تشرح الأدوات الحاسوبية التي يمكن استخدامها في بنائه تاركة ذلك للمتخصصين في المجال الحاسوبي، إذ ترى أن مثل هذا المعجم لا يصنعه اللغويون فحسب وإنما يجب تكوين فرق عمل من تخصصات مختلفة.

وفي فصل بعنوان "مدونات المتعلمين" لعبدالله لفيفي (١٤٣٦هـ) في كتاب "المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها"، شرح المؤلف أهمية الاستفادة من مدونات المتعلمين Learner Corpus في بناء المعاجم اللغوية، إذ بيَّن أن صناعة المعاجم الموجهة لمتعلمي اللغة - الإنجليزية على وجه التحديد - من أوائل المواد التعليمية التي استفادت من الدراسات المعنية بتحليل مدونات المتعلمين، أما في الوقت الحالي فأغلب معاجم الطلاب تعتمد على هذا النوع من المدونات، خصوصاً المعاجم أحادية اللغة (Monolingual

learners' dictionaries). كما ذكر منهجية الاستفادة من نتائج تحليل مدونات المتعلمين في إثراء معاجم الطلاب، وأن من الصور الشائعة لذلك الاستفادة من مخرجات تحليل الأخطاء اللغوية على مستوى المفردات أو التراكيب، وتزويد قراء المعجم بتنبيهات حول هذه الأخطاء في المداخل المعجمية المرتبطة بها، مع بيان صور استخدامها الصحيح.

أما المعتز بالله السعيد (٢٠١٦م) في بحثه المعنون له بـــ "تطبيقات متقدمة لصناعة معجم اللغة العربية للناطقين بغيرها للمستويين التعليميين B1 وB2" فقد كان أكثر تركيزاً ممن سبقه إذ أتى على مجموعة من أساليب صناعة المعاجم التعليمية العربية، وأورد مجموعة من التطبيقات تحت عنوان فرعي حمل اسم "تطبيقات متقدمة لصناعة المعجم التعليمي المنشود"، وفيه قسم هذه التطبيقات إلى ثلاثة أقسام هي: (١) الموارد: وهي التطبيقات اللغوية التي تُستمد منها مادة المعجم، وتشمل: المدونة اللغوية والمكنز اللغوي. (٢) أساليب التحليل: وتمثلها التطبيقات الإحصائية التي تضبط مخرجات المعجم، وتوفر المعلومات الإحصائية بشأن استعمال وشيوع المفردات، وتشمل: القياسات الببليومترية، والنحو العددي. (٣) أدوات المعالجة: وتمثلها التطبيقات الحاسوبية التي يمكن توظيفها في استخلاص مادة المعجم من الموارد، واستنباط المعلومات الإحصائية عبر أساليب التحليل، وتشمل: تطبيق الفهرسة الآلية، وتطبيق الكشاف السياقي.

وفي بحث آخر للفيفي وآخرين (١٤٣٧هـ) بعنوان "معجم المفردات الشائعة لمتعلمي العربية (شائع)" أورد المؤلفون نبذة عن أهمية بناء معجم عربي حديث لمتعلمي اللغة العربية الناطقين بغيرها يستفيد من التقدم الكبير في مجال اللغويات الحاسوبية والمدونات اللغوية، وذلك لتقديم مادة معجمية تختصر

على متعلم اللغة العربية من الناطقين بغيرها الوقت والجهد اللازمين لتعلم مفردات اللغة ومعانيها ومشتقاتها، كما تحدث الباحثون عن مشروعهم لمعجم المفردات الشائعة لمتعلمي العربية "شائع"، وهو معجم بُني على نتائج التحليل الإحصائي لشيوع الكلمات والمعاني والاشتقاقات في "المدونة العربية" (Althubaity, 2014)، وتمتد فكرة هذا الشيوع لجميع جوانب المعجم تقريباً، حيث يُبرز المعجمُ المفردات الأكثر شيوعاً في اللغة العربية، مع تصنيفها حسب درجة شيوعها، ويقدم كذلك المعاني الأكثر شيوعاً لكل مفردة، مرتبةً حسب شيوع استخدامها، وكذلك الحال في الاشتقاقات التي يُبرز المعجم الشائع منها لكل مدخل معجمي، كما أن صياغة التعريفات تحت كل مدخل معجمي تعتمد على قائمة الكلمات الشائعة المستمدة من المدونة اللغوية. وقد طور الفريق في سبيل بناء هذا المعجم عدة أدوات حاسوبية، منها أداة للمساعدة على تصنيف المعاني وهي جزء من لوحة التحكم في المعجم، كما أنها تقوم بترتيب عرض المعاني تلقائياً لكل مفردة اعتماداً على نسب شيوعها، ومن هذه الأدوات أداة تساعد على إبراز مدى شيوع الكلمات المستعملة في كتابة التعريفات. ومع عدم شمول هذه الدراسة لكثير من الأدوات الحاسوبية في حقل صناعة المعاجم، إلا أن من أهم ما يميزها وجود جانب عملي استفاد من تلك الأدوات الحاسوبية المشروحة فيها، مما قد يفيد القارئ حول جدوى استعمالها في مجال بناء المعاجم اللغوية لمتعلمي اللغة من غير الناطقين بها.

في نهاية هذا العرض السريع للدراسات السابقة في مجال التقنيات الحاسوبية لصناعة المعاجم، يمكننا أن نلحظ أن القليل منها فقط ركز على الجوانب التقنية التي باتت تدخل في صناعة المعاجم الحديثة الموجهة لمتعلمي اللغة من الناطقين بغيرها، إذ إن بناء هذه المعاجم بحاجة إلى أدوات خاصة قد

لا تحتاجها بعض المعاجم العامة الموجهة إلى أبناء اللغة، مع ضرورة الإشارة إلى أن سرعة ظهور وتطور مثل هذه الأدوات يجعل الإلمام بها أو شرح جميع خصائصها متعذراً في دراسة واحدة، ولذلك ستشمل دراستنا تلك التقنيات التي باتت أساسية في صناعة المعاجم بشكل عام وصناعة المعاجم اللغوية لمتعلمي اللغة على وجه التحديد، مع شرح أبرز خصائصها واستعمالاتها، متمنين أن تتوالى الدراسات التي تستقصي هذه التقنيات وتواكب ما يستجد منها.

٥ أدوات بناء معاجم المتعلمين

٥,١ نستعرض في هذا الجزء مجموعة من التقنيات الحاسوبية التي باتت جزءاً مهماً في صناعة المعاجم الحديثة، والتي سهلت على المختصين كثيراً من مراحل هذه الصناعة، ونقلتها إلى مستويات متقدمة جداً.

٥,٢ تحليل سلوك المفردة Word behaviour:

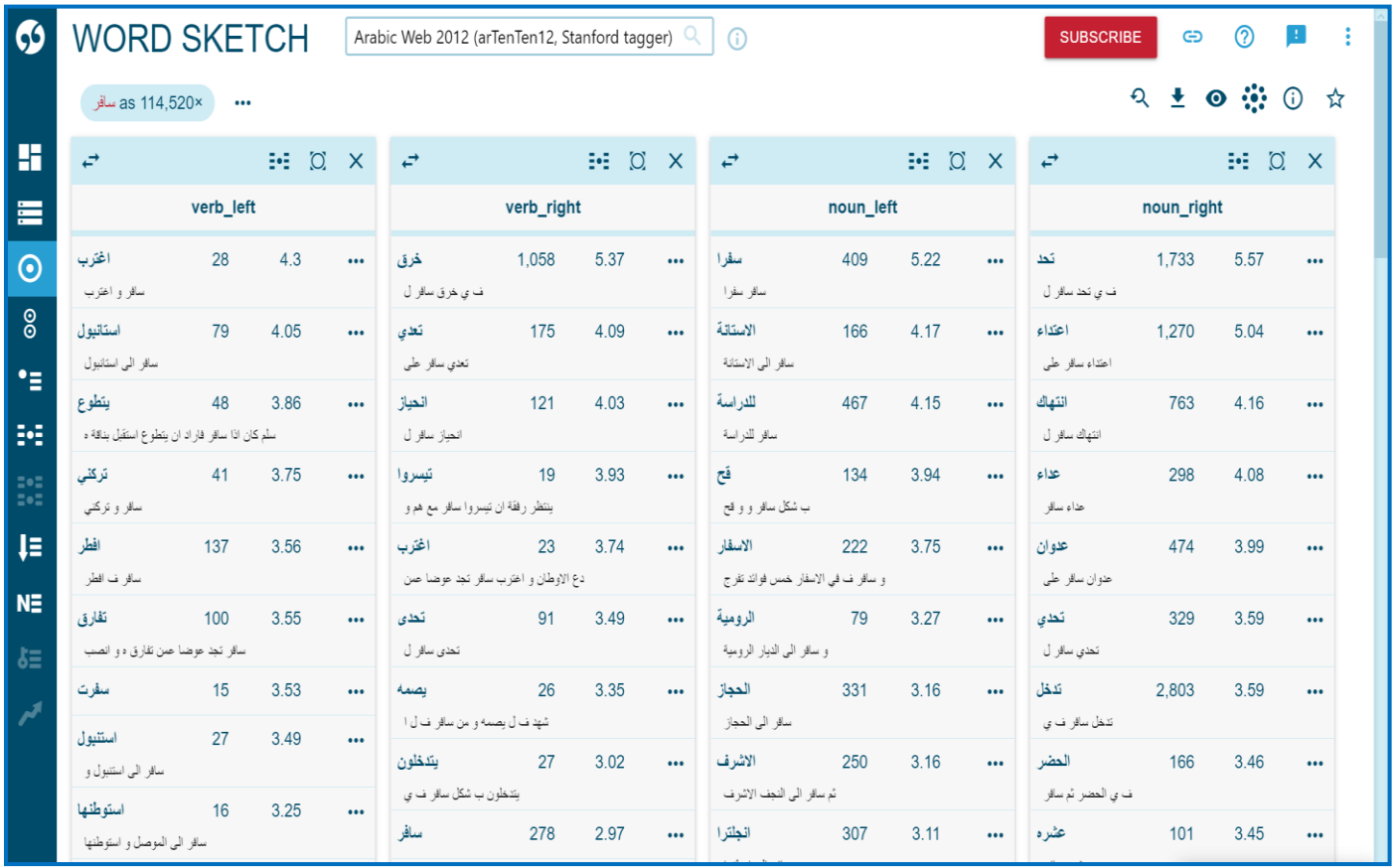
يشير إليه كيلقارف Kilgarriff وآخرون بمصطلح Word Sketches، ويُعرفونه بسلوك المفردة من حيث القواعد اللغوية والتلازم اللفظي (Kilgarriff *et al.*, 2004). ويقصد به ما يشيع استخدامه مع كل مفردة من قواعد لغوية وتلازم لفظي، كشيوع مجيء بعض الأسماء في تراكيب نحوية محددة، ومجيء أحرف جر بعينها بعد بعض الأفعال، وهكذا. وهو أقرب ما يكون لرسم مخطط لاستخدام الكلمة في سياقاتها اللغوية، وذلك من خلال جداول تُظهر ما يرد قبلها أو بعدها من مفردات أو عبارات، بغرض التعرف على سلوك المفردة في نصوص المدونة اللغوية والأشكال المختلفة للسياقات التي ترد فيها، كما تُظهر هذه الجداول عدد التكرار الفعلي لكل شكل من هذه الأشكال في المدونة اللغوية، إضافة إلى نسبة هذا التكرار إلى حجم المدونة

لتسهيل إجراء المقارنات مع مدونات لغوية أخرى. ويمكن استعراض الكلمات السابقة واللاحقة إجمالاً، أو بتوزيعها على قوائم فرعية حسب نوع الكلمة (اسم، فعل، حرف، ظرف، ...).

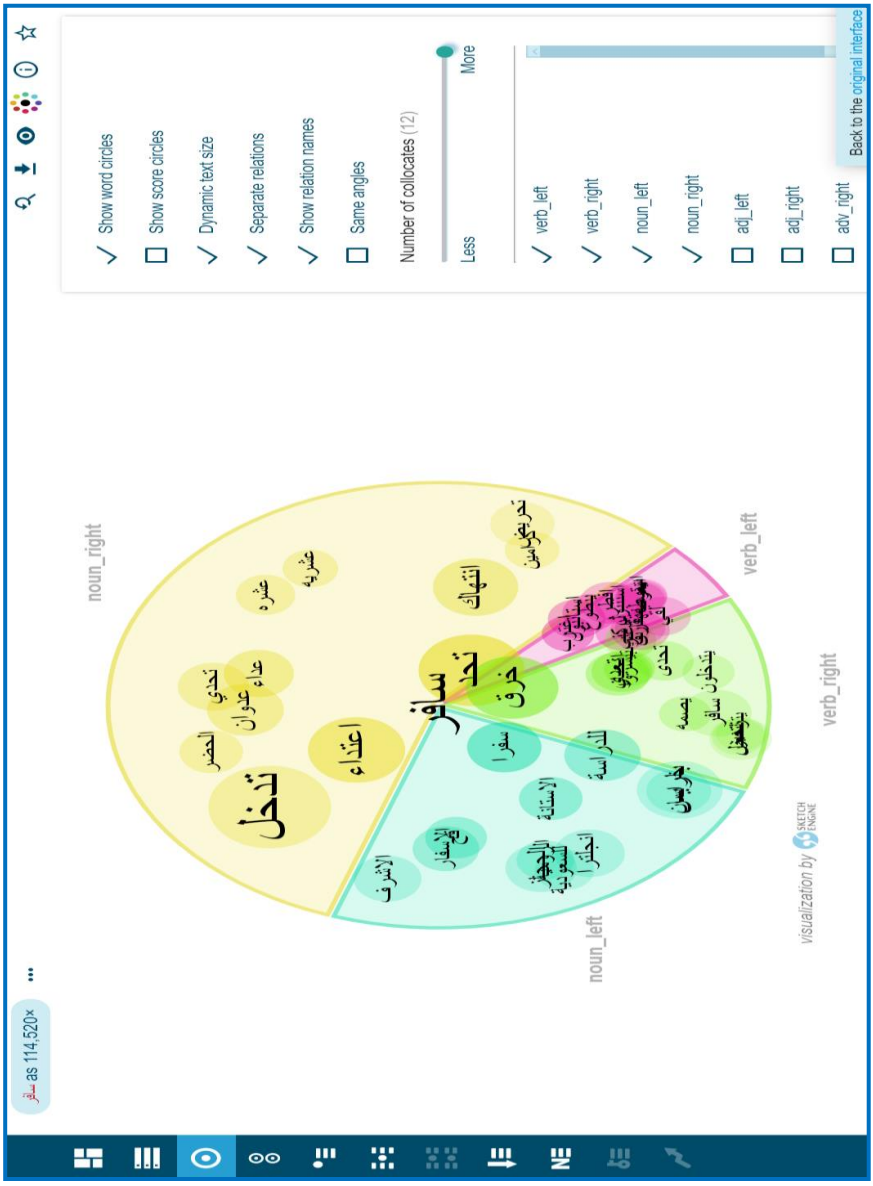
ويفيد هذا النوع من التحليل السياقي لسلوك المفردات في الكشف عن معانيها المختلفة، والتغييرات التي تطرأ على هذه المعاني وفق السياقات التي ترد فيها، كما تفيد في دراسة التلازم اللغوي من ناحية المعنى أو الدلالة Collocation مثل مجيء "**إلى**" بعد الفعل "**انتقل**"، أو من ناحية التركيب النحوي Collogation مثل مجيء حروف الجر بعد الأفعال.

من المواقع التي توفر أدوات حاسوبية لتحليل سلوك المفردات موقع Sketch Engine[1]، وموقع ArabiCorpus. فموقع Sketch Engine (Kilgarrif 2014; Kilgarriff et al., 2004) لديه أداة باسم " Word Sketche" لتحليل سلوك المفردات، وهي تظهر مخططاً لاستخدام المفردة محل البحث في جداول توضح ما يرد قبلها أو بعدها من المتلازمات موزعة وفق وسم هذه المتلازمات في المدونة اللغوية (الشكل ١)، وهناك خيار آخر لإظهار نتائج التحليل بطريقة بصرية تُسهيل قراءتها ومقارنتها (الشكل ٢).

١ https://www.sketchengine.eu


WORD SKETCH
Arabic Web 2012 (arTenTen12, Stanford tagger)
SUBSCRIBE
سافر as 114,520×
verb_left
verb_right
noun_left
noun_right
اغترب 28 4.3
سافر و اغترب
استانبول 79 4.05
سافر الى استانبول
ينطوع 48 3.86
سلم كان اذا سافر فاراد ان يتطوع استقبل بداقة ه
تركني 41 3.75
سافر و تركني
افطر 137 3.56
سافر ف افطر
تفارق 100 3.55
سافر تجد عوضا عن تفارق ه و انصب
سفرت 15 3.53
استنبول 27 3.49
سافر الى استنبول و
استوطنها 16 3.25
سافر الى الموصل و استوطنها
خرق 1,058 5.37
ف ي خرق سافر ل
تعدي 175 4.09
تعدي سافر على
انحياز 121 4.03
انحياز سافر ل
تيسروا 19 3.93
ينتظر رفقة ان تيسروا سافر مع هم و
اغترب 23 3.74
دع الاوطان و اغترب سافر تجد عوضا عمن
تحدى 91 3.49
تحدى سافر ل
يصمه 26 3.35
شهد ف ل يصمه و من سافر ف ل ا
يتدخلون 27 3.02
يتدخلون ب شكل سافر ف ي
سافر 278 2.97
سفرا 409 5.22
سافر سفرا
الاستانة 166 4.17
سافر الى الاستانة
للدراسة 467 4.15
سافر للدراسة
فح 134 3.94
ب شكل سافر و و فح
الاسفار 222 3.75
و سافر ف في الاسفار خمس فوائد تفرج
الرومية 79 3.27
و سافر الى الديار الرومية
الحجاز 331 3.16
سافر الى الحجاز
الاشرف 250 3.16
ثم سافر الى النجف الاشرف
انجلترا 307 3.11
تحد 1,733 5.57
ف ي تحد سافر ل
اعتداء 1,270 5.04
اعتداء سافر على
انتهاك 763 4.16
انتهاك سافر ل
عداء 298 4.08
عداء سافر
عدوان 474 3.99
عدوان سافر على
تحدي 329 3.59
تحدي سافر ل
تدخل 2,803 3.59
تدخل سافر ف ي
الحضر 166 3.46
ف ي الحضر ثم سافر
عشره 101 3.45

الشكل ١ : تحليل سلوك كلمة "سافر" باستخدام Sketch Engine



الشكل ١ : نموذج آخر بصري لتحليل سلوك المفردات في Sketch Engine

كما تحسب أداة Word Sketches عدد تكرار التلازم في المدونة اللغوية، إضافة إلى قوة التلازم من خلال عدة مقاييس مثل MI-score، وAssociation score، وlogDice. ويعتبر هذا الأخير أفضلها من حيث دقة النتائج، وهو يشير إلى قوة التلازم بين المفردة الرئيسة محل البحث والمفردة الملازمة لها بناءً على معادلة لقياس مدى شيوع كل منهما إضافة إلى شيوع المركب ككل، وهذا المقياس طوره رايكلي (Rychlý, 2008) – وهو أحد الأعضاء المؤسسين لموقع Sketch Engine – اعتماداً على مقياس Dice الذي كان يعطي أرقاماً صغيرة جداً، واستخدم له المعادلة الآتية:

$$logDice = 14 + \log_2 D = 14 + \log_2 \frac{2f_{xy}}{f_x + f_y}$$

ثم عدلت هذه المعادلة لاحقاً من قبل موقع Sketch Engine إلى الصيغة الآتية:

$$14 + \log_2 \text{Dice}\left(\frac{||w_1, R, w_2||}{||w_1, R, *||}, \frac{||w_1, R, w_2||}{||*, *, w_2||}\right) = 14 + \log_2 \frac{2 \cdot ||w_1, R, w_2||}{||w_1, R, *|| + ||*, *, w_2||}$$

كما أن هذا المقياس لا يتأثر بحجم المدونة اللغوية ولذا يمكن استعماله لقياس مدى التلازم بين مفردتين في أكثر من مدونة بغض النظر عن اختلاف أحجامها.

وأما موقع AraiCorpus (Parkinson, 2015) فيوفر تحليلاً مشابهاً لأداة Word Sketches في Sketch Engine، ولكن مع خيارات أقل، إذ يكتفي بعرض الكلمات السابقة واللاحقة لمفردة البحث في قوائم إجمالية دون فرزها

وفق النوع الصرفي، كما يعرض عدد مرات ورودها مع مفردة البحث في المدونة اللغوية (الشكل ٣)، وهو يكتفي بترتيب التلازم وفق عدد التكرار ولا يستعمل أياً من المقاييس الإحصائية التي يستعملها Sketch Engine لقياس قوة التلازم.

search results for ntql | نتقل in All **summary** **citations** **subsections**
download citations

words before and after

lists of before and after words occuring at least twice
click on the word for citations including that word before or after

word before	occurences	word after	occurences
ثم	1,573	من	2,860
أن	1,066	إلى	2,786
الذي	624	الى	2,361
ان	612	إلي	307
التي	504	في	306
قد	329	بعد	242
عندما	204	بعدها	226
حيث	186	الي	197
وقد	172	بين	163
	143	إليه	144
ذلك	141	هذه	142

الشكل ٣: تحليل سلوك كلمة "انتقل" باستخدام ArabiCorpus

٥,٣ قوائم شيوع المفردات Frequency Word Lists:

يمكن تعريف قوائم شيوع المفردات بأنها قوائم تضم المفردات الأكثر شيوعاً – نطقاً أو كتابةً – في مدونة لغوية معينة وفق نتائج التحليل الإحصائي. ولا يكفي حساب عدد مرات تكرار المفردة في المدونة اللغوية لاعتبارها شائعة، بل لا بد من حساب ما يسمى بنسبة الشيوع (شيوع المفردة

في كل مليون كلمة مثلاً)، وهذه النسبة تأخذ في الاعتبار عدد مرات التكرار إضافة إلى حجم المدونة التي وردت فيها، وذلك وفق المعادلة الآتية:

$$\text{Frequency permillion words}$$

نسبة شيوع المفردة في المليون

$$= \frac{\text{Word frequency تكرار المفردة}}{\text{Corpus size حجم المدونة}} \; x\; 1{,}000{,}000$$

ولذلك فإن نسب الشيوع تتأثر عادة بمعايير تصميم المدونة مثل: الحجم، ونوع النصوص (مكتوبة أو منطوقة)، وزمنها، والأوعية الأدبية، وغيرها.

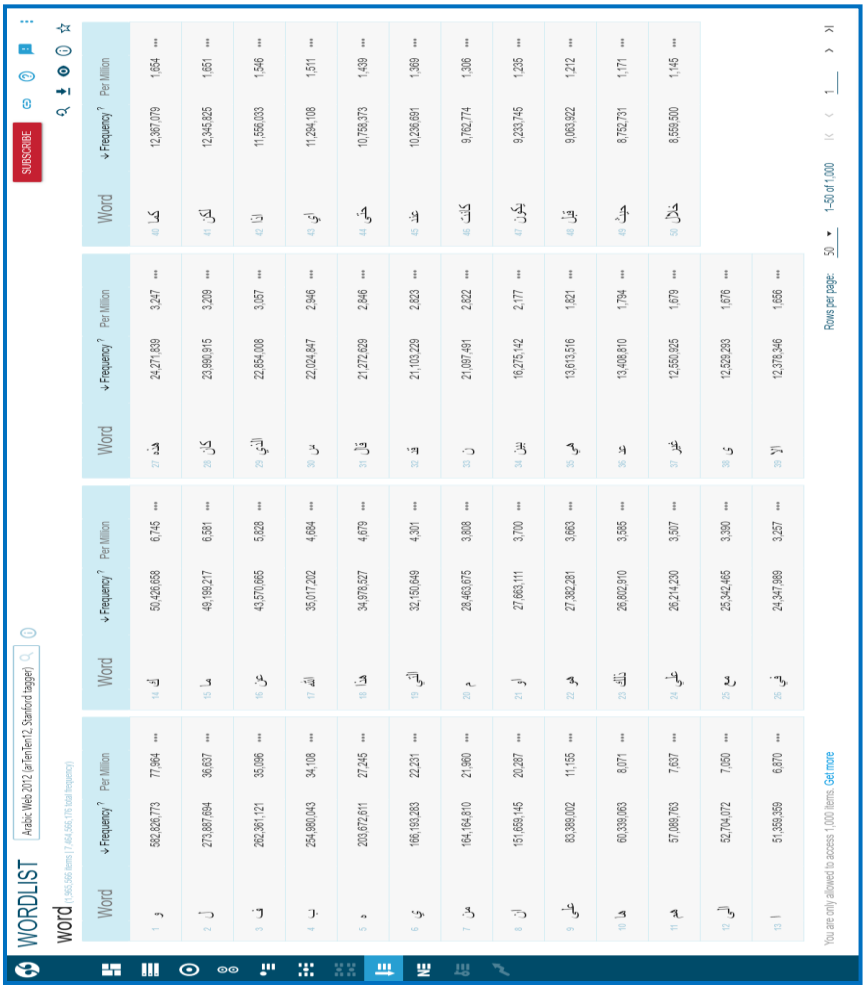
٥,٤ وقد أثبتت الدراسات أن الكلمات الأكثر شيوعاً تغطي نسبة كبيرة من الاستخدام اللغوي، ففي اللغة الإنجليزية على سبيل المثال تمثل الألف مفردة الأولى في قوائم شيوع المفردات ما نسبته ٧٢٪ من الاستخدام اللغوي العام (Cobb, 2002)، وتمثل الثلاثة آلاف مفردة الأكثر شيوعاً ما نسبته ٩٠ -٩٥٪ من المفردات المستخدمة في اللغة العامة (Laufer, 1989; Nation, 1990, 2001).

وتوجد كثير من التطبيقات الحاسوبية والمواقع الإلكترونية التي تساعد على استخراج قوائم الشيوع من المدونات ومنها:

- The KACST (King Abdulaziz City for Science and Technology) Arabic Corpora Processing Tool *Khawas* "غواص" (Al-thubaity and Al-Mazrua, 2014; Al-thubaity et al. 2014)
- aConCorde (Roberts, 2014; Roberts et al., 2006)
- AntConc (Anthony, 2005, 2014a,b)
- WordSmith Tools (Scott, 2008, 2012)

- Sketch Engine (Kilgarrif 2014; Kilgarriff et al., 2004)

ويُظهر ويُظهر الشكل ٤ المفردات الخمسين الأكثر شيوعاً في المدونة اللغوية Arabic Web 2012 (arTenTen12, Stanford Tagger) باستعمال أداة Wordlist في Sketch Engine، وإلى جوار كل مفردة يمكن ملاحظة عدد مرات التكرار Frequency ثم نسبة الشيوع في المليون Per Million.



الشكل ٢ : قائمة المفردات الشائعة في مدونة Arabic Web 2012 بواسطة أداة Wordlist في Sketch Engine

من المعاجم التي اهتمت بميزة الشيوع في بنائها معجم COBUILD (Sinclair, 1987)) الذي يشير إلى المفردات الأكثر شيوعاً برموز خاصة، كما يقوم على الشيوع في ترتيب المعاني تحت المداخل المعجمية (Stein, 1989). ومثل هذا المعجم في الاستفادة من منهجية الشيوع معجم Longman dictionary of contemporary English (Murphy, 2009)، ومعجم Macmillan English dictionary for advanced learners (Rundell, 2007)، وغيرهما، إذ يُشار إلى المداخل المعجمية الأكثر شيوعاً برموز تدل على مدى شيوعها في الاستخدام اللغوي ليوليها الطالب عناية أكبر في بداية تعلمه للغة، فمعجم Longman Dictionary of Contemporary English على سبيل المثال يُبرز ثلاثة آلاف مدخل من مداخله المعجمية التي تمثل الألفاظ الأكثر شيوعاً في اللغة الإنجليزية، وهو يميِّز بين نوعين من الشيوع: المفردات الشائعة في الكلام Speaking، والمفردات الشائعة في الكتابة Writing، كما أنه يقسم كل مجموعة إلى ثلاث درجات من الشيوع: الكلمات الألف الأولى والثانية والثالثة، فكلمة "Teach" (يُعلِّم / يُدرِّس) مثلاً رُمز لها بالعلامات التالية W2 S1، وهذا يعني أنها من الألف الثانية الأكثر شيوعاً في الكتابة، ومن الألف الأولى الأكثر شيوعاً في الكلام.

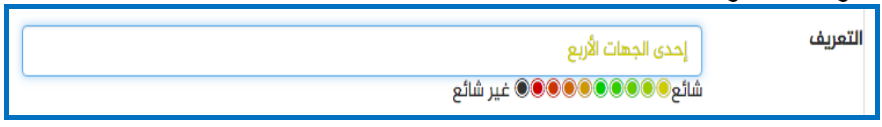
٥,٥ قياس شيوع المفردات في شرح المداخل المعجمية:

هي أداة تستخدم لتسهيل كتابة التعريفات في شرح المداخل المعجمية، خصوصاً في المعاجم الموجهة للناطقين بغير اللغة العربية، وذلك من خلال

إظهار مدى شيوع المفردات المستعملة في الشرح مقارنة بقائمة جاهزة للمفردات مرتبة حسب الشيوع.

وقد رأت بعض المعاجم الحديثة مثل معجم Longman Dictionary of Contemporary English (Murphy, 2009) أن تعتمد في كتابة شرح المداخل المعجمية Definitions على المفردات الواردة في قائمة الشيوع الخاصة بالمعجم، والتي تضم ثلاثة آلاف كلمة هي الأكثر شيوعاً في اللغة الإنجليزية، وبذلك يسهل على الطالب قراءة هذه المداخل وفهمها إن كان قد ألم بالمفردات المشمولة في قائمة الشيوع تلك.

إن وجود أدوات حاسوبية لقياس الشيوع في شرح المداخل المعجمية يُسهِّل على صناع المعاجم كتابة الشرح وخصوصاً في المعاجم الطلابية، حيث تُظهر هذه الأدوات درجة شيوع المفردات المستخدمة بمجرد إدخال نص الشرح، وذلك اعتماداً على قائمة شيوع سبق تعريفها في هذه الأداة. الشكل ٥ مثالاً لأداة حاسوبية استخدمت لكتابة تعريفات المداخل المعجمية في معجم المفردات الشائعة لمتعلمي العربية "شائع" (الفيفي وآخرون، ١٤٣٧هـ)، إذ تعتمد على قائمة فيها ٣٠٠٠ مفردة، قسمت حسب الشيوع إلى عشرة مستويات في كل منها ٣٠٠ كلمة، وبناء على هذا التقسيم تساعد الأداة على إبراز مدى شيوع الكلمات المستعملة في كتابة التعريف من خلال تغيير لون الخط – ضمن سلسلة من الألوان – ليدل على درجة شيوع كل مفردة من مفردات التعريف.

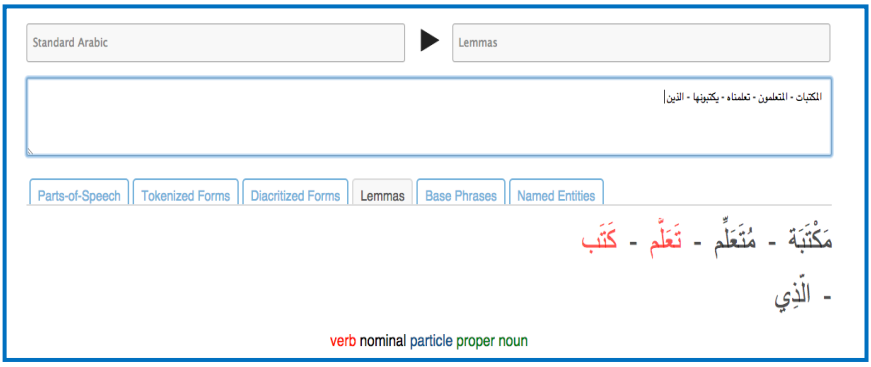


الشكل ٥ : الأداة المستخدمة لقياس شيوع المفردات في شرح المداخل المعجمية لمعجم "شائع"

٥,٦ مستخرج الأصول المعجمية Lemmatiser:

هـو أداة حاسـوبية تسـتعمل عـادة ضـمن تطبيقـات التحليـل الصـرفي الحاسوبي، ووظيفتها التعرف على الأصل المعجمي لأي مفردة Lemma، وهي تسهل على الناطقين بغير العربية البحث في المعاجم اللغوية دون الحاجة إلى معرفة القواعد الصرفية للاشتقاق التي تستعمل عادة لمعرفة أصل الكلمة وبالتالي الوصول إلى المدخل المعجمي المطلوب، إذ يتطلب البحث عن المفردات في المعاجم العربية الورقية التي تعتمد الجذع أو الجذر إلمام الطالب بطريقة إرجاع الكلمات إلى أصلها من أجل الوصول إلى الكلمات الرئيسة التي تمثل المداخل المعجمية (Head Word)، وإلا كان من الصعب على المتعلم الوصول إلى المداخل المعجمية التي يبحث عنها، فكلمة "المكتبات" نجدها في الغالب تحت المدخل المعجمي "مَكْتَبَة" بعد تجريدها من حرف التعريف (ال) وإرجاعها إلى مفردها، وكلمة "يكتبون" نجدها تحت المدخل المعجمي "كَتَبَ" بعد تجريدها من ضمير الجمع المذكر (الواو) وعلامة رفع الفعل (النون) وإرجاعها للزمن الماضي، وهكذا في كثير من الكلمات؛ والطالب الناطق بغير العربية قد يواجه مشكلة في استخدام المعاجم إن لم يكن ملماً بالقواعد الصرفية التي تعينه على معرفة أصل الكلمة، ومن هنا فإن أداة استخراج الأصول المعجمية في المعاجم الإلكترونية تقوم بهذا الدور نيابة عن الطالب، فتستقبل أي كلمة صحيحة إملائياً، سواء أكانت هي الأصل الذي يمثل اللفظة الأساسية للمدخل المعجمي أو مشتقة منه، وسواء أكانت مجردة أم مزيدة، فتُرجعها لأصلها المعجمي، وتأخذُ الطالب إلى المدخل الخاص بها مباشرة، وبعض الأدوات تعرض مجموعة من المداخل المعجمية التي لها علاقة بالكلمة المدخلة إما للتأكد من المدخل المقصود قبل الذهاب لأي منها، أو في حال تعلق الكلمة بأكثر من مدخل مثل "رجل" التي قد تكون بمعنى إنسان

مُذكَّر (رَجُل) أو عضو المشي (رِجْل)؛ وفكرة هذه الأداة مبنية على المعالجة الصرفية للكلمات، ومعرفة أصول الاشتقاقات، والحروف الأصلية والمزيدة في الكلمة، وما أضيف من سوابق أو لواحق، فتصلُ بذلك إلى أصل الكلمة ومنه للمدخل المعجمي المناسب، كما تختصر على الطالب الجهد وتقلل من احتمال الذهاب لمدخل خاطئ أو عدم إيجاد المدخل المعجمي، كما تدخل ضمن تركيب هذه الأداة أدوات أخرى مثل المصحح الإملائي الذي يقترح التصويبات المناسبة عند وجود خطأ في إدخال الكلمة، ومزيل التشكيل الذي يقوم بجزء من عملية أوسع تعرف بتسوية الكلمات Normalisation، وغيرها من الأدوات؛ ويُظهر الشكل ٦ نموذجاً للمحلل الصرفي MADAMIRA (Pasha et al., 2014) الذي يوفر ميزة استخراج الأصول المعجمية للكلمات العربية.



الشكل ٦: نموذج لاستخراج الأصول المعجمية باستخدام أداة MADAMIRA

٥,٧ كشاف السياقات Concordancer:

أداة حاسوبية تستعمل للبحث في نصوص المدونات اللغوية للكشف عن المفردات في سياقاتها اللغوية من أجل تحليلها ودراستها، ويشمل ذلك جوانب عدة منها الصرفية والنحوية والدلالية والتداولية وغيرها، وتستخدم كشافات

السياقات عند بناء المعاجم لأغراض عدة منها استخراج وتصنيف المعاني، فقبل وجود المدونات اللغوية كان المختصون في المعاجم يوظفون خبراتهم اللغوية والمعجمية عند وضع المعاني وتصنيفها، وهو ما يعرف بالحدس اللغوي Linguistic Intuition، أما في المعاجم القائمة على المدونات فإن الباحثين يعمدون إلى استخراج عدد كبير من الأمثلة السياقية للفظة المعجمية من إحدى المدونات اللغوية العامة، بغرض كشف معانيها وتصنيفها بناء على الاستخدام الفعلي لها، وهذا بالطبع لا يلغي أهمية جانب الحدس اللغوي سواء أكان عند تحديد المعاني أم عند تصنيفها، وهذا ما عبر عنه Hanks 1990)) بأنه تفاعل دقيق بين الحدس اللغوي والدليل الواقعي في المعجمية القائمة على المدونات؛ وإضافة إلى ما سبق فإن تحليل الأمثلة السياقية يساعد على تصنيف المعاني حسب شيوعها، ويضيف له بعداً إحصائياً يزيد من دقة هذا التصنيف.

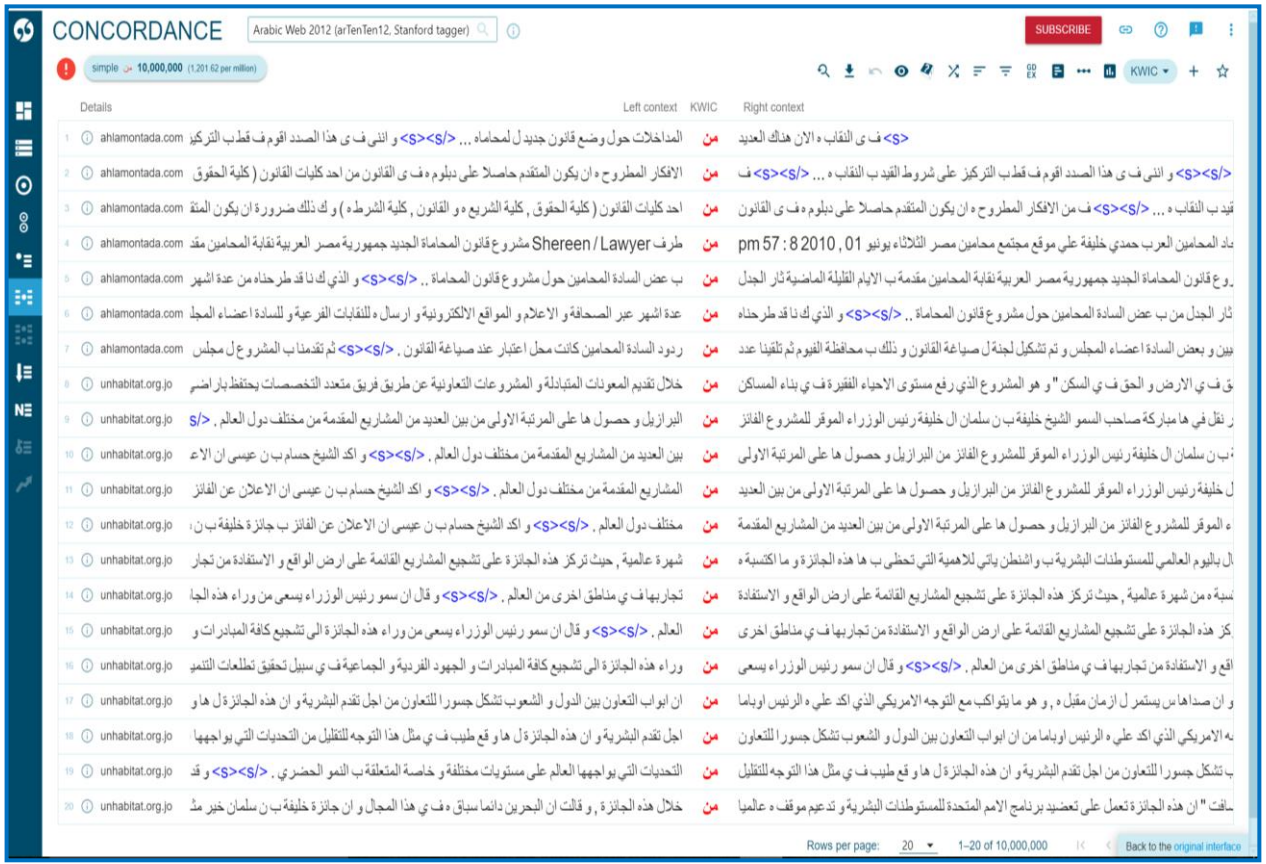
وقد أجرى الفيفي وأتويل (Alfaifi and Atwell, 2015) دراسة مقارنة لعدد من كشافات السياقات، ووضعوا عدة معايير للنظر في مدى إمكانية استخدام هذه الكشافات مع النصوص العربية، مثل قراءة الملفات النصية بتنسيق UTF-8 وتنسيق UTF-16، وإظهار النص العربي مع التشكيل بطريقة صحيحة، وإمكانية تسوية الكلمات من ناحية التشكيل والهمزة (Normalisation)، وغيرها من المعايير، وقد وجد الباحثان أن ثلاثة من البرامج المفحوصة حققت أعلى درجة في مطابقتها لهذه المعايير، وهي:

١. برنامج غواص (Althubaity and Almazrua, 2014; Althubaity et al. 2014).

٢. موقع Sketch Engine (Kilgarrif 2014; Kilgarriff et al., 2004).

٣. برنامج AconCorde (Roberts, 2014; Roberts et al., 2006).

الشكل ٧: مثال للكشاف السياقي في "غواص"



الشكل ٨: مثال للكشاف السياقي في "Sketch Engine"

الشكل ٩ : مثال للكشاف السياقي في "AconCorde"

٥,٨ المدونات اللغوية العامة General Corpora :

قد يرى البعض أن المدونات اللغوية الخام مجرد مصادر للنصوص اللغوية أكثر منها تقنيات حاسوبية ، لكنها عندما تكون مبنية وفق معايير تصميم تجعل جميع أجزائها مقروءة آلياً كالمعلومات الوصفية Metadata والوسم Tagging والتحشية Annotation وترميز هياكل الملفات Markup إضافة إلى النص نفسه Text الذي قد يتم ترميزه ضمن هيكلة الملف إلى عنوان Title ونص أساسي Mine text مقسم إلى فقرات Paragraphs وجمل Sentences وكلمات Word وهكذا ، فإنه في هذه الحالة يمكن اعتبار المدونة ضمن التقنيات التي لا غنى للمعاجم اللغوية الحديثة عنها.

ويعرف سينكلير Sinclair (1996م) المدونة اللغوية بأنها "مجموعة من النصوص اللغوية التي يتم جمعها وتصنيفها وفق معايير لغوية محددة لتكون

عينة لغوية". ويمكن الإشارة إلى المدونة اللغوية العامة بأنها تلك التي "أنشئت لأغراض مختلفة ومن نصوص متباينة. ومن أمثلتها: المدونة العربية[1] والمدونة البريطانية[2] British National Corpus" (صالح، ١٤٣٦هـ)، وتجمع مواد المدونات العامة من مصادر لغوية عدة دون حصرها في مصدر بعينه؛ ومن هنا فإن المدونات العامة مصدرٌ مناسبٌ للأبحاث التي تعنى بدراسة الاستخدام العام للغة، ومن ذلك المعاجم اللغوية، فقد أضحت كثير من هذه المعاجم بمختلف أنواعها تقوم على تحليل الاستخدام اللغوي في المدونات اللغوية العامة ومنها على سبيل المثال: معجم لونجمان للإنجليزية المعاصرة Longman Dictionary of Contemporary English (Procter, 1978)، ومعجم كولنز Collins Cobuild English Language Dictionary (Collins Cobuild, 1988)، ومعجم أكسفورد Oxford English Dictionary ((OED, 1989))، ومعجم Webster's Third International Dictionary of the English Language (Webster's Dictionary, 1961).وذلك لاستخلاص المعلومات المطلوبة للمدخل المعجمي، مثل: مشتقات اللفظ مع درجة شيوع كل مشتق وأمثلة استخدامه، والمعاني الواردة تحت كل مدخل مع أمثلة لكل معنى ومدى شيوعه، والتراكيب المتعلقة بالمدخل المعجمي مع معانيها الخاصة إن وجدت وأمثلة استخدامها، إضافة إلى التعبيرات السياقية Idioms، والمتلازمات اللغوية Collocations، والمرادفات Synonyms،

---

١ المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية (http://www.kacstac.org.sa)

٢ BNC (http://www.natcorp.ox.ac.uk)

والأضداد Antonyms ، وغير ذلك من المعلومات اللغوية التي يقدمها المدخل المعجمي لقارئه.

الشكل ١٠ : المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية (Althubaity, 2014)

الشكل ١١ : المدونة الوطنية البريطانية (The British National Corpus, 2007)

ولا يمكن العمل على المدونات اللغوية بمنأى عن الأدوات الحاسوبية التي تعين على تحليل نصوصها وتقريب محتواها ، فيعتقد الباحثون ومنهم إيفيسون

(Evison, 2010) بأن المدونة اللغوية في ذاتها لا تحتوي على معلومات جديدة عن اللغة، ولكن باستخدام البرمجيات الحاسوبية فإنها يمكن أن تزودنا بتصور جديد عن الجوانب المألوفة لدينا؛ ومن أهم هذه البرمجيات أدوات المعالجة الآلية الصرفية والنحوية والدلالية وغيرها، ومنها أيضاً أدوات تحليل المدونات التي ورد الحديث عنها في النقاط السابقة كأدوات تحليل سلوك المفردة، وأدوات استخلاص قوائم الشيوع، وكشاف السياقات وغيرها.

٥,٩ مدونات المتعلمين Learner Corpora:

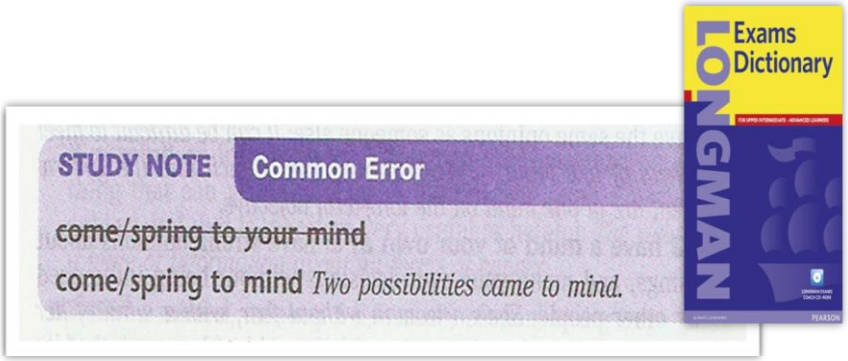
لا فرق من حيث البنية بين مدونات المتعلمين والمدونات العامة، ولذا فإن تعريف قرانجر 2002) Grangerم) لمدونات المتعلمين - وهي أشهر من عرّف هذا النوع - كان مبنياً على تعريف سينكلير 1996) Sinclairم) للمدونات العامة. إلا أن ما يميز بين هذين النوعين أمران: الأول أن مصدر البيانات لمدونات المتعلمين ينحصر في تلك المواد التي يحررها متعلمو اللغة دون غيرهم، بينما تجمع نصوص المدونات العامة من مصادر متنوعة، والثاني أن الغاية من بناء مدونات المتعلمين تدور في الغالب حول اكتساب اللغة وتعليمها، في حين تستخدم المدونات العامة لأغراض عدة لا حصر لها. ومن هنا تعرف قرانجر مدونات المتعلمين بأنها "مجموعة حاسوبية من البيانات النصية الواقعية للغة الثانية أو الأجنبية، والتي تم جمعها وفقاً لمعايير تصميم واضحة، ولغرض محدد في مجال اكتساب اللغة الثانية، أو تعليم اللغة الأجنبية، ويتم وسم هذه النصوص بطريقة معيارية ومتجانسة، مع توثيق أصلها ومصدر الحصول عليها" (قرانجر، ٢٠٠٢: ٧).

دخلت مدونات المتعلمين في صناعة المعاجم الموجهة لمتعلمي اللغة، وخصوصاً المعاجم الأحادية (Monolingual learners' dictionaries)،

حيث ينبه المعجمُ طلابَ اللغة في بعض مداخله المعجمية إلى الأخطاء الشائعة التي تقع من المتعلمين عند استخدامهم لمفردات هذه المداخل، سواء أكانت تلك الأخطاء في كتابة المفردة (خطأ إملائي)، أو في استخدامها مفردةً (خطأ صرفي)، أو في استخدمها مركبة مع غيرها (خطأ نحوي)، أو في إشارتها للمعنى المقصود في السياق (خطأ دلالي)، أو غير ذلك من الأخطاء، مع توجيه الطالب للاستخدام الصحيح؛ ويستنتجُ صانعو المعاجم هذه الأخطاء ومدى شيوعها بتحليلهم لمدونات المتعلمين خصوصاً تلك المشتملة على وسم للأخطاء اللغوية، ولذلك يشيع هذا النوع من الوسم في مدونات المتعلمين لأهميته في هذا المجال (Gillard and Gadsby, 1998; Granger, 2003; Nesselhauf, 2004).

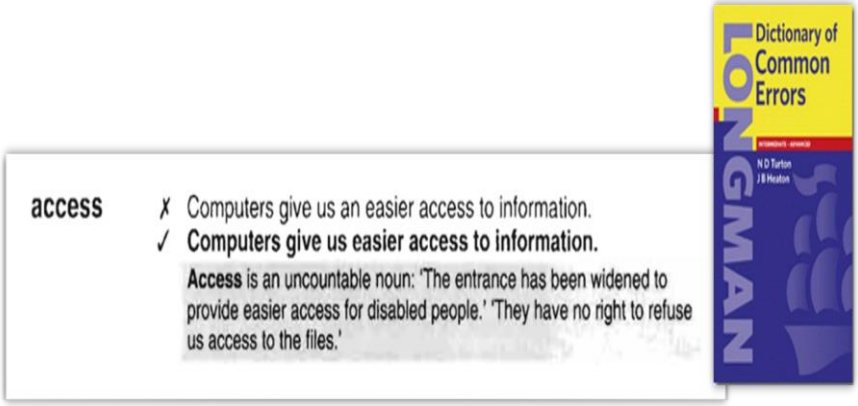
ويعد معجم Longman Essential Activator (سمرز Summers، 1997م) أول معجم يعتمد على مدونة متعلمين، حيث استُخدِمت "مدونة لونجمان" (شبكة مدونة لونجمان، ٢٠١٢م) في تأليف هذا المعجم (جيلارد وجاسبي Gillard and Gadsby، 1998م، وقرانجر، ٢٠٠٣م ب، ونسلهاف، ٢٠٠٤م).

يعرض لنا الشكل ١٢مثالاً لخطأ شائع من "معجم لونجمان للاختبارات" Longman Dictionary Exams (بولن Bullon، ٢٠٠٦م: ٩٦٩)، وقد تم وضعه على شكل ملحوظة في نهاية المدخل المعجمي لكلمة "mind" للتحذير من الأساليب الخاطئة لاستخدام هذه الكلمة والتنبيه على البدائل الصحيحة لها.



الشكل ١٢ : مثال لخطأ شائع من "معجم لونجمان للاختبارات"

وقد توسع البحث في تأليف معاجم الطلاب المبنية على مدونات المتعلمين حتى ظهرت معاجم مستقلة للتنبيه على أخطاء المتعلمين وتصويبها مثل "معجم لونجمان للأخطاء الشائعة"(٣) Longman Dictionary of Common Errors (تورتن وهيتن Turton and Heaton، ١٩٩٦ : ٥)، انظر الشكل ١٣.



الشكل ١٣ : مثال لخطأ شائع من "معجم لونجمان للأخطاء الشائعة"

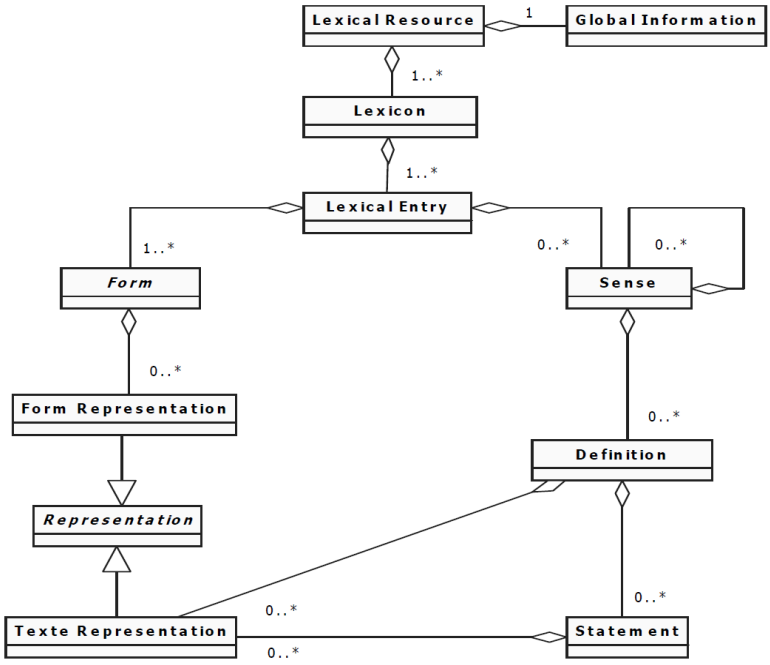
والمعاجم التي تعنى بالتنبيه على الأخطاء الشائعة قد وُجدت بكثرة في اللغة العربية، ومنها "معجم الأخطاء الشائعة" (العدناني، ١٩٨٣م)، و"المعجم الوجيز في الأخطاء الشائعة والإجازات اللغوية" (محمد، ١٤٢٦هـ)،

و"معجم تقويم اللغة وتخليصها من الأخطاء الشائعة" (أَمُّون، التاريخ غير معروف)، وهي في الغالب موجهة للناطقين باللغة العربية، كما أنها لم تستفد من التطور الحاصل في التقنيات الحاسوبية لصناعة المعاجم الحديثة.

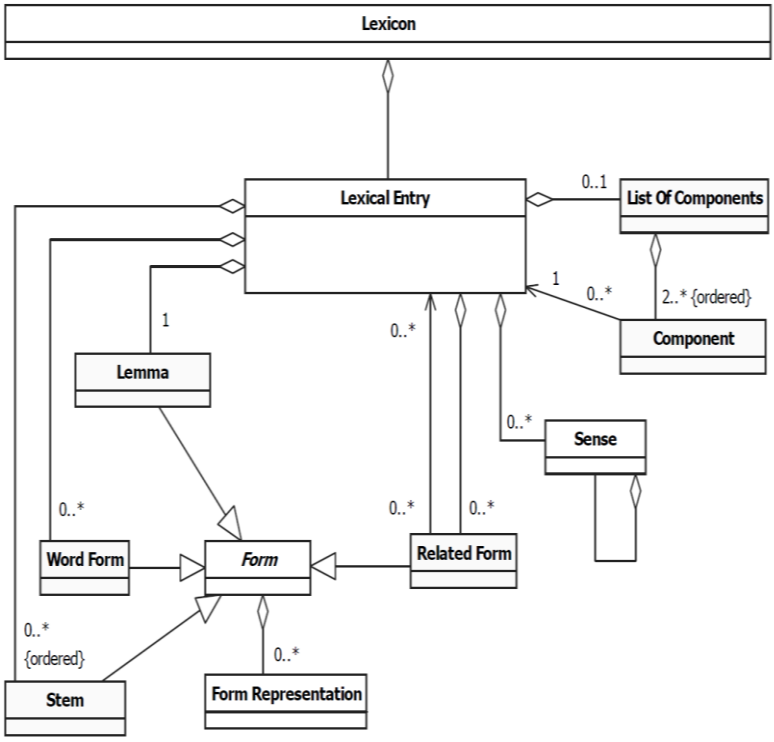
٥,١٠ إطار ترميز المعجم

وهو عبارة عن آلية معيارية لوسم المعلومات اللغوية في المعاجم الحاسوبية لتوحيد التعامل مع هيكلها ومع بياناتها لتمكين مختلف التطبيقات الحاسوبية من الاستفادة من تلك المعاجم وبأكثر من طريقة. ويستهدف هذا الإطار نوعين من المعاجم: المعاجم المصممة للمعالجة الآلية للغة العربية Natural Language Processing (NLP)، والمعاجم المقروءة آلياً Machine-Readable Dictionaries (MRD). ويقصد بهذا الأخير تلك المعاجم التي صممت في البداية لتكون معاجم ورقية. ويعد كتاب LMF Lexical Markup Framework (Francopoulo and Paroubek, 2013) المرجع الرئيس لهذا الإطار إذ يقدم شرحاً وافياً عن آلية بنائه واستعماله في سبعة عشر فصلاً شارك في تأليفها أكثر من أربعين مختصاً.

ولهذه الآلية نموذج رئيس Core Model يمثل المعلومات الأساسية للمعجم (الشكل ١٤)، تتصل به عدة نماذج فرعية لتمثيل المعلومات المعجمية الصرفية والنحوية والدلالية، ويعرض الشكل ١٥ النموذج الفرعي للمعلومات الصرفية.

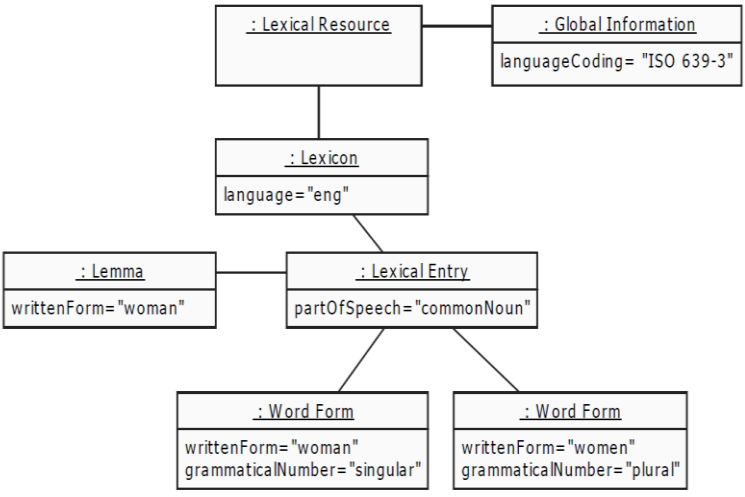


الشكل ١٤ : النموذج الرئيس لإطار ترميز المعجم (Francopoulo and Paroubek 2013: 21)



الشكل ١٥ : النموذج الفرعي لترميز المعلومات الصرفية

ويوضح الشكل ١٦ مثالاً لهيكل البيانات الصرفية لكلمة امرأة Woman، والتي تشتمل على معلومات منها: اللغة، وكلمة المدخل المعجمي، ونوعها، وطريقة كتابتها مفردةً ومجموعةً.



الشكل ١٦ : مثال لترميز المعلومات الصرفية لكلمة امرأة Woman

ويمكن ترميز هذه المعلومات وفق إطار ترميز المعجم بالطريقة الآتية:

<"LexicalResource dtdVersion="16>

<GlobalInformation>

</"feat att="languageCoding" val="ISO 639-3>

<GlobalInformation/>

<Lexicon>

</"feat att="language" val="eng>

<LexicalEntry>

</"feat att="partOfSpeech" val="commonNoun>

<Lemma>

</"feat att="writtenForm" val="woman>

<Lemma/>

<WordForm>

</"feat att="writtenForm" val="woman>

</"feat att="grammaticalNumber" val="singular>

<WordForm/>

<WordForm>

</"feat att="writtenForm" val="women>

</"feat att="grammaticalNumber" val="plural>

<WordForm/>

<LexicalEntry/>

<Lexicon/>

<LexicalResource/>

٦ استخدام أدوات بناء معاجم المتعلمين

من خلال الاستعراض السابق لأهم التقنيات الحاسوبية التي باتت من الركائز الرئيسة في صناعة المعاجم، يمكننا أن نلحظ ما ذكره العارف (١٤٢٨هـ) وأوردناه في الإطار النظري للبحث من أن استعمال مثل هذه الأدوات يتطلب مهارات خاصة – وخصوصاً في مجال الحاسب الآلي ومعالجة اللغة العربية آلياً والإحصاء والذكاء الاصطناعي والمنطق والرياضيات – قد لا تتوفر لكثير من اللغوين المشتغلين بمجال صناعة المعاجم العربية، وهو ما يوجب الاختيار بين أمرين: إما أن يشتمل الفريق الذي يعمل في صناعة المعجم على متخصصين لديهم إلمام بهذه التقنيات وآليات توظيفها بالشكل

الصحيح، إضافة إلى المتخصصين في المجالات الأخرى غير الحاسوبية التي يتطلبها المعجم؛ وإما أن يطور اللغوي المشتغل بنفسه في صناعة المعاجم من مهاراته في هذه التقنيات، ولو بشكل يسير يساعده على استعمال المهم منها في بناء أساسات معجمه، وهذا الخيار يتطلب وجود ورش عمل يقدمها المتخصصون في هذه التقنيات، سواء أكانوا من العاملين في مجال صناعة المعاجم، أم ممن يستخدمها في مجال عمله لأغراض أخرى كمعالجة اللغة الطبيعية آلياً.

كما أن هناك جانباً أكثر أهمية للتمكن من استخدام هذه التقنيات في صناعة المعاجم العربية، وهو قدرتها على معالجة النصوص العربية والتعامل معها بشكل سليم، ففيما يخص أدوات تحليل المدونات اللغوية على سبيل المثال يمكن الرجوع إلى دراسة الفيفي وأتويل (Alfaifi & Atwell, 2016) المشتملة على تقييم لمجموعة من أدوات تحليل المدونات اللغوية العربية وكشافات السياقات، إذ كان بعضها قاصراً في تعامله مع النصوص العربية وخصوصاً عندما يتضمن النص غير الحروف كحركات التشكيل والشدة والتنوين مثلاً. كما اشتملت دراسة الفيفي (Alfaifi, 2015) على تقييم لمجموعة من الأدوات الحاسوبية لوسم الأخطاء في المدونات اللغوية لمتعلمي اللغة العربية، ووجدت الدراسة أن جميع الأدوات المشمولة فيها لم تكن مؤهلة بالقدر الكافي للتعامل مع النصوص العربية ووسم الأخطاء بالشكل الصحيح. مثل هذه الإشكالات توجب على المتخصصين في مجال الحاسب الآلي واللغويات الحاسوبية العمل مع أصحاب هذه الأدوات على تطويرها وتصحيح جوانب الضعف فيها ليمكنها التعامل مع اللغة العربية، وقد يكون من الممكن بناء أدوات جديدة ذات قدرة على معالجة النص العربي؛ ليستفيد

منها المهتمون بالمعالجة الآلية للغة العربية من شتى التخصصات، والمعجميون منهم على وجه الخصوص، إذ لا شك أنها ستسهم في الارتقاء بصناعات كثيرة منها صناعة المعاجم العربية التي لا زالت متأخرة بشكل واضح عن نظيرتها في كثير من اللغات الأخرى.

* * *

الخاتمة:

تناولت هذه الدراسة مجموعة من التقنيات الحاسوبية الحديثة المستخدمة في صناعة المعاجم اللغوية خصوصاً تلك الموجهة لمتعلمي اللغة من الناطقين بغيرها، وذلك لإبراز الدور المهم الذي يمكن أن يضيفه استخدام هذه التقنيات، ومنها تحليل سلوك المفردات الذي يضيف معلومات إحصائية أساسية للمداخل المعجمية خصوصاً لقياس قوة التلازم، وقوائم شيوع المفردات التي أثبتت أهميتها في أكثر من مكان في المعجم مثل ترتيب المتلازمات في المداخل المعجمية، وكتابة التعريفات للمداخل، يسندها في ذلك أداة قياس شيوع المفردات التي تسهل قياس مدى الشيوع أثناء الكتابة. من التقنيات التي استعرضتها الدراسة كذلك مستخرج الأصول المعجمية الذي بات جزءاً رئيساً في المعاجم الأجنبية بينما لم يرَ النور إلى الآن في معاجمنا العربية رغم وجوده في بعض المحللات الصرفية العربية، ومنها كشاف السياقات الذي قد يعدّ أكثر هذه التقنيات استعمالاً من قبل اللغويين لوضوح فائدته وكذلك آلية استخدامه التي لا تتطلب خبرة حاسوبية كبيرة أو معرفة إحصائية متعمّقة. أما المدونات اللغوية سواء العامة منها أو تلك التي تتضمن نصوصاً لمتعلمي اللغة، فيمكن اعتبارها إلى الآن حبيسة الأبحاث اللغوية في مجالات مثل الصرف والنحو وتحليل الخطاب، ولم تتجاوزها إلى الدراسات المعجمية أو المشاريع العملية لصناعة المعاجم سوى ما ندر.

ولعل في استعراض الدراسة لهذه التقنيات ما يفتح آفاقاً جديدة لدى بعض المشتغلين بتأليف المعاجم أو المهتمين بهذا الحقل، ويشجع على مزيد من البحث والتأليف لصالح متعلمي اللغة الناطقين بغيرها، الذين لم ينالوا نصيبهم من المعاجم العربية التي تناسب احتياجاتهم وتراعي قدراتهم.

* * *

المراجع

السعيد، المعتز بالله (٢٠١٦م) تطبيقات متقدمة لصناعة معجم اللغة العربية للناطقين بغيرها للمستويين التعليميين B1 وB2، بحث منشور في مؤتمر إسطنبول الدولي الثاني: تعليم اللغة العربية للناطقين بغيرها: إضاءات ومعالم، ٨ -٩ / ١٠ / ٢٠١٦م، إسطنبول.

الفيفي، عبدالله (١٤٣٦هـ) مدونات المتعلمين، في كتاب المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها، تحرير العصيمي، صالح بن فهد. الطبعة الأولى، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض.

الفيفي، عبدالله، سمير الرحيلي، محمد القحطاني، عبدالرحمن العصيمي، أيمن الغامدي (١٤٣٧هـ) معجم المفردات الشائعة لمتعلمي العربية "شائع"، مؤتمر اتجاهات حديثة في تعليم العربية لغة ثانية، ٢٣ -٢٤ ربيع الثاني ١٤٣٧هـ (٢ -٣ February 2016)، معهد اللغويات العربية، جامعة الملك سعود، الرياض.

المنظمة العربية للتربية والثقافة والعلوم (١٩٨٩م) المعجم العربي الأساسي، طبعة لاروس.

حمادة، سلوى (٢٠٠٩م) المعالجة الآلية للغة العربية: المشاكل والحلول. الطبعة الأولى، دار غريب للطباعة والنشر والتوزيع، القاهرة.

صالح، محمود إسماعيل (١٤٣٦هـ) المدونات اللغوية وكيفية الإفادة منها، في كتاب المدونات اللغوية العربية: بناؤها وطرائق الإفادة منها، تحرير العصيمي، صالح بن فهد. الطبعة الأولى، مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، الرياض.

عمر، أحمد مختار (١٤١٨هـ) صناعة المعجم الحديث. الطبعة الأولى، عالم الكتب، القاهرة.

علي، نبيل (١٩٨٨م) اللغة العربية والحاسوب (دراسة بحثية)، دار تعريب.

الموسى، نهاد (٢٠٠٠م) العربية: نحو توصيف جديد في ضوء اللسانيات الحاسوبية. الطبعة الأولى، المؤسسة العربية للدراسات والنشر، بيروت.

* * *

Analysis and Disambiguation of Arabic. In proceedings of the International Conference on Language Resources and Evaluation (LREC2014), Reykjavik, Iceland. http://www.lrec-conf.org/proceedings/lrec2014/pdf/593_Paper.pdf

Procter, Paul (1978). Longman Dictionary of Contemporary English. Longman Group Limited, Harlow and London, England.

Roberts, A., Al-Sulaiti, L. and Atwell, E. (2006). aConCorde: Towards an open-source, extendable concordancer for Arabic. Corpora, vol. 1, pp 39-60.

Roberts, Andrew (2014). aConCorde [Computer Software]. Retrieved 6 April 2014, http://www.andy-roberts.net/coding/aconcorde

Rychlý, Pavel (2008). A lexicographer-friendly association score. Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN, 6–9.

Scott, Mike (2008). Developing Wordsmith. International Journal of English Studies. 8 (1), 95-106.

Scott, Mike (2012). WordSmith Tools version 6 [Computer Software], Liverpool: Lexical Analysis Software, Retrieved 16 September 2014, http://www.lexically.net/wordsmith

Sinclair, John (1996). EAGLES. Preliminary recommendations on Corpus Typology, Retrieved from: http://www.ilc.cnr.it/EAGLES/corpustyp/corpustyp.html

The British National Corpus, (2007). Version 3, BNC XML Edition, Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Webster's Dictionary (1961). Webster's Third International Dictionary of the English Language. Spring-field, MA: Merriam.

❊ ❊ ❊

Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). The Sketch Engine. In the proceedings of the Euralex, 6-10 July 2004, pp 105-116. Lorient, France.

Kilgarriff, Adam (2014). Sketch Engine [Computer Software]. Retrieved 6 April 2014, http://www.sketchengine.co.uk/

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), Second Language Vocabulary Acquisition (pp. 20–34). New York, NY: Cambridge University Press.

McEnery, A. M. (2003). Corpus linguistics. In R. Mitkov (Ed.), The Oxford handbook of computational linguistics (pp. 448–463). Oxford, UK: Oxford University Press.

Murphy, Michael. (2009). Longman dictionary of contemporary English (5th Ed.). Harlow: Pearson Education.

Nation, I. S. P. (1990). Teaching and learning vocabulary. Rowley, MA: Newbury House.

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge, England: Cambridge University Press.

Nesselhauf, Nadja (2004). Learner Corpora and Their Potential in Language Teaching. In: Sinclair, J. (Ed.), How to Use Corpora in Language Teaching. Amsterdam & Philadelphia, Benjamins, pp. 125-152.

OED. (1989). Oxford English Dictionary. Oxford: Clarendon.

Parkinson, D. (2015). arabiCorpus. Retrieved 08 February 2015 from http://arabicorpus.byu.edu

Pasha, Arfath, Al-Badrashiny, Mohamed, Diab, Mona, El Kholy, Ahmed, Eskander, Ramy, Habash, Nizar, Pooleery, Manoj, Rambow, Owen, and Roth, Ryan M. (2014). MADAMIRA: A Fast, Comprehensive Tool for Morphological

Anthony, Laurence (2014b). AntConc 3.4.2 - Readme. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software/antconc341/AntConc_readme.pdf

Cobb, Tom. (2002). 'Why and how to use frequency lists to learn words'. The Compleat Lexical Tutor for Datadriven Learning. Retrieved 11 December 2015, from http://www.lextutor.ca/research.

Collins Cobuild. (1988). Collins Cobuild English Language Dictionary. Glasgow: Collins.

Evison, Jane (2010). What are the basics of analysing a corpus? In: Anne O'Keeffe and Michael McCarthy (Eds.), The Routledge Handbook of Corpus Linguistics. Routledge, London, UK.

Francopoulo, Gil and Paroubek, Patrick (2013). LMF Lexical Markup Framework. John Wiley & Sons: Hoboken.

Gillard, Patrick and Gadsby, Adam (1998). Using a learners' corpus in compiling ELT dictionaries. In Granger (Ed.) Learner English on Computer. Longman, London and New York, pp. 159-171.

Granger, Sylviane (2002). A Bird's-eye View of Computer Learner Corpus Research. In S. Granger, J. Hung and S. Petch-Tyson (Eds.), Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Amsterdam & Philadelphia, Benjamins, pp. 3-33.

Granger, Sylviane (2003). Error-tagged learner corpora and CALL: A promising synergy. CALICO Journal, 20(3), 465–480.

Hanks, Patrick (1990). Intuition and Evidence in Lexicography. In Tomaszczyk, Jerzy and Barbara Lewandowska-Tomaszczyk (Eds.), Meaning and Lexicography. John Benjamins Publishing: Amsterdam/Philadelphia.

Hanks, Patrick (2003). Lexicography. In Mitkov, Ruslan (Ed.), The Oxford Handbook of Computational Linguistics. New York: Oxford University Press.

Saleh, Mahmoud Ismail. Al-mudawnat Al-lughawiyah wa kayfiyat al-ifadah minha (Linguistic Corpora). In Arabic Corpora: Structure and uses. 1st edition, King Abdullah Ibn Abdulaziz International Center for Arabic. Al-Osaimi, Saleh Ibn Fahd (ed.). 1st edition, King Abdullah bin Abdulaziz International Center for Arabic Language, Riyadh, 1436.

The Arab League Education, Cultural and Scientific Organization ALECSO. Al-Mu'jam Al-Arabi Al-Asassi (The Basic Arabic Dictionary), Larose, 1989.

English References

Alfaifi, Abdullah (2015). Building the Arabic Learner Corpus and a System for Arabic Error Annotation. PhD thesis, University of Leeds.

Alfaifi, Abdullah and Atwell, Eric (2015). Comparative evaluation of tools for Arabic corpora search and analysis. International Journal of Speech Technology. DOI: 10.1007/s10772-015-9285-5.

Al-Thubaity, A., and Al-Mazrua, M. (2014). Khawas: Arabic Corpora Processing Tool USER GUIDE. Retrieved 6 April 2014, from: http://sourceforge.net/projects/kacst-acptool/files/?source=navbar

Al-Thubaity, A., Khan, M., Al-Mazrua, M., and Almoussa, M. (2014). KACST Arabic Corpora Processing Tool "Khawas" [Computer Software]. Retrieved 6 April 2014, from: http://kacst-acptool.sourceforge.net/

Althubaity, Abdulmohsen (2014). A 700M+ Arabic corpus: KACST Arabic corpus design and construction. Language Resources and Evaluation: 1–31.

Anthony, Laurence (2005). AntCone: Design and development of a freeware corpus analysis toolkitfor the technical writing classroom. Proceedings of IPCC International Professional Communication Conference, pp 729-737.

Anthony, Laurence (2014a). AntConc, (Version 3.4.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

List of References:

Al-Fifi, Abdullah, Samir Al-Rehaili, Mohammed Al-Qahtani, Abdulrahman Al-Osaimi and Ayman Al-Ghamdi. Mu'jam al-mufradat al-sha'i'ah limuta'alimi al'arabiyah "sha'i'" (Dictionary of Common Vocabulary for Learners of Arabic "sha'i'", Conference of Modern Trends in Teaching Arabic as a Second Language, 23-24 Rabi Al-Thani 1437 (2-3 February 2016), Institute of Arabic Linguistics, King Saud University, Riyadh, 1437.

Al-Fifi, Abdullah. Learner Corpora. In Arabic Corpora: Structure and uses. 1st edition, King Abdullah Ibn Abdulaziz International Center for Arabic, Riyadh, 1436.

Ali, Nabil. Allughah al-'Arabiyah wa Al-Hasoub (The Arabic language and computer (research study), Dar Ta'rib, 1988.

Al-Mousa, Nihad. Al'Arabiyah: Nahwa Tawseef Jadid fi daw' al-lisaniyat al-hasoubiyah. (Arabic: Towards a new Description in the light of computational linguistics. 1st edition, Arab Foundation for Studies and Publishing, Beirut, 2000.

Al-Saeed, Al-Moataz Billah. Tatbiqat mutaqadimmah lisina'at mu'jam Allughah al-Arabiyah lil-natiqeen bighayriha lillmistawayan al-ta'limiyayn B1 wa B2 (Advanced Applications for Compiling Arabic Dictionary for Foreign Speakers: Levels B1 and B2. Istanbul International Conference II: Teaching Arabic to Foreign Speakers: Illuminations and Milestones, 8-9/10/2016, Istanbul, 2016.

Hamada, Salwa. Almu'alajah al-Aliyah lillughah al'Arabiyah: almashakil wa alhulul (The Arabic Language Processing: Problems and Solutions). 1st edition, Dar Gharib for Printing, Publishing and Distribution, Cairo. 2009.

Omar, Ahmed Mukhtar. Sina'at al-mu'jam al-hadith (the Industry of modern lexicon). 1st edition, 'alam alkutub, Cairo, 1418.

Computational Technologies of Dictionaries for non-Native Language Learners

Dr. Abdullah Alfaifi
Assistance Professor in Computational Linguistics
Al-Imam Mohammad Ibn Saud Islamic University

Abstract:

This paper reviews a range of computer techniques utilized by specialists in the compilation of modern dictionaries, namely those directed to non-native language learners. Among these techniques is the behavioral analysis of words in their linguistic contexts, which helps detect the collocational forms that are supposed to be included in the lexicon entries, in addition to measuring the collocation strength according to which these collocations are to be ordered. Also, one of these techniques is the common-vocabulary lists and the tools for measuring their commonness, which is used in writing the explanation of lexicon entries. Another technique includes vocabulary-root lemmatization, which is mainly used in electronic dictionaries, especially students' dictionaries, as well as the context-finder technique, which has become essential for researchers in language blogs, whether they are general or related to language learners. In this study, the researcher sought to link these computer techniques to the industry of modern lexicons, especially those directed to language learners, to highlight the role of modern technological advancements and the feasibility of utilizing them in an industry is as important as that of language lexicons.

Key words: Dictionary, Computational Technologies, Concordancer, Lexical Entries, Corpora, Word Frequency.