مجمع الملك سلمان العالمي للغة العربية
King Salman Global Academy for Arabic Language

مجلة اللسانيات العربية، العدد ١٩، ذو الحجة، ١٤٤٥هـ
ردمد (ورقي): ٩٩٥٥-١٦٥٨ ردمد (رقمي): ٩٨٥٨-١٦٥٨

اللسانيات العربية
The Arabic Linguistics Journal

The Arabic Linguistics Journal, No. 19; july, 2024
ISSN: 1658-9955 E-ISSN: 1658-9858

arabiclisa@ksaa.gov.sa

استثمار لسانيات المدونات في صناعة المعاجم المدرسية

Investing Corpus Linguistics in school Lexicography

حافظ معاشي[1]* حكيمة خمار[2]

[2،1] قسم اللغة العربية، كلية الآداب والعلوم الإنسانية، جامعة محمد الخامس، الرباط، المغرب

توثيق البحث APA Citation:

معاشي، حافظ، وخمار، حكيمة. (2024). استثمار لسانيات المدونات في صناعة المعاجم المدرسية. *مجلة اللسانيات العربية*، *19*، 86-110.

استقبل في: 24-07-1445/ **رُوجع في:** 29-09-1445/ **قُبل في** 08-11-1445/ **نُشر في:** 25-12-1445

Received on: 2024-02-05 / **Revised on:** 2024-04-08 / **Accepted on:** 2024-05-16 / **Published on:** 2024-07-01.

الملخص

أحدثت لسانيات المدونات ثورة في مجال صناعة المعاجم، من خلال تغيير بعض النظريات والممارسات المعجمية، وتطبيق مجموعة من المناهج والمقاربات المستخدمة في تحليل ودراسة البيانات اللغوية. يهدف هذا البحث إلى إبراز علاقة لسانيات المدونات بمجالات من قبيل الصناعة المعجمية، واللسانيات الحاسوبية، والتعليم. ويُركز البحث على كيفية استثمار لسانيات المدونات في المجال التعليمي من خلال صناعة معاجم مدرسية تواكب المقررات التعليمية، وتضم المفردات والمصطلحات المستعملة في الكتب الدراسية. ويقدم البحث نموذجًا عن كيفية بناء مدونة المعجم المدرسي انطلاقًا من الكتب الدراسية المقررة عبر مراحل مختلفة، وبمساعدة أدوات تقوم على المعالجة الآلية للغات الطبيعية. ويُظهر البحث أن لسانيات المدونات منهجية فعالة في تطوير صناعة معاجم مدرسية تلبي احتياجات الطلاب وتساعدهم على تنمية رصيدهم اللغوي والمعرفي وتطويره وتحسين مهاراتهم اللغوية.

الكلمات المفتاحية: لسانيات المدونات، المعجم المدرسي، صناعة المعاجم، المدونة، توسيم المدونات.

Abstract

Corpus Linguistics has revolutionized the field of Lexicography by changing traditional theories and practices and introducing innovative methodologies for studying and analyzing language data. This paper aims to demonstrate the relationship between corpus linguistics and fields such as Lexicography, Computational Linguistics, and Education. Its primary objective is to explore how corpus linguistics can be leveraged in the educational field to build school dictionaries that align with educational curricula and include vocabulary and words from textbooks. The study proposes a model for building a school dictionary corpus based on textbooks prescribed at different levels using Natural Language Processing tools. It demonstrates that corpus linguistics is an effective methodology for developing school dictionaries that meet the needs of students and help them develop and enhance their linguistic and cognitive abilities and improve their language skills.

Keywords: Corpus Linguistics, School Dictionary, Lexicography, Corpus, Corpus tagging.

*المؤلف المراسل *Corresponding author*
Email: hafid.maachi@um5r.ac.ma

2024© حقوق النشر والملكية الفكرية محفوظة لمجلة اللسانيات العربية وللمؤلفين بموجب ترخيص:
Creative Commons Attribution 4.0 International License.

مقدمة

تشهد ساحة التعليم في البلدان العربية غيابا ملحوظا لمعاجم مدرسية تواكب المقررات التعليمية وتضم المفردات والمصطلحات المستعملة في الكتب الدراسية، وتنبني على دراسات ومنهجيات ومقاربات علمية وتربوية تهم الصناعة المعجمية في هذا المجال. وتكتسي المعاجم المدرسية أهمية بالغة لجميع مكونات المنظومة التعليمية، حيث تعد من الأدوات التعليمية والبيداغوجية التي تسهم في تطوير الرصيد اللغوي والمعرفي للتلاميذ، وتساعدهم على تجاوز الإشكالات اللغوية والمعرفية التي تواجههم خلال مسارهم الدراسي، وتمكنهم من اكتساب مجموعة من المهارات. وعليه، فإن لسانيات المدونات (Corpus Linguistics) تعد المنهجية أو الأسلوب (Methodology) اللساني الأنسب لدراسة الكتب الدراسية وتحليلها، واستخدامها كمدونات (Corpus) يتولى المعجميّون تحليلها ودراستها باعتماد أدوات وبرامج حاسوبية تقوم على تقنيات المعالجة الآلية للغات الطبيعية (Natural Language processing)، وتساعدهم على استخراج المواد المعجمية والمعلومات المتعلقة بها. وتنهض لسانيات المدونات بأدوار فعالة في مجالات اللسانيات التطبيقية، والصناعة المعجمية (Lexicography) من ضمن هذه المجالات التي تسهم فيها لسانيات المدونات عبر مجموعة من المناهج والمقاربات المستخدمة بشكل فعال لتطوير هذا المجال.

يناقش هذا البحث في المحور الأول مفهوم المعجم المدرسي، وفي المحور الثاني مفهوم لسانيات المدونات. ويتناول المحور الثالث دور لسانيات المدونات في صناعة المعاجم. وأما المحور الرابع، فيتطرق إلى مدى الاستفادة من لسانيات المدونات في صناعة المعاجم المدرسية، ويخصص المحور الخامس والأخير لتقديم نموذج عن كيفية بناء مدونة المعجم المدرسي.

1. مفهوم المعجم المدرسي

تعدّ المعاجم المدرسية من الوسائل والأدوات التعليمية والبيداغوجية الأساسية في العملية التعلمية والعلمية، وتضطلع بمجموعة من الأدوار والوظائف المهمة، ولها مواصفات وخصائص تجعل منها أداة تعليمية محورية، إذ تسهم في عملية الاكتساب والتعلم اللغوي والمعرفي. ورغم ذلك، فإن الحقل التعليمي في الدول العربية يشهد غيابا شبه تام لمعاجم مدرسية مبنية على أساس محتوى الكتب الدراسية المقررة، وتتضمن المفردات والمصطلحات التي يحتاجها التلميذ لتطوير رصيده اللغوي والمعرفي وتنميته، وتساعده على تجاوز جملة من الإشكاليات اللغوية والمعرفية التي تواجهه خلال مساره الدراسي، وتكسبه مجموعة من المهارات الضرورية التي تؤهله لمواصلة مسيرته التعليمية بنجاح.

وللمعاجم المدرسية في العالم العربي تسميات مختلفة ومتعددة، فنجد منها المعجم المدرسي، والمعجم التعليمي، ومعجم الناشئة، والمعجم الطلابي، والقاموس المدرسي، والمعجم المرحلي، وغيرها من التسميات. وتدل نسبة الصفة (مدرسي) إلى (المعجم) على الارتباط بالمدرسة والمناهج والمقررات الدراسية والمستويات التعليمية (لحسن ومرادي، 2020، ص 211). وقد أشار عباس الصوري (1988) إلى أن المعجم المدرسي يرتبط بالمناهج والمقررات الدراسية التي تدرس في مستوى معين، ويعكس المضامين التعليمية التي يتعرض لها المتعلم، وبالتالي الحاجة للبحث في المعجم المدرسي عن المضامين الواردة لاستجلاء ما غمض منها أو إغناء رصيده منها (ص 21).

أما عبد الغني أبو العزم (2010) فيعرّف المعجم المدرسي بأنه "أداة ثقافية بامتياز، لارتباطه ارتباطا عضويا بالعملية التعليمية في جميع مراحلها، أي من التعليم الأساسي بما في ذلك روض الأطفال، مرورا بالتعليم الإعدادي والثانوي، ووصولا إلى التعليم الجامعي." (ص 1). وبالتالي فإن فالمعجم المدرسي هو مجموع الوحدات المعجمية المتداولة فعليا في الكتب المدرسية، والسياق التعليميّ المقاميّ والمقاليّ الذي تأتي ضمنه هذه الكتب (قريش، 2007، ص 177).

ويؤكد محمد الرفيق (2020) أن من الضروري أن يتوفر لكل مرحلة دراسية معاجمها الخاصة التي تراعي خصوصية المتعلم النفسية والعمرية وما يحتاجه من معرفة، وأن تضم مكونات المعجم المدرسي مداخل معجمية وتعريفات تتناسب مع سن التلميذ ومرحلته التعليمية (ص 52). وهذا ما أكده أيضا أبو العزم (2010)، حيث أشار إلى أن لكل مرحلة تعليمية معجمها الخاص والمتميز بإطار ومستوى ومعايير وشروط لإنجازه (ص 41).

يعدّ المعجم المدرسي أداة تعليمية ذات أهمية بالغة في تطوير القدرات اللغوية والمعرفية لدى التلاميذ، وتحسين مهاراتهم اللغوية، وتجاوز الإشكالات اللغوية والمعرفية التي تواجههم. وعلى الرغم من ذلك، هناك غياب شبه تام للمعجم المدرسي في المنظومة التعليمية العربية، مما يُشكل تحدّيًا كبيرًا يهدد جيل المستقبل.

وقد أكد عبد العلي الودغيري (2001) أنّ من الضروري إعادة النظر في الكيفية التي تبنى بها المعاجم المدرسية، والاستفادة من مناهج وطرق علمية حديثة تعنى بالصناعة المعجمية واستثمارها في مجال التعليم من جهة، ومن خلال معالجة مجموعة من الإشكاليات التي تعاني منها المعاجم المدرسية القليلة الموجودة في العالم العربي، والتي تتسم بالاحتكام إلى الذوق الشخصي والرأي الذاتي، ولم تبُن على أساس علمي مدروس، لا سيما من حيث اختيار مدونتها وانتقاء المناسب من رصيدها ومداخلها، ولم تنطلق من تصور دقيق مبني على دراسة ميدانية إحصائية لمعرفة حاجات المتعلم بعد تحديد مستواه، ونوعيته والهدف من تعليميه (ص 234).

وفي السياق نفسه، أشار عبد العزيز قريش (2007) إلى أن الساحة التعليمية في المغرب تفتقر إلى معجم مدرسي مستقل عن الكتب الدراسية، وقد أكد أن غياب هذا المعجم ينعكس على مردودية التلميذ بصفة خاصة والمنظومة التعليمية بصفة عامة، وأن لذلك عدة مظاهر، مثل إهمال السياق وحيثياته المقامية والمقالية خلال البحث في المعاجم اللغوية عن المفردات المضمنة في الكتب الدراسية، وإغفال المعلومات الصوتية والصرفية والتركيبية الخاصة بالمفردات الواردة في المعاجم المدرسية المتاحة، وانشطار المعجم المدرسي بين اللغة العامة واللغة الخاصة، وانحصار المعجم المدرسي غالبا في شرح الكلمات والمفردات والاقتصار على ذلك، والإسهاب في إكساب المتعلم اللغة، وإغفال الدراسات العلمية والمناهج والمقاربات التعليمية ذات الصلة بالبيداغوجيا والتواصل، وعدم الاهتمام بالصناعة المعجمية المدرسية والتأليف المدرسي، وعدم توفر المعجم المدرسي على المضامين الواردة في الكتب الدراسية المقررة وفي المناهج التعليمية (ص 181 و182).

يُعدّ وجود المعجم المدرسي داخل المنظومة التعليمية ضرورة قصوى نظرًا إلى الأدوار والوظائف المهمة التي يُؤدّيها، وكذلك السمات والخصائص التي تجعله أداة تعليمية جوهرية. وبالتالي، فإنّ بناء مدونة على أساس الكتب الدراسية واتّباع نهج لسانيات المدونات في دراستها وتحليلها من شأنه أن يسهم في إعداد معاجم مدرسية تناسب مراحل التلميذ العمرية والتعليمية وتلبي احتياجاته اللغوية بشكل أفضل.

2. مفهوم لسانيات المدونات

منذ ستينيات القرن العشرين، شهد مجال لسانيات المدونات تقدما ملحوظا نتيجة للتطور التكنولوجي الذي مر به ميدان الحاسوب، من خلال تخزين كميات ضخمة من البيانات اللغوية واسترجاعها وتحليلها. فأصبح التقليد التجريبي لتحليل اللغة بمثابة الأساس للنموذج الكمي والنوعي المعروف باسم لسانيات المدونات (McEnery وآخرون، 2019، ص 74). وأشار

كينيدي (Kennedy، 2014) إلى أنه على مدى العقود الثلاثة الماضية، أدى تجميع وتحليل المدونات اللسانية في قواعد البيانات المحوسبة إلى ظهور حقل علمي جديد يطلق عليه اسم لسانيات المدونات (ص 01).

يقابل مصطلح مدونة باللغة الإنجليزية (Corpus)، وهي كلمة لاتينية تعني الجسد، وجمعها (Corpora) أو (Corpuses)، ولها عدة مقابلات في العربية، منها: المدونات اللغوية، والذخائر النصية، والذخائر اللغوية، والمكتنزات النصية، والمتون اللغوية (الدكروري، 2018، ص 28). وتتألف المدونة من مجموعة كبيرة من النصوص اللغوية في صيغة إلكترونية، يتم اختيارها بناء على معايير خارجية لتمثل - قدر الإمكان - لغة أو مجموعة متنوعة من اللغات بوصفها مصدرا للبيانات في البحث اللغوي (Sinclair، 2005، ص 16).

حسب ديفيد كريستال (Crystal، 2008)، فإن لسانيات المدونات هي العلم الذي ينطلق من الوصف اللغوي للنصوص اللغوية أو يتخذ من الشواهد اللغوية الواقعية داخل النصوص نقطة انطلاق لتحديد فرضيات البحث اللغوي (ص 117). ويشير علي القاسمي (2019) إلى أن لسانيات المدونات هي العلم الذي يدرس الظواهر اللغوية من خلال مدونة لسانية أو مجموعة من النصوص تقرأ آليا بواسطة الحاسوب، كما يدرس اللغة في استخدامها الفعلي، أي من خلال نصوص مقتبسة من العالم الحقيقي (ص 708).

ويرى المعتز بالله السعيد (2019) أن لسانيات المدونات منهج لغوي حديث نسبيا، يهدف إلى وصف واقع اللغة، اعتمادا على مجموعة من النصوص التي تمثل الواقع اللغوي، وذلك من خلال مناهج مختلفة من التحليل اللغوي، من بينها الوصفي، والمعياري، والتاريخي، والمقارن، والتقابلي، وبمستوياته الصوتية، والصرفية، والتركيبية، والدلالية، والمعجمية، كما يهدف إلى التحقق من فرضيات قائمة حول لغة معينة أو مجموعة من اللغات المشتركة في بعض خصائصها، وأداة البحث في هذا المنهج هي المدونات اللسانية باعتبارها عينة البحث الأساسية التي يتعامل معها آليا (ص 53).

تعد لسانيات المدونات أسلوبا منهجيا (Methodology)، يقوم في دراسته للغة على أساس تجريبي (Empirical basis)، مع التركيز على التحليل الكمي أو النوعي (Quantitative and Qualitative analyses) لمجموعة من النصوص المكتوبة أو المنطوقة. وتتجلى أهمية أساليب لسانيات المدونات من خلال تقديمها أدلة تدعم الافتراضات النظرية، وتساعد على تحسين فهم الظواهر اللغوية المدروسة مسبقا، وتكشف جوانب للتحقيق جديدةً لم تكن ملحوظة من قبل. وبالإضافة إلى ذلك، تتصل لسانيات المدونات بشكل وثيق بمختلف مجالات اللسانيات التطبيقية (Adamou، 2019، ص 01). كما يمكن استخدام هذه المنهجية في أي حقل من اللسانيات، لكنها لا تنتمي إلى مجال محدد من مجالاتها (McEnery and Wilson، 2001، ص 02 - Biber وآخرون، 1998، ص 3 و4). ومن وجهة النظر العلمية، يمكن القول إن أسلوب المدونات اللسانية هو الأسلوب الأقوى في دراسة اللغة، نظرا إلى إمكانية القياس والتحقق من المتوصل إليه (Leech، 1991، ص 106).

وأشار بيبر وآخرون (Biber وآخرون، 1998) إلى أن المدونات اللسانية تعتمد مقارباتٍ منهجيةً في تحليل اللغة، وتتميز بأربع خصائص رئيسية، تتمثل في أنها:

- تجريبية وواقعية (Empirical) في تحليلها للنصوص القائمة على الاستخدام الفعلي للغة؛
- تستخدم مجموعة ضخمة من النصوص أساسا لعملية التحليل؛
- تستثمر الحاسوب في التحليل؛
- تعتمد على كل من المنهج الكمي والمنهج النوعي في التحليل (ص 04).

تنهض المدونات اللسانية بأدوار جوهرية ضمن مجال اللسانيات، حيث تستخدم مجموعة من المناهج والمقاربات الخاصة بها لدراسة اللغة واستكشافها (McEnery and Hardie، 2012، ص 01). وتسهم في معرفة بنية اللغة واستخدامها وتنوعها، من خلال توفير أدلة تجريبية للظواهر اللغوية، مما يسمح للباحثين بوصف استخدام اللغة بناء على البيانات اللغوية الفعلية (Biber وآخرون، 1998، ص 25). كما تساعد الباحثين على استكشاف ودراسة التباينات في اللغة عبر ضروب مختلفة من الأنواع والسجلات والفترات الزمنية، وبذلك تقدم رؤية شاملة حول كيفية تحول اللغة وتغيرها مع مرور الزمن (Leech وآخرون، 2001). وللسانيات المدونات كذلك أدوار مهمة في مجال صناعة المعاجم إذ توفر البيانات الضرورية لإعداد مداخل معجمية تتضمن معاني الكلمات والمرادفات اللفظية وأمثلة على استخداماتها، وتقدم تحليلًا للجوانب العملية للاستخدام اللغوي وبنية الخطاب يعين الباحثين في فهم كيفية عمل اللغة ضمن سياق التواصل (Stubbs، 2001). ولها كذلك دور في تطوير النماذج الحاسوبية من خلال توفير بيانات لغوية لتدريب نماذج حاسوبية تدخل ضمن تعلم الآلة وتقييم خوارزميات المعالجة الآلية للغات الطبيعية (Manning and Schütze، 1999). وللسانيات المدونات أيضا دور مهم في مجال تعليم اللغات، إذ تقدّم أمثلة أصلية وفعلية على استخدام اللغة لأغراض التدريس والتعليم (O'Keeffe وآخرون، 2007). كما تتضافر من خلالها مجالات متعددة التخصصات، إذ تتيح للسانيين العمل مع الباحثين في علم النفس وعلم الاجتماع وغيرها من المجالات (McEnery and Gabrielatos، 2006).

هناك أنواع مختلفة من المدونات اللسانية، تختلف حسب استخدامها والهدف منها، ومنها: المدونات العامة، والمدونات المتخصصة، ومدونات المتعلمين، والمدونات التربوية، والمدونات التاريخية، والمدونات المتوازية، والمدونات المقارنة وغيرها، ويعتمد استخدام مدونة بعينها على الهدف من المدونة (Bennett، 2010، ص 13).

وتعتمد لسانيات المدونات في دراسة اللغة وتحليلها المناهج الكمية والنوعية القائمة على الأساليب الإحصائية، بالإضافة إلى مقاربات قائمة على المدونات وموجهة بالمدونات. وقد استفادت لسانيات المدونات من التطورات التكنولوجية التي شهدتها علوم الحاسوب، وخاصة ميدان المعالجة الآلية للغات الطبيعية، الذي أسهم بأدوات تخدم المدونات اللسانية، وتعنى بتحليل البيانات اللغوية، وتساعد المعجميين على استخراج المعلومات اللغوية التي يحتاجونها ودراستها وتحليلها. وكان للسانيات المدونات منذ بدايتها دور في تغيير الممارسات المعجمية، من خلال استخدام المدونات اللسانيات بوصفها أدوات قيمة، يعتمد عليها المعجميون لاستخراج المواد المعجمية والمعلومات المتعلقة بها.

3. لسانيات المدونات ودورها في الصناعة المعجمية

مع نهاية القرن العشرين، شهدت النظريات والممارسات المعجمية تغيرات جوهرية، نتيجة للتطورات التكنولوجية التي عرفها مجال تحليل البيانات اللغوية، والتي سمحت باستخدام التقنيات الحاسوبية في مجال الصناعة المعجمية، مما أدى إلى التحسين من الكفاءة والدقة بشكل كبير (Teubert، 2004، ص 16). وقد أحدثت المدونات اللسانية ثورة كبيرة ضمن الممارسة المعجمية، حيث أثارت نقاشا حيويا حول القيمة النسبية لأدلة المدونات اللسانية بالمقارنة مع الحدس اللغوي (Rundell and Stock، 2008، ص 28).

حسب جيوفري ويليامز (Williams، 2003، ص 05) فإن لسانيات المدونات أسهمت في تغيير الممارسات المعجمية، وذلك من خلال إتاحة إمكانية الوصول إلى كميات كبيرة من البيانات اللغوية التي تمثل الاستخدام الفعلي والواقعي للغة، وقد

استخدم المعجميون بيانات تجريبية منذ فترة طويلة، حتى قبل ظهور لسانيات المدونات. ومع توافر المدونات اللسانية ومجموعات أخرى من النصوص المقروءة آليا، أمكن للمعجميين أن يحصلوا من الحاسوب على أمثلة لاستخدام الكلمات والعبارات من ملايين الكلمات النصية في غضون بضع ثوان. وقد مكّن هذا ليس فقط من إنتاج المعاجم وتنقيحها بوتيرة أسرع من ذي قبل، بل أدى أيضا إلى تعريفات أحدث وأكثر شمولا، من خلال تحليل عينة أكبر من الأمثلة الطبيعية (McEnery and Wilson، 2001، ص 106 و107).

أسهمت لسانيات المدونات ولا تزال في مجال الصناعة المعجمية من خلال مناهجها في دراسة البيانات اللغوية لمدونات المعاجم وتحليلها، إذ تشكّل موردا تستقي منه المعاجم المداخل المعجمية والمعلومات المتعلقة بها. وقد اعتبرت الصناعة المعجمية المجال الرئيسي الثاني الذي أسهمت فيه لسانيات المدونات بتوفيرها أساليب جديدة وسعت نطاق البحث بالكامل (Teubert، 2001، ص 127). وقد أشار سنكلير (Sinclair، 2004) إلى أن المدونات اللسانية استخدمت لتطوير معاجم متخصصة، مثل المعاجم القانونية والطبية والعلمية، وهي معاجم تساعد المهنيين والباحثين بشكل كبير، كل في مجاله. كما أكد أن بناء مشروع معجم أساسه المدونات اللسانية خطوة إيجابية، لكنه يتطلب تقييمًا دقيقًا من المعجميين لدرجة الموثوقية ضمانا لدقة المعجم.

1.3. مقاربات لسانيات المدونات

تعد المعاجم والمدونات اللسانية نوعين من الموارد اللغوية، حيث يمثل كل واحد منهما مرحلة مختلفة من البناء، يستخدم المعجميون الأدلة والمعلومات الموجودة في المدونة اللسانية ثم يعالجونها من خلال التحليل والتصفية والفرز والاختيار؛ لإنشاء المداخل المعجمية والمعلومات المتعلقة بها (Sinclair، 2004). وتشمل منهجية لسانيات المدونات أسلوبا نسقيا لإنشاء المدونات اللسانية وتحليلها وتفسيرها، توفر من خلاله للسانيين والباحثين أدوات قيّمة لاستكشاف اللغة من منظور قائم على البيانات والأدلة. وتعدّ المقاربة القائمة على المدونات (Corpus-Based Approach) من بين هذه المناهج التي توظف فيها المدونات اللسانية أساسا تجريبيا للمعجميين لاستخراج البيانات واكتشاف الظواهر اللغوية دون افتراضات أو توقعات (Nesi، 2000، ص 839). أما في المقاربة الموجهة بالمدونة اللسانية (Corpus-Driven Approach)، فتستخدم المدونة اللسانية بوصفها قائمة أو مستودعًا لجرد البيانات اللغوية، لاستخراج المواد المناسبة لدعم المعرفة البديهية، والتحقق من التوقعات، والسماح بتحديد كمية الظواهر اللغوية، وإيجاد دليل على النظريات الحالية أو لاسترجاع عينات توضيحية. إنها طريقة يتم فيها استجواب المدونة، واستخدام بياناتها لتأكيد التفسيرات والافتراضات اللغوية المحدَّدة مسبقًا (Tognini-Bonelli، 2001، ص 81-84).

2.3. أدوات المعالجة الآلية للغات الطبيعية ولسانيات المدونات

اتخذت أغلب المعاجم الحديثة في القرن الواحد والعشرين من بيانات المدونات اللسانية نقطة انطلاق لها، حيث أصبح المعجمي المعاصر يستعمل عددا من التقنيات الحديثة المتطورة التي تشتمل على:

- أجهزة حاسوب شخصية، ذات سعة تخزينية هائلة، ومعالجات قوية، ووصلات إنترنت سريعة؛
- بيانات المدونات اللسانية، التي تتم معالجتها باستخدام أدوات برمجية طورت عن طريق تقنيات المعالجة الآلية للغة الطبيعية، والوصول إليها من خلال برامج استعلام مخصصة؛

- برنامج لإدخال نصوص المعجم في قواعد البيانات التي تخزن المعلومات اللغوية وتديرها (Atkins and Rundell، 2008، ص 03).

أشار ميشيل راندل وآدم كيلغاريف (Rundell and Kilgarriff، 2011) إلى أن هناك مجموعة متنوعة من الأدوات البرمجية تحلل المدونات، وتساعد في عدة جوانب من بينها: تحديد البنيات الصرفية والتركيبية والدلالية للوحدات اللغوية الموجودة في النص، وتحديد الكلمات وسياقاتها، واختيار قائمة من الأمثلة والشواهد، وتسهيل عملية تحرير المعجم وإنجازها وغيرها من الوظائف. وتعمل المقاربة الحالية في تطوير المعاجم على أن توفر للمعجميين برامج تحليل البيانات المستمدة من المدونات بشكل مفيد ومسبق التحضير. ويضيف ميشيل راندل وآدم كيلغاريف أن هناك تطورات حديثة في مجال لسانيات المدونات، تتجلى في إمكانية إنشاء نموذج للمعجم باستخدام برامج المدونات، ويتم ذلك عبْر تحديد بيانات ذات صلة، ويقوم البرنامج فعليا بملء الحقول المناسبة في قاعدة بيانات المعجم، ويبقى للمعجميين فرصة للتحقق من البيانات، أو إمكانية تحسينها أو رفض القرارات التي اتخذها البرنامج (ص 278).

وأكد توبرت (Teubert، 2001) أنه مع توفر مدونات كبيرة وبرامج حاسوبية متطورة لاستكشافها، أصبحت لسانيات المدونات جزءا من الصناعة المعجمية. وأضاف أن الجيل القادم من المعاجم، سواء أحادية اللغة أو ثنائية اللغة، يحتاج إلى التحقق من المدونة، إن لم تكن قائمة بالكامل عليها (ص 151). وهذا ما يؤكده كينيدي (Kennedy، 2014)، حينما أقر بأهمية وجود مدونة مخزنة في الحاسوب نظرا إلى دورها في تسهيل البحث على الموارد والمعلومات اللغوية وفرزها وإحصائها، إما أساسا للوصف اللغوي أو لمعالجة القضايا والإشكاليات المتعلقة باللغة. وتعددت أبحاث لسانيات المدونات لتسهم بمنهجيتها في تحليل البيانات اللغوية ودراستها في صناعة المعاجم، والقواعد الوصفية، والدراسات المقارنة التعاقبية والتزامنية لأنواع الكلام، وكذا في التطبيقات الأسلوبية والتربوية وغيرها من التطبيقات (ص 11).

1.2.3. معلومات المدونة اللسانية

تحتوي المدونات اللسانية على أنواع ثلاثة من المعلومات، تساعد في تحليل البيانات اللغوية الموجودة في المدونة ودراستها، وتتجلى هذه المعلومات في البيانات الوصفية (Metadata)، والترميز النصي (Textual markup)، والتحشية اللغوية (Linguistic annotation). كما توفر البيانات الوصفية معلومات حول النص، مثل الكاتب، وتاريخ النشر، والمكان، واللغة، والموضوع، وغيرها من التفاصيل الأخرى. ويتضمن الترميز النصي، معلومات تم ترميزها في النص بشكل غير مباشر للكلمات الفعلية. إذ يمكن للترميز النصي، على سبيل المثال، أن يخبرنا عن مكان بداية الكتابة ونهايتها بخط الحروف المائلة في النص المكتوب، ويشير إلى متى يبدأ وينتهي المتحدثون في التعبير عن أنفسهم في مدونات الكلام المنطوق (McEnery and Hardie، 2012، ص 29). وتشمل التحشية اللغوية إضافة المعلومات اللغوية إلى المدونة، من قبيل التوسيم بأقسام الكلام (POS tagging)، وذلك بإضافة مجموعة من الرموز أو العلامات التي تحدد بالنسبة لكل وحدة معجمية في النص نوع الكلمة (فعل، اسم، صفة...إلخ)، وبذلك يتحول النص إلى مستودع للمعلومات اللغوية. وتقوم التحشية اللغوية بتحديد المعلومات الواردة في النص العادي على نحو يمكّن من الوصول إليها واستخدامها بطريقة أسهل (McEnery and Wilson، 2001، ص 30).

تستخدم البيانات الوصفية (Metadata) أساسا لتصنيف المدونة وتنظيمها، لفهم سياق النص الزماني والمكاني، وكذلك تحديد هوية الكاتب أو المتحدث. ويعدّ الترميز النصي (Textual markup) من العناصر المهمة في تحليل اللغة الطبيعية، إذ

يتيح هذا النوع من المعلومات تحديد هيكل النص وتنسيقه، ويشمل الترميز النصي الوسوم اللغوية مثل HTML أو XML التي تحدد الفقرات والعناوين والروابط وغيرها من العناصر التي تشكل الهيكل الأساسي للنص. وتشير التحشية اللغوية (Linguistic annotation) إلى إضافة وسوم أو علامات لغوية (Tags) إلى مكونات النص، لتحديد هيكله اللغوي والمعاني اللغوية المختلفة، وتمثل التحشية مختلف مستويات اللغة الصوتية منها، والصواتية، والصرفية، والتركيبية، والدلالية، والتداولية، وغيرها من المستويات مما يساعد على تحديد نوع التوسيم اللغوي الذي من خلاله يتحدد الهدف من المدونة.

2.2.3. توسيم المدونات اللسانية

تعد المدونة اللسانية أداة فعالة في البحث اللساني، وخاصة إذا توفرت موسومة، فإنها ستشكل عندئذ مادة غنية بالمعلومات اللغوية المتنوعة، يستفيد منها الباحثون والمختصون على اختلاف تخصصاتهم وتوجهاتهم. وقد أشار ماكنري وآخرون (McEnery وآخرون، 2019) إلى أنه منذ عام 1998، شهد مجال توسيم المدونات تقدما كبيرا نتيجة لتطور التقنيات الحاسوبية التي أدت إلى توسيع نطاق استخدام البيانات اللغوية، وأصبحت الأدوات التي يمكن للغويين استخدامها لتوسيم المدونات مدمجة بشكل جيد في حزم برامج مستقلة عن المدونات اللسانية (ص 82). يتجلى توسيم المدونات من خلال إضافة المعلومات اللغوية إلى الوحدات اللغوية المتضمنة في مدونة النصوص المكتوبة أو الكلام المنطوق، وتتمثل المعلومات اللغوية في: المعلومات المعجمية، والصرفية، والنحوية، والدلالية، والتداولية، والخطابية، وغيرها، ويمكن إضافة المعلومات الصوتية والفونولوجية في حالة مدونة الكلام المنطوق (Lu، 2014، ص 05).

يعدّ توسيم المدونات اللسانية بالبيانات اللغوية الإضافية هو الأكثر استخداما، وذلك لعدة أسباب، منها:

- أنه يسهم في إثراء النصوص وتعزيز القيمة اللغوية من خلال إضافة بيانات داخل النص.
- أنه يوفر قوائم جاهزة للعديد من الوسوم اللغوية، مما يسهل استخدامها في التوسيم.
- أنه يتسم بسهولة الإضافة خاصة مع وجود برامج حاسوبية خاصة لبعض أساليب التوسيم، مما يسهل عملية إدراجها بالطريقة المناسبة (الفيفي، 2023، ص 287 - 288).

تهدف الوسوم إلى تبسيط المعلومات التي تضاف إلى المدونات اللسانية، وذلك عبر إنشاء جدول يصنف هذه الوسوم ويوضح دلالتها، وقد يشمل شرحًا لكيفية استخدامها في بعض الحالات، كالحالات التي فيها لبس أو تداخل بين الوسوم أو عند استثناء بعض الحالات من تطبيق وسم محدد (الفيفي، 2023، ص 287).

وتلتقي لسانيات المدونات مع اللسانيات الحاسوبية (Computational linguistics) في بعض المجالات، بحيث تسهم المعالجة الآلية للغات الطبيعية في البحث في المادة اللغوية وتحليلها، واستخراج مجموعة من الظواهر اللغوية (القاسمي،2019، ص 708). وتقوم بعدة مهام ووظائف مختلفة مثل تجزيء النص (Tokenization)، والتضمين (Embedding)، وتوسيم أقسام الكلام (POS tagging)، والتعرف على الكيانات المسماة (NER)، وتصنيف النصوص (Text Classification)، ونمذجة اللغة (Language Modelling) وغيرها من الوظائف المهمة. ويحتاج الباحثون والمطورون في هذا المجال إلى أدوات فعالة لتنفيذ هذه المهام، ونتيجة للتطورات الحديثة في المعالجة الآلية للغات الطبيعية، أصبح هنالك الآن العديد من الأدوات والأساليب لتجزيء النصوص، وتضمين الكلمات، وتوسيم أقسام الكلام وغير ذلك. ومن

الأدوات الشائعة التي تتوفر على أساليب مختلفة لأداء هذه الوظائف نجد NLTK، وcoreNLP، وspaCy، وAl-lenNLP، وFlair، وStanza وغيرها من الأدوات المتنوعة (Sarker، 2021، ص 01).

يعدّ توسيم أقسام الكلام (POS tagging) من أكثر الأنواع شيوعًا واستخداما، نظرا إلى ارتباطه الوثيق بالعديد من دراسات لسانيات المدونات، وتقديم الدعم لعمليات التوسيم الأخرى، مثل التأصيل (Lemmatization)، والتشجير التركيبي (Syntactic parsing)، والتحشية الدلالية (Semantic annotation) وما إلى ذلك. ويتضمن هذا النوع من التوسيم تخصيص رمز أو علامة لكل هيكل كلمة (Token)، مما يسهم في تحديد الحد الأدنى لقسم الكلام في الكلمة (Gries and Berez، 2017، ص383). ويعدّ تحديد أقسام الكلام (Parts of Speech) في الجملة خطوة حاسمة في مهام المعالجة الآلية للغات الطبيعية (NLP)، حيث تتجلى في إضافة معلومات قيّمة عن الكلمات المستخدمة في الجملة، بما في ذلك النحو، والدلالات، والترجمة، وغيرها. ويرتبط توسيم أقسام الكلام (POS tagging) ارتباطًا وثيقًا بلسانيات المدونات، ويتمثل في تحديد قسم الكلام الذي تنتمي إليه كل كلمة في الجملة. ويتم ذلك بواسطة موسّم قسم الكلام (POS tagger) وهو برنامج حاسوبي خاص يقرأ كلمات اللغة ويحدد قسم الكلام بالنسبة إلى كل كلمة (Kanakaraddi and Nandyal، 2018، ص 01).

لذلك، لا يقتصر نشاط تجميع المدونات على جمع النصوص اللغوية المناسبة من مجموعة واسعة من البيانات اللغوية لمجالات مختلفة، لكن من المهم ربطها وتزويدها بالبيانات الوصفية، وإثراء بياناتها بواسطة أدوات تعمل بتقنيات المعالجة الآلية للغات الطبيعية، وتضيف إلى الوحدات اللغوية المضمنة وسوما لغوية أو "علامات"، وتعتمد أنواع الوسوم على نوع اللغة وعلى النظرية المطبقة. وتضم الممارسة العملية وسوما من أقسام الكلام وتصنيفاته الفرعية، والمورفولوجية، بما في ذلك علامات الترقيم. وهناك أيضا أنواع أخرى من الوسوم، كالوسوم النحوية والدلالية والتداولية وغيرها.

3.2.3. تحليل المدونات اللسانية

يحتاج المعجميون لتحليل بيانات المدونة إلى برنامج حاسوبي يمكّنهم من البحث عن البيانات ومعالجتها وحفظها، ويسمى هذا البرنامج "أداة المدونة"، وتعد هذه الأداة مفتاحا أساسيا للتحليل المعجمي، فالمدونة بدون برامج حاسوبية جيدة لن يستفاد منها بالقدر المطلوب (Kilgarriff and Kosem، 2012، ص 31). وتتوفر هذه البرامج على أدوات متنوعة تعمل بتقنيات المعالجة الآلية للغات الطبيعية، وتؤدي وظائف محددة حسب نوعية التحليل والأداة المستخدمة.

تعد المدونات اللسانية أساسا تجريبيا لدراسة بعض الظواهر اللغوية والإجابة عن مجموعة من الأسئلة البحثية، ويكون حجم البيانات اللغوية التي تتضمنها المدونات اللسانية كبيرا جدا، مما يجعل من الصعب تحليلها يدويا أو بالعين المجردة في إطار زمني معقول. لذلك، تستخدم أدوات حاسوبية متطورة تسمح للمستخدمين بالبحث في المدونات اللسانية بسرعة وبشكل موثوق. ومن هذه الأدوات: المكشافات السياقية (Concordancer) التي تسمح للمستخدمين بالنظر إلى الكلمات في سياقها، وقوائم التكرار أو التردد (Frequency List) التي تعنى بإنتاج بيانات حول درجة تكرار الوحدات اللغوية أو ترددها، وتسرد جميع الكلمات الواردة في المدونة اللسانية، وتحدد لكل كلمة عدد مرات تكرارها. وتمثل البيانات التي توفرها أدوات الكشافات السياقية وأدوات التردد والتكرار على التوالي شكلين مختلفين من التحليل النوعي والكمي، اللذين يعدّان من المناهج المعتمدة في لسانيات المدونات (McEnery and Hardie، 2012، ص 1 و2).

لقد غيرت أجهزة الحاسوب بفضل تخزين النصوص في شكلها الإلكتروني -قابلة للقراءة آليا، أداء المهام المتضمنة في تحليل السياقات، وكذا تحديد مواقع تكرار كلمة بعينها مع سرد سياقاتها بسرعة كبيرة وبشكل موثوق تمامًا. وقد حدث هذا التغيير بفضل برامج حاسوبية متطورة تسمى "مولدات الكشافات السياقية" (Concordance Generators)، أو "برامج الكشاف السياقي" (Concordancing Software)، أو بشكل عام "المكشافات السياقية" (Concordancer) وهي تعمل خاصة لتحقيق هذا الهدف (Rundell and Stock، 2008، ص 29).

تعتمد لسانيات المدونات على أدوات متنوعة، تقوم بمجموعة من العمليات التحليلية والإحصائية يطبقها برنامج حاسوبي على المدونة اللسانية قيد الدراسة، سواء كان هذا البرنامج الحاسوبي مستقلا أو مدمجا على المدونة مباشرة، ومن أهم هذه الأدوات:

- قائمة الكلمات (WordList): وتعرض كلمات المدونة اللسانية مع نسبة تردد كل كلمة وتكرارها في المدونة (الشهري، 2022، صفحة 80)؛

- التصاحب اللفظي أو التلازم اللفظي (Collocation): ويعرّف بالكلمات المتلازمة أو المتصاحبة (Collocate) التي تظهر مع الكلمة المحورية قيد البحث على امتداد تتابعي من كلمة إلى خمس عشرة كلمة إلى يمين الكلمة المحورية ويسارها، وتسمح هاته الأداة بمعرفة معنى الكلمة المحورية واستعمالها من خلال متلازماتها اللفظية (الشهري، 2022، صفحة 81)؛

- الكشاف السياقي (Concordance): وهو بمثابة قائمة تضم الكلمة أو العبارة مع بعض الكلمات في السياق في كلا جانبيها، مما يمكّن من معرفة ميول استخدام الكلمة من خلال سياقها، وتجمع هاته الأداة بين التحليل الكمي والنوعي (الشهري، 2022، صفحة 81)؛

- التتابعات اللفظية (N-Grams): أي تسلسل من الكلمات، في هذا السياق، يستخدم مصطلح (Unigram) للإشارة إلى كلمة واحدة، ومصطلح (Bigram) إلى تسلسل من كلمتين، ومصطلح (Trigram) إلى تسلسل من ثلاث كلمات، وهكذا. وتستخدم هذه التتابعات اللفظية أساسا لمعرفة تردد العبارات المكونة من كلمة واحدة Unigram، أو كلمتين Bigram، أو ثلاث كلمات Trigram (الدكروري، 2018، صفحة 116).

وهناك أدوات أخرى بالإضافة إلى هذه الأدوات، تختلف حسب وظائفها أو العمليات التحليلية والإحصائية المراد القيام بها. وقد أصبحت لسانيات المدونات، حسب توبرت (Teubert، 2001)، جزءًا لا يتجزأ من ميدان صناعة المعاجم، نظرا إلى توفر مجموعة واسعة من المدونات وزيادة عدد البرمجيات المتطورة التي تهتم بدراسة الوحدات اللغوية وتعريفها من خلال فحص السياق (ص 151).

4. استثمار لسانيات المدونات في بناء المعاجم المدرسية

أشارت كرانكر (Granger، 2018) إلى أن المدونات اللسانية تعد نقطة تحول حقيقية في عالم صناعة المعاجم، وأنها لم تقدم حتى الآن كل الإمكانات الكاملة التي يمكنها تقديمها في العديد من المعاجم المتخصصة، ومنها على وجه الخصوص المعجم التعليمي أو المدرسي، من خلال بعض المقاربات، كالمقاربة القائمة على المدونة، والمقاربة الموجهة بالمدونة (ص 22 و23).

ذكر مايكل راندل (Michael Rundel) رئيس تحرير معجم ماكميلان المدرسي (Macmillan School Dictionary, 2004) في مقدمة المعجم أنهم يستخدمون برامج حاسوبية متطورة لتحليل مدونة تتألف من 20 مليون كلمة من الكلمات الواردة في الكتب المدرسية ومناهج الامتحانات في جميع المواضيع، من الزراعة إلى علم الحيوان. وقد ساعده البحث على اكتساب فهم عميق للمصطلحات والمفاهيم الأساسية اللازمة لدراسة مختلف المواد المدرسية. واستنادا إلى هذه النتائج، تم تحديد أكثر الكلمات

استخداما في الكتب المدرسية مما يتصل بعلوم النبات، والدراسات المتعلقة بالثقافات، والبيئة، وتكنولوجيا المعلومات. ويتضمن هذا المعجم المدرسي هذه الكلمات، ويوفر المعلومات اللازمة لكل كلمة على أساس معلومات موثوقة من بيانات هذه المدونة.

يتجلى الهدف الرئيسي من المعاجم المدرسية في كونه أداة تساعد على الفهم، والتعلم، والاكتساب، وتطوير الكفايات اللغوية والمعرفية للتلميذ من خلال تزويده بما يحتاج من معلومات لغوية ضمن محيطه التعليمي. وبخصوص رجوع التلميذ إلى المعاجم، يرى أحمد محمد معتوق (1996) أن التلميذ "قد يعثر أثناء قراءة الموضوعات المقررة أو أثناء القراءة الحرة على كلمات غريبة لا يتمكن من إدراك معانيها من خلال سياق الكلام الذي يقرأه، أو أنه لا تتوافر لديه الفرصة لسؤال المدرّس عن معانيها، مما قد يؤدي إلى الفهم الخاطئ." (ص 188). فمن الضروري في المعاجم المدرسية، أن يكون عرض معاني الكلمات سياقيا، أي تحديد معنى الكلمات من خلال السياق، وبالتالي يجب أن تسعى المعاجم إلى إظهار تلك السياقات من خلال الاستخدام الحقيقي، وهذا ما توفره لنا لسانيات المدونات (Williams، 2003، ص 04).

يعدّ المعجم المدرسي أداة مهمة تسهم في تنمية الثروة اللغوية وتطويرها، حيث يشجع تقديم تعاريف متناسبة مع السياق الدراسي للتلاميذ، على استكشاف اللغة في إطار دراستهم، مما يُسهِم في تحسين كفاياتهم اللغوية والمعرفية. ويرى مايكل راندل (Michael Rundel، 2004) أن المدونات توفر تفاصيل حول المفاهيم المدرجة في الكتب المدرسية المستخدمة في الفصل الدراسي، وتوضح كيفية شرح هذه المفاهيم من خلال سياقاتها الفعلية. وذلك يُسهِم في إعداد نماذج خاصة تتعلق بالتعاريف بطريقة تجعلها سهلة الفهم والمتابعة.

وقد أكد أبو العزم (2010) ضرورة الاعتماد على منهجيات معجمية لصناعة معاجم مدرسية، وذلك بالاشتغال في ضوء رصيد لغوي أساسي، أو الاشتغال في ضوء مدونة لغوية توافق قدرات التلميذ اللغوية، لاستخراج المداخل المعجمية وتشكيلها (ص 48 و49).

إن استثمار نهج لسانيات المدونات في بناء المعاجم المدرسية من خلال تطبيق المقاربة القائمة على الكتب المدرسية واستخدامها كمدونة للمعجم المدرسي، سيسهم في حصر المواد اللغوية التي يحتاجها التلاميذ، واستخراج المداخل المعجمية من خلالها، وسيعزز جودة التعريفات المبنية على السياق، ومن تقديم الأمثلة والشواهد، ومعلومات لغوية غنية أخرى. وقد أكد سنكلير (Sinclair، 2004) أهمية المعاجم التي تخدم احتياجات متعلمي اللغة، وتستخدم المدونات لتقديم تعريفات واضحة وأمثلة سياقية ومعلومات عن تواتر الكلمات واستخدامها.

من جهتخه، أكد الحاج صالح (2005) أن العمل في إطار مدونة لسانية يعدّ ضرورة أساسية لجمع المادة المعجمية، وقد أشار إلى شروط وخصائص مهمة، كالتخطيط الجيد لجمع المدونة، الذي يعدّ ركنا وشرطا أساسيا، وكذا حصر نصوصها من اللغة المستعملة في المحيط التعليمي للمتعلم، وأن تتسم بالشمول والتنوع لاستخدام الشواهد والأمثلة منها (ص 21).

أشار راندل وستوك (Rundell and Stock، 1992) إلى أن المعاجم التقليدية التي سبقت ظهور لسانيات المدونات واعتمدت مصادر أخرى تواجه تحديات بارزة. فمعجم كولينز (Collins) الإنجليزي (الصادر في طبعته الأولى عام 1979، أي قبل توفر دلائل المدونات اللسانية)، على سبيل المثال، قد اعتمد لغة الناطقين الأصليين، وقدم خمس تعاريف لتغطية المجالات المحددة، وهو مرتب بشكل تقليدي، ولا يفرق بين استخدام الصفة أو النعت أو الاسم (ص 25).

أحدثت لسانيات المدونات ثورة في عالم صناعة المعاجم، حيث تم الاستناد على النظرة السياقية للمعنى، وتطبيقها على الممارسات المعجمية من خلال معاجم COBUILD. وفتحت لسانيات المدونات آفاقا جديدة لتحليل الكلمات في السياق

لإثبات استخدامها، مما أدى إلى تغيير في شكل المعاجم، وتنسيقها، وهيكلها. وسمح هذا النهج بنقل المعرفة السياقية بشكل أفضل إلى المستخدمين، ورغم كون المشاريع المعجمية الكبرى قد اعتمدت على تكنولوجيا الكمبيوتر، فإن لسانيات المدونات أحدثت ثورة أكبر في بيانات المصدر، إذ فتحت أبواب الوصول إلى كميات ضخمة من البيانات الواقعية. وكان سنكلير وفريق COBUILD من جامعة برمنغهام أول من أدرك إمكانات المدونات في صناعة المعاجم (Williams، 2003، ص 01).

إن استخدام أسلوب لسانيات المدونات في دراسة الكتب الدراسية وتحليلها من شأنه أن يسهم في تنمية القدرات اللغوية والمعرفية لدى التلاميذ، من خلال حصر السياق اللغوي التعليمي الذي يحتاجونه، والذي وجب عليهم معرفته واكتسابه. وبالتالي فإن استخدام هذا النهج في صناعة المعاجم المدرسية سيساعد المعجميين في أثناء بناء معاجمهم. وبالتالي، فإن الهدف من البحث المعجمي يجب أن يحصر في تجميع الوحدات اللغوية من متون الكتب الدراسية (البوشيخي، 1998، ص 25).

هناك حاجة ملحة إلى إعادة النظر في كيفية صناعة المعاجم المدرسية، باستثمار مناهج وطرق علمية وتقنية حديثة تركز على الصناعة المعجمية وتوجهها نحو مجال التعليم. وينبغي الاستفادة من التحديات التي واجهتها المعاجم المدرسية القليلة في العالم العربي، التي لم تتبن أسسا علمية مدروسة، خصوصا من حيث اختيار مدوناتها أو اختيار المحتوى المناسب. لذا، يجب على المعاجم المدرسية الحديثة أن تأخذ في اعتبارها النهج العلمي والتقني في عمليات اختيار المفردات وتعريفها، وذلك بالاعتماد على دراسات ميدانية تحليلية لاحتياجات المتعلمين، وتوجيه اهتمام خاص لمستوى المتعلم واحتياجاته اللغوية والمفهوم الذي يبحث عنه، واختيار المفردات والمصطلحات اعتمادا على الواقع اللغوي الذي يعيشه التلاميذ.

5. نموذج إعداد مدونة المعجم المدرسي

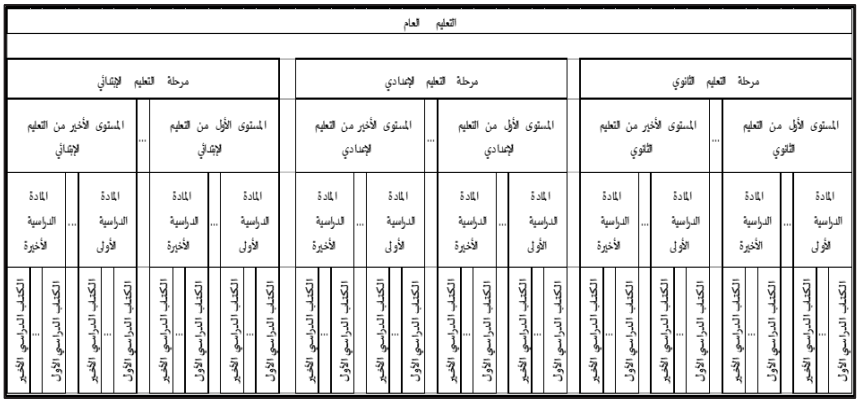
يسعى هذا البحث إلى تقديم نموذج لإعداد مدونة للمعجم المدرسي؛ وجدير بالتذكير هنا أن معايير بناء مدونة لسانية يستند إلى الاحتياجات اللغوية التي تسعى المدونة لتلبيتها (Williams، 2002، ص 46). فالحاجة ماسة لبناء مدونة تتضمن جميع الكتب الدراسية المقررة لجميع مراحل التعليم العام (قبل الجامعي)، والتي من خلالها سيستخلص المعجميون المداخل المعجمية، والمعلومات المتعلقة بها، لبناء معاجم مدرسية موجهة لتلاميذ التعليم العام. تضم هذه المعاجم المدرسية الكلمات والمفردات الواردة في الكتب الدراسية المقررة، التي يحتاجها ويستخدمها التلاميذ في الفصول الدراسية، والتي من خلالها تساهم في تنمية رصيدهم اللغوي والمعرفي وتطويره، وتساهم في إكسابهم مجموعة من المهارات الضرورية، وتجاوز جملة من الإشكاليات التي يواجهونها خلال مراحل التعليم المختلفة.

أشار أحمد مختار عمر في كتابه صناعة المعجم الحديث (2009)، أن هناك طريقتين لجمع المادة الأساسية الملائمة لمعاجم المتعلمين. وتتمثل الطريقة الأولى في حصر المادة المعجمية انطلاقا من الكتب الدراسية؛ وتتمثل الطريقة الثانية في جمع المادة من الكتب الدراسية، ومن مواد أخرى تساعد على تنمية حصيلة التلميذ في أثناء تدرجه في المراحل العمرية (ص 45 و46). وبالتالي فإن حصر المادة المعجمية وجمعها من خلال الكتب الدراسية هو الهدف الرئيسي من إنشاء هذه المدونة.

يظهر الجدول رقم 1 مراحل التعليم العام ومستوياته، والمواد الدراسية والكتب الدراسية المنتمية لكل مستوى، والتي ستشكل جميعها مدونة المعجم المدرسي.

جدول رقم 01

بنية مراحل التعليم العام



يشير كل من أتكينز وراندل (Atkins and Rundell، 2008) إلى أن هناك ثلاثة جوانب رئيسية لإنشاء مدونة لسانية، وهي:

- التصميم: اختيار النصوص التي ستشكل المدونة اللسانية.
- جمع البيانات: الحصول على هذه النصوص.
- الترميز: تحويل النصوص إلى تنسيق مشترك، وتجهيزها للاستخدام في نظام استعلام المدونة (Corpus Querying System) أو نظام استعلام النصوص (Text Query System) (ص 53).

تعد عملية حصر الكتب الدراسية وجمعها من المراحل الأساسية لإنشاء مدونة المعجم المدرسي، حيث إن الكتب الدراسية تعد بدورها مدونات، إذا توفرت مسبقا في صيغة إلكترونية - مقروءة آليا على الحاسوب، أما إذا تعذر توفرها بهذه الصيغة، فيُمكن تحريرها في الحاسوب باستخدام برامج متخصصة مثل Microsoft Word، وLibreOffice Writer، وNotepad++، وSublime Text، وGoogle Docs، وEmacs، وVim وغيرها من البرامج. وتجدر الإشارة إلى وجود طرق أخرى تستخدم لتخزين نصوص الكتب الدراسية إلكترونيا بدون تحريرها يدويا، وهي طرق تعتمد تقنيات المعالجة الآلية للغات الطبيعية، ومنها استخدام الماسح الضوئي (Optical Character Reader) أو كما يرمز له (OCR)، حيث يقوم البرنامج بمسح ضوئي للنص المطبوع، ثم تحويله إلى نص رقمي بصيغة بسيطة (Plain text)، أو استعمال نظام تحويل النص الشفوي إلى نص مكتوب (Speech to text)، بالإضافة إلى طرق أخرى مثل تحويل الكتب الدراسية بصيغة PDF إلى نصوص قابلة للتعديل عليها بواسطة برامج تحرير ملفات PDF مثل: (PDF converter).

بعد توفير جميع الكتب الدراسية في صيغة إلكترونية، تُنقح نصوصها وتنقى من الزوائد، وتُزال الشوائب التي من شأنها أن تؤثر في عمليات التحليل والإحصاء. وفي مرحلة موالية، يتم تحويل تمديد الملفات النصية إلى (txt.) وبترميز Unicode UTF8 أو Unicode UTF16. بعد ذلك، تضاف أنواع ثلاثة من المعلومات، هي المعلومات الوصفية (Metadata) لتصنيف الملفات النصية وتنظيمها، وتحديد السياق الزمني والمكاني للنص، كتحديد هوية الكاتب، وعنوان الكتاب الدراسي، وتاريخ الطبع والنشر وغيرها من المعلومات. ثم تأتي مرحلة الترميز النصي (Textual markup) الذي يتيح تحديد هيكل النص وتنسيقه. ويتم الترميز النصي بواسطة لغات الترميز، التي تقوم بوصف البيانات هيكليا، مما يجعلها قابلة للفهم سواء من البشر أو من البرامج. وتتكون لغات الترميز من علامات تحدد عناصر البيانات وخصائصها. ومن لغات الترميز نجد HTML، XML، JSON، SVG، MathML وغيرها من لغات الترميز التي تحدد الفقرات، والعناوين، والروابط، وغيرها من العناصر التي تشكل الهيكل الأساسي للنص. أما المرحلة الأخيرة، فتتمثل في التحشية اللغوية (Linguistic annotation) والتي تعنى بإضافة وسوم أو علامات لغوية إلى مكونات النص، لتحديد هيكله اللغوي والمعاني اللغوية المختلفة.

يتم توسيم مكونات النصوص بالمعلومات اللغوية باستعمال إحدى الأدوات الحاسوبية القائمة على تقنيات المعالجة الآلية للغات الطبيعية، لتصبح الملفات النصية لجميع الكتب الدراسية المقررة موسومة، ويوضح الشكل رقم 01 باختصار هذه المراحل.

شكل رقم 01

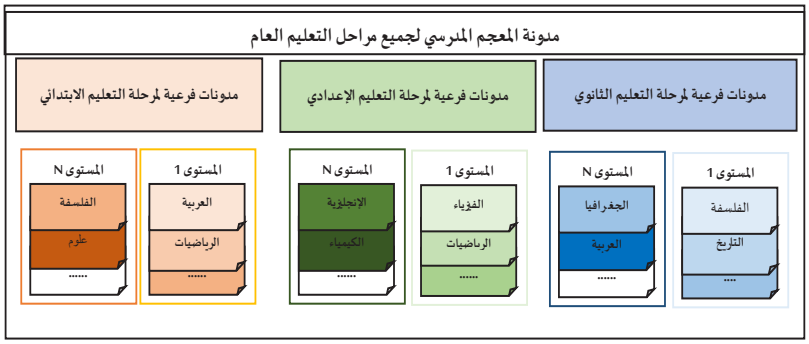
مراحل إنشاء مدونة المعجم المدرسي من الكتب الدراسية

عند الانتهاء من مرحلة بناء مدونة المعجم المدرسي، يجب تنظيمها وترتيبها من خلال وضع الكتب الدراسية ضمن الملفات حسب المراحل، والمستويات والمواد، كما هو مبين في الشكلين رقم 02 و03. يتجلى الهدف من هذه العملية التنظيمية في مساعدة

المعجميين في أثناء بناء المعاجم المدرسية على تحديد المرحلة الدراسية، والمستوى، والمادة الدراسية، كما تسهم هذه العملية التنظيمية في دراسة كل مدونة فرعية على حدة وتحليلها، وذلك لاستخراج المواد المعجمية التي تخص كل مستوى دراسي بعينه.

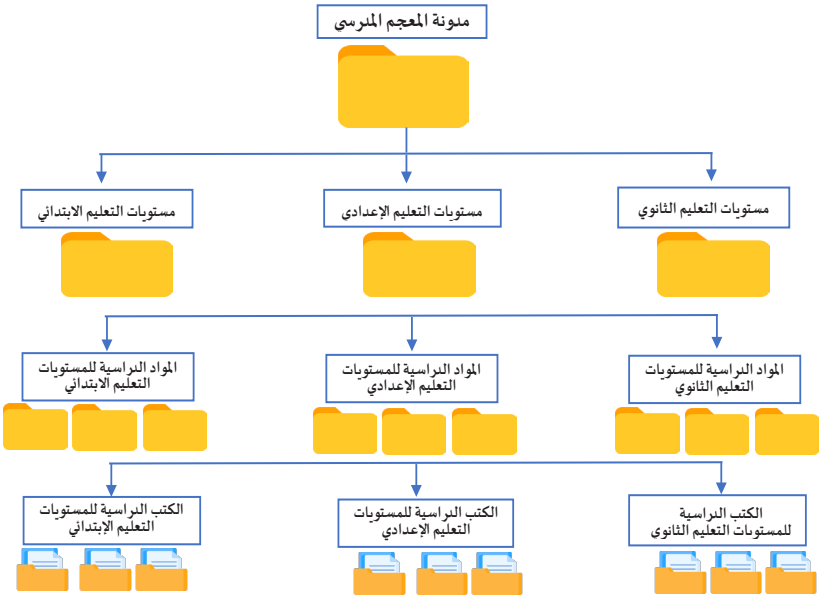
شكل رقم 02

تنظيم مدونة المعجم المدرسي ومدوناتها الفرعية



شكل رقم 03

نظرة عامة حول مكونات مدونة المعجم المدرسي



بعد ذلك، تأتي مرحلة ترتيب مدونة المعجم المدرسي وتنظيمها، وكذا مدوناتها الفرعية وترتيبها حسب المراحل، والمستويات، والمواد، ثم الكتب الدراسية، كما يبين الشكل رقم 03، وذلك لتسهيل عمل المعجمي لاستخراج المداخل المعجمية والمعلومات المتعلقة بها.

يحتاج المعجميون إلى أدوات حاسوبية تقوم بالتوسيم الآلي لأقسام الكلام (Parts of speech)، أو كما هو معروف بتوسيم أقسام الكلام (POS tagging)، أو بتوسيم الأصل (Lemma)، وذلك من خلال عملية التأصيل المعجمي (Lemmatisation). ويعدّ توسيم أقسام الكلام أشهر هذه الأنواع وأكثرها استعمالا (الدكروري، 2018، ص 66 و67). ولتوسيم أقسام الكلام، هناك مجموعة قليلة من الأدوات البرمجية التي تقوم بتوسيم النصوص المكتوبة باللغة العربية آليا، والتي يظهرها الجدول رقم 02، أما بالنسبة إلى اللغتين الإنجليزية والفرنسية، فهناك العديد من الأدوات والبرامج التي تحقق هذا الهدف، وتسمح هذه الأدوات البرمجية، بتحديد العناصر اللغوية في النصوص وتسميتها (مثل: الجمل والكلمات) وتوسيمها بمعلومات لغوية، مثل أقسام الكلام (POS tagging) (اسم، فعل، صفة...إلخ).

جدول رقم 02

الأدوات البرمجية التي تستخدم لتوسيم النصوص باللغة العربية ولغات أخرى.

الأداة	وظيفتها
Stanford Arabic NLP	توفر مكتبة Stanford NLP نماذج لتوسيم النصوص وتحليل اللغة العربية، بما في ذلك تحديد أقسام الكلام (POS Tagging) بواسطة الأداة Stanford Arabic POS Tagger.
CAMeL Tools	مجموعة أدوات برمجة بلغة البايثون توفر مجموعة متنوعة من أدوات معالجة اللغة الطبيعية للغة العربية، بما في ذلك التجزيء، والتجديع، والتأصيل، وتوسيم بأقسام الكلام.
Farasa	تقدم المكتبة البرمجية المفتوحة المصدر مجموعة واسعة من خدمات معالجة اللغة العربية، بما في ذلك توسيم النصوص وتحديد أقسام الكلام (POS Tagging).
Madamira	يقدم تحليلًا للغة العربية، بما في ذلك توسيم النصوص وتحديد أنواع الكلمات.
Kalimat NLP	مشروع مفتوح المصدر يستخدم لتحليل اللغة العربية، ويمكن أن يشمل توسيم النصوص.
SARF	أداة لمعالجة اللغة العربية، وتوفير خدمات توسيم النصوص، وتحديد أنواع الكلمات.
Arabic Treebank Tokenizer	توفر هذه الأداة مجموعة من الأدوات لتوسيم النصوص باللغة العربية، وهو جزء من مشروع Penn Arabic Treebank.
ANAA	مكتبة برمجية مفتوحة المصدر لتحليل النصوص باللغة العربية، تدعم توسيم الكلمات وتحديد أنواعها.

بعد توسيم نصوص مدونة المعجم المدرسي نختار برنامجًا مناسبًا لدراسة المدونة وتحليلها. ونذكر أن هناك مجموعة من البرامج والأدوات لمعالجة المدونات وتحليلها، وتدعم تلك البرامج والأدوات اللغة العربية بالإضافة إلى اللغات الأخرى، مثل: Sketch Engine، وaConCorde، وAntConc، وWordSmith، وGhawwas، وIntelliText Corpus Queries (Alfaifi and Atwell، 2016، ص 348). وتعدّ هذه الأدوات برامج مستقلة مصممة للبحث في مجموعات النصوص، باستثناء الأداتين Sketch Engine وIntelliText اللتين تعتمدان على استعلامات عبر الويب. وتشمل المعايير المعتمدة لتقييم هذه الأدوات قراءة الملفات النصية العربية بتنسيقات UTF8 Unicode أو UTF16 وعرض التشكيل، وضبط دقة عرض الحروف العربية من اليمين إلى اليسار، وتوحيد التشكيل أو الهمزة، وتوفير واجهة مستخدم باللغة العربية، وتمكين المستخدمين من فتح أو تحميل مجموعاتهم الشخصية باللغة العربية (Almujaiwel and Al-Thubaity، 2016، ص 01 و02). ويعرض الجدول رقم 03، هذه البرامج والأدوات والتي تقدم عدة وظائف، وتوفر مجموعة من الأساليب المتنوعة لتحليل ودراسة المدونات بشكل فعال.

جدول رقم 03

البرامج والأدوات لمعالجة وتحليل المدونات، التي تدعم اللغة العربية بالإضافة إلى اللغات الأخرى

البرنامج / الأداة	الوصف	الوظائف المتوفرة	المنصات	مجاني/غير مجاني
AConCorde	أداة تكشيف السياق متعدد اللغات (الإنجليزية والعربية)	الكشاف السياقي، قائمة الكلمات.	Linux, Mac, Windows	مجاني
AntConc	مجموعة أدوات تحليل المدونة	قائمة الكلمات، الكشاف السياقي، الكلمات المفتاحية.	Linux, Mac, Windows	مجاني
Sketch Engine	برنامج لإدارة المدونات وتحليل النصوص لعديد من اللغات منها اللغة العربية.	الكشاف السياقي، التوسيم، إنجاز عينات، البحث، العرض المرئي، قائمة الكلمات، الكلمات المفتاحية، الجمع، تحليل النصوص، التتابع اللفظي، التصاحب اللفظي، الإحصاء، التقطيع والتجزيء، تحليل تردد الكلمات، استخراج المعلومات، إنشاء المدونات، معالجة النصوص، استخراج الكلمات المفتاحية، الكشف عن الكلمات المفتاحية في السياق وغيرها من الأدوات.	Web	غير مجاني، يتيح النسخة التجريبية المجانية لمدة 30 يومًا
Wordsmith	واحد من مجموعات أدوات المدونات الأكثر رسوخًا والتي توفر مجموعة متنوعة من الوظائف كتحليل النصوص بلغات عديدة منها اللغة العربية.	الكشاف السياقي، قائمة الكلمات، إحصائيات، الكلمات المفتاحية.	Windows	غير مجاني
Ghawwas	نظام لمعالجة وتحليل المدونات باللغات العربية والإنجليزية والفرنسية.	تكرار الكلمات في المدونة، الكشاف السياقي، التلازم اللفظي والكلمات المفتاحية.	Linux, Mac, Windows	مجاني
IntelliText Corpus Queries	هو نظام قائم على الويب طُوِّرَ من قبل مركز دراسات الترجمة (CTS) في جامعة ليدز بغرض تسهيل التدريس وتعزيزه والبحث في مختلف مجالات العلوم الإنسانية.	الكشاف السياقي، التلازم اللفظي، اللواحق، مقارنة التكرارات، الكلمات المفتاحية والعبارات.	Web	--

بعد اختيار وتحديد أحد البرامج المذكورة في الجدول رقم 03، يتم تحميل مدونة المعجم المدرسي في البرنامج المحدد لتحليلها ودراستها من طرف المعجميين، لاستخراج المواد المعجمية والمعلومات المتعلق بها، باستعمال أدوات البرنامج. ويسهم استخدام

المدونة الموسومة بأقسام الكلام، في استخراج المداخل المعجمية، والتعريف بمعاني الكلمات من خلال سياقتها الفعلية، واستخراج الأمثلة والشواهد، واستخراج المرادفات والأضداد، ومعرفة المصاحبات اللفظية للكلمات، والتتابع اللفظي، واستخراج الكلمات المفتاحية، واستخراج قوائم الكلمات مع نسب ترددها وتكرارها، بالإضافة إلى معلومات كثيرة أخرى تزود بها المدونات من خلال برامج حاسوبية تتوفر على الكثير من الأدوات التي تتميز بوظائف متنوعة ومختلفة.

ومن البرامج الشهيرة والفعالة التي تخدم المدونات، نجد برنامج سكيتش إنجين[1] (Sketch Engine) الذي يدعم مجموعة من اللغات من بينها اللغة العربية، إذ يتضمن مجموعة من الأدوات القائمة على تقنيات المعالجة الآلية للغات الطبيعية، والتي تساعد المعجميين في بناء معاجمهم، وهناك العديد من المعاجم الشهيرة التي استخدمت سكيتش إنجين في إنجاز مشاريعها (الدكروري، 2018، ص 44)، ومن بينها:

- معجم روبير (Le Robert) للغة الفرنسية (2015) ؛
- معجم كولينز (Collins) متعدد اللغات (2015) ؛
- معجم ماكميلان (Macmillan) للغة الإنجليزية (2015) ؛

بالإضافة إلى مجموعة من دور النشر التي تعتمد في إنجاز معاجمها على برنامج سكيتش إنجين، من بينها:

- دار نشر جامعة أوكسفورد (Oxford University Press)؛
- دار نشر جامعة كمبريدج (Cambridge University Press).

تتوفر أدوات سكيتش إنجين على عدة مزايا من بينها:

- استخراج قوائم الكلمات مع نسب ترددها وتكراراها داخل المدونة؛
- معرفة معاني الكلمات من خلال كشف سياقاتها الفعلية والحقيقية؛
- استخراج المرادفات والأضداد؛
- استخراج المصاحبات اللفظية؛
- استخراج المتتابعات اللفظية؛
- استخراج الأمثلة والشواهد؛
- استخراج الكلمات المفتاحية في السياق؛
- استخراج الكلمات حسب قسم الكلام؛
- البحث بواسطة الجذر، الأصل؛
- البحث بواسطة أقسام الكلام؛
- استخراج قوائم أقسام الكلام؛
- استخراج العبارات الشائعة مع نسب ترددها وتكراراتها داخل المدونة؛
- استخراج المداخل المعجمية؛
- عرض البحث من خلال سحابة الكلمات،

[1] https://www.sketchengine.eu/

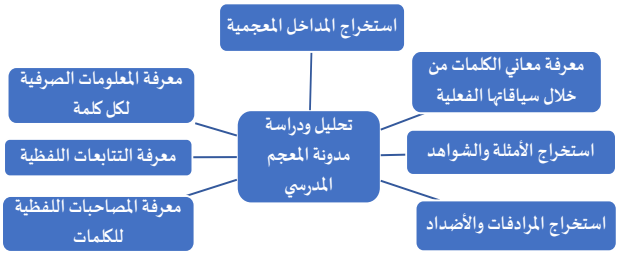
- يتوفر على نظام كتابة القاموس(Dictionary Writing System).

ومساهمات أخرى تختلف حسب نوعية الأدوات البرمجية ووظائفها.

ويلخص الشكل رقم 04 مجموعة من المعلومات التي تقدمها المدونات للمعجميين في أثناء صناعة المعاجم، في حين يلخص الشكل رقم 05 المراحل المتبعة بعد جمع الكتب الدراسية في صيغة إلكترونية، واستخدامها كمدونات فرعية لمدونة المعجم المدرسي.

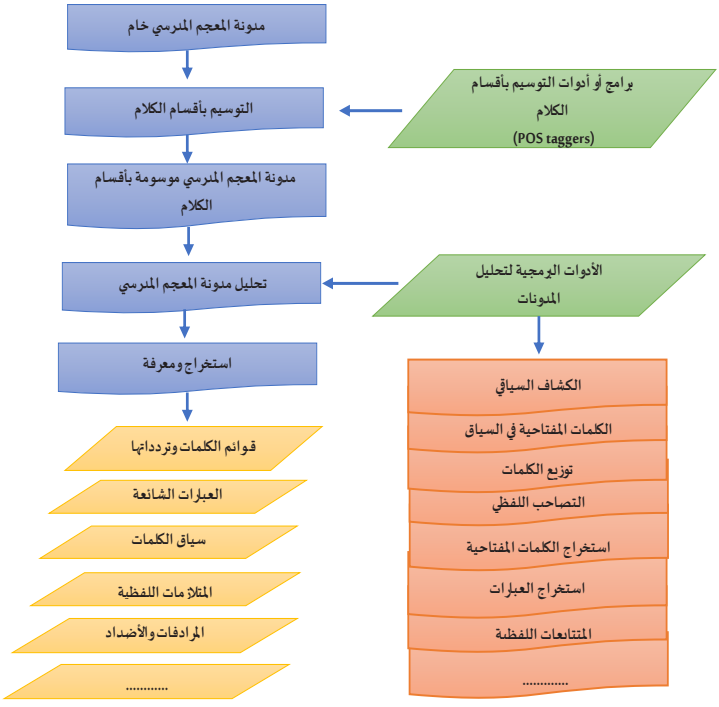
شكل رقم 04

المعلومات التي تقدمها المدونات من خلال الأدوات البرمجية.



شكل رقم 05

المراحل المتبعة بعد جمع مدونة المعجم المدرسي انطلاقا من الكتب الدراسية



خاتمة

إن استثمار نهج لسانيات المدونات في مجال التعليم، وتحديدا في مجال الصناعة المعجمية المدرسية، من شأنه أن يسهم بشكل كبير في صناعة معاجم مدرسية، تتناسب مع مستويات التلاميذ التعليمية والعمرية، وتلبي احتياجاتهم اللغوية والمعرفية، وتنمي رصيدهم اللغوي والمعرفي.

ومن المهم، أن تقوم صناعة المعاجم المدرسية على مناهج علمية، تتصل بتحليل البيانات اللغوية واستنتاج الأمثلة العملية من السياقات الواقعية في الوسط التعليمي. لذا، يجب أن تتخذ هذه المعاجم من لسانيات المدونات واقعا تكنولوجيا يمكن الاستفادة منه بشكل أفضل في تحليل البيانات اللغوية وتقديم تعريفات دقيقة وأمثلة غنية. لذلك، يتعين على المعجميين أن يعكفوا على تبني مناهج علمية حديثة والاستفادة من مجالات أخرى كاللسانيات الحاسوبية، ومجال المعالجة الآلية للغات الطبيعية خاصة، وذلك لتحسين جودة المعاجم وجعلها تلبي احتياجات التلاميذ بشكل أفضل.

ختاما، يمكن القول إن صناعة معاجم مدرسية على أساس مدونة تضم جميع الكتب الدراسية من المشاريع المهمة والضخمة التي تتطلب تخطيطا مسبقا وتنفيذا على مراحل زمنية مختلفة، وتحتاج إلى تضافر جهود الفاعلين في الشأن التربوي والتعليمي والمعنيين بقضايا اللغة والثقافة والتعليم؛ كما أن إنجازه يستدعي مشاركة مجموعة متنوعة من الكفاءات والخبراء من حاسوبيين، ولسانيين، ومعجميين وغيرهم من المتخصصين (Tarp and Gouws، 2012، ص 334 و335).

المراجع العربية

أبو العزم، عبد الغني. (2010). وظيفة المعجم المدرسي للتعليم الأساسي. *الندوة الدولية التكوينية حول المعجم المدرسي: مادته وآليات صناعته (2009)* (الصفحات 39-51). بوزريعة، الجزائر: مركز البحث العلمي والتقني لتطوير اللغة العربية.

البوشيخي، عز الدين. (1998). خصائص الصناعة المعجمية الحديثة وأهدافها العلمية والتكنولوجية. *اللسان العربي* (46).

الحاج صالح، عبد الرحمان. (2005). المعجم العربي والاستعمال الحقيقي للغة العربية. *مجلة المجمع الجزائري للغة العربية* (1)، 13-25.

الرفيق، محمد. (2020). المعجم المدرسي دراسة تحليلية للوافر الوجيز (الإصدار 1). عالم الكتب الحديث للنشر والتوزيع.

الدكروري، أيمن. (2018). المدونات اللغوية ودورها في معالجة النصوص العربية (الإصدار 1). (مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية، المحرر) الرياض، المملكة العربية السعودية: دار وجوه للنشر والتوزيع.

السعيد، المعتز بالله. (2019). المدونات اللغوية. تأليف محسن رشوان، والمعتز بالله السعيد، *الموارد اللغوية الحاسوبية* (الصفحات 51-92). الرياض: مركز الملك عبد الله بن عبد العزيز الدولي لخدمة اللغة العربية.

الشهري، علي غرامة. (2022). لسانيات المدونات: تعريف ومجالات إفادة. *المجلة العربية للعلوم ونشر الأبحاث ، 8* (3)، 77-86.

الصوري، عباس. (1988). في الممارسة المعجمية للمتن اللغوي. *اللسان العربي* (45)، 9-32.

الفيفي، عبد الله بن يحيى. (2023). وسم المدونات اللغوية: المفهوم والمجالات *Humanities and Social Sciences Series, 38* (1).

الفيلالي، عبد الكريم. (2014). نحو بناء معدم مدرسي موسوعي. *مجلة علوم التربية* (53)، 152-153.

القاسمي، علي. (2019). *علم المصطلح: أسسه النظرية وتطبيقاته العملية* (الإصدار 2). بيروت: مطبعة لبنان ناشرون.

الودغيري، عبد العلي. (2001). *دراسات معجمية: نحو قاموس عربي تاريخي وقضايا أخرى*. الدار البيضاء: مطبعة النجاح الجديدة.

عمر لحسن، وكريم مرادي. (2020). الخصائص المعجمية للمعجم المدرسي الجزائري. *مجلة الصوتيات*، 16 (02)، 255-272.

قريش، عبد العزيز. (2007). الإشكاليات في المعجم المدرسي. تأليف أحمد بريسول، وكنزة بنعمر، *المعجم العربي العصري وإشكالاته* (الصفحات 175-193). الرباط: معهد الدراسات والأبحاث للتعريب.

مختار، عمر أحمد. (2009). *صناعة المعجم الحديث* (الإصدار 2). القاهرة: عالم الكتب.

معتوق، أحمد محمد. (1996). الحصيلة اللغوية: أهميتها - مصادرها - وسائل تنميتها. الكويت: عالم المعرفة.

Almujaiwel, S., & Al-Thubaity, A. (2016, August). Arabic corpus processing tools for corpus linguistics and language teaching. In *Makalah Seminar. Online. Tersedia di https://www. researchgate. net/publication/309351881(Diunduh 15 September 2016)*.

Alfaifi, A., & Atwell, E. (2016). Comparative evaluation of tools for Arabic corpora search and analysis. *International Journal of Speech Technology*, *19*, 347-357.

Adamou E. (2019). Corpus linguistic methods. J. Darquennes, J. Salmons & W. Vandenbussche (eds.) Language contact. Boston & Berlin: Mouton de Gruyter.

Atkins, B. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. OUP Oxford.

Bennett, G. (2010). *Using CORPORA in the Language Learning Classroom: Corpus Linguistics for Teachers.* The University of Michigan Press: https://doi.org/10.3998/mpub.371534.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonitics* (6 ed.). Blackwell Publishing.

Granger, S. (2018). Has lexicography reaped the full benefit of the (learner) corpus revolution?. In *The XVIII EURALEX International Congress* (p. 208).

Gries, S. T., & Berez, A. L. (2017). Linguistic annotation in/for corpus linguistics. *Handbook of linguistic annotation*, 379-409.

Kanakaraddi, S. G., & Nandyal, S. S. (2018, March). Survey on parts of speech tagger techniques. In 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-6). IEEE.

Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.

Kilgarriff, A., & Kosem, I. (2012). *Corpus tools for lexicographers* (pp. 31-55). na.

Leech, G. (1991, August 4-8). Corpora and theories of linguistic performance. (J. Svartvik, Ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm* , 105-126.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus.*

Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing.

McEnery, T., & Wilson, A. (1996). Corpus Linguistics: An Introduction. Edinburgh University Press.

McEnery, T., & Gabrielatos, C. (2006). English corpus linguistics. *The handbook of English linguistics*, 33-71.

McEnery T. & Wilson A. (2001). *Corpus linguistics: an introduction* (2nd ed.). Edinburgh University Press.

McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, *39*, 74-92.

McEnery T. & Hardie A. (2012). *Corpus linguistics: method theory and practice*. Cambridge Université Press.

Nesi, H. & Warwick, U. K. (2000). Electronic dictionaries in second language vocabulary comprehension and acquisition: The state of the art. In *IX EURALEX International Conference* (pp. 839-847).

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Rundell, M & Stock, P. (1992). The corpus revolution. *English Today*, *8*, pp 21-32 doi:10.1017/S0266078400006520

Rundell, M. and Kilgarriff, A. (2011). 'Automating the creation of dictionaries: where will it all end?', *in A Taste for Corpora. A tribute to Professor Sylviane Granger.* Meunier F., De Cock S., Gilquin G. and Paquot M. (eds.), Benjamins. 257-281.

Rundell, M. (Ed.). (2004). *Macmillan school dictionary: for advanced learners*. Macmillan Education.

Sarker, S. (2021). Bnlp: Natural language processing toolkit for Bengali language. *arXiv preprint arXiv:2102.00405*.

Sinclair, J. (2004). *Trust the text: Language, corpus, and discourse*. Routledge.

Sinclair, J. (2005). Corpus and text-basic principles. *Developing linguistic corpora: A guide to good practice*, *92*, 1-16.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. John Wiley & Sons.

Tarp, S., & Gouws, R. H. (2012). School dictionaries for first-language learners. *Lexikos*, *22*, 333-351.

Teubert, W. (2001). Corpus linguistics and lexicography. *International journal of corpus linguistics*, *6*.

Tognini-Bonelli, E. (2001). Corpus linguistics at work. *Corpus Linguistics at Work*, 1-236.

Williams, G. (2002). In search of representativity in specialised corpora: Categorisation through collocation. *International Journal of Corpus Linguistics*, *7*(1), 43-64.

Williams, G. (2003). From meaning to words and back: Corpus linguistics and specialised lexicography. *Asp. la revue du GERAS*, (39-40), 91-106. https://doi.org/10.4000/asp.1320.

بيانات الباحثين

1. حافظ معاشي، طالب دكتوراه، ينتمي إلى مختبر البحث (تدريس اللغة العربية ودراسات اللسانيات التطبيقية)، ضمن تكوين البحث (اللسانيات وقضايا اللغة العربية)، في كلية الآداب والعلوم الإنسانية، جامعة محمد الخامس الرباط، المغرب، حاصل على ماستر المعالجة الآلية للغات العامة واللغات الخاصة من نفس الكلية. تدور اهتماماته البحثية حول المعالجة الآلية للغات، ولسانيات المدونات، وصناعة المعاجم المدرسية.

2. حكيمة خمار، أستاذة التعليم العالي، قسم اللغة العربية، في كلية الآداب والعلوم الإنسانية، جامعة محمد الخامس الرباط، المغرب، حاصلة على درجة الدكتوراه في اللسانيات، من جامعة ابن طفيل بالقنيطرة، سنة 2009 ودبلوم الدراسات العليا (دكتوراه السلك الثالث) في اللسانيات من جامعة محمد الخامس بالرباط، سنة 1999، تدور اهتماماتها البحثية حول مجالات اللسانيات الحديثة العامة والخاصة وخاصة آفاق تطور البحث العلمي في الدراسات اللغوية (المعجم- الدلالة -التركيب -اللسانيات التطبيقية-المعالجة الآلية للغة الطبيعية...).

AUTHORS BIODATA

1. Hafid Maachi, is a Ph.D. student affiliated with the Research Laboratory (Arabic Teaching and Applied Linguistic Studies), at the Research Formation (Linguistics and Arabic Language Issues), Faculty of Arts and Human Sciences, University Mohammed V, Rabat (Morocco). He holds a Master's degree in Automated Processing of General and Special Languages from the same faculty. His research interests revolve around Natural Language Processing, Corpus Linguistics, and School Lexicography.

2. Hakima Khamar is a Professor of Higher Education in the Department of Arabic Language, Faculty of Arts and Humanities, University Mohammed V, Rabat (Morocco). She holds a Ph.D. in Linguistics, from Ibn Tofail University (Kenitra, Morocco), which she obtained in 2009, and a postgraduate degree in Linguistics from University Mohammed V, in Rabat (1999). Her research interests include general and specialized linguistics, with a focus on linguistic studies such as lexicon, semantics, syntax, applied linguistics, and natural language processing.

معرف أوركيد (**ORCID**): 0009-0000-9909-9706

معرف أوركيد(**ORCID**): 0009-0007-1082-0933

Email: hafid.maachi@um5r.ac.ma

Email : hakima.khamar@flsh.um5.ac.ma