جامعة الأزهر
كلية اللغة العربية بإيتاي البارود
المجلة العلمية

مدونة الثروة اللفظية في المملكة العربية السعودية: البناء اللغوي وآفاق الاستفادة التعليمية والبحثية

إعداد

عبد العزيز بن عبدالله صالح المهيوبي

قسم الإعداد اللغوي، معهد تعليم اللغة العربية، جامعة الإمام محمد بن سعود الإسلامية، المملكة العربية السعودية

(العدد الثامن والثلاثون)

(الإصدار الثالث .. أغسطس)

(١٤٤٧هـ - ٢٠٢٥م)

علمية - محكمة - ربع سنوية

الترقيم الدولي: ISSN 2535-177X

مدونة الثروة اللفظية في المملكة العربية السعودية: البناء اللغوي وآفاق الاستفادة التعليمية والبحثية.

عبدالعزيز بن عبدالله صالح المهيوبي.

قسم الإعداد اللغوي، معهد تعليم اللغة العربية، جامعة الإمام محمد بن سعود الإسلامية، المملكة العربية السعودية.

البريد الإلكتروني: aamahyoubi@imamu.edu.sa

الملخص:

تناولت هذه الدراسة مدونة "الثروة اللفظية" في المملكة العربية السعودية، وهي مدونة لغوية تعليمية محوسبة توثق الأداء اللغوي الواقعي لطلاب وطالبات مراحل التعليم العام في المملكة عام ٢٠٠٦ م. وقد جُمعت نصوص المدونة من مصادر متنوعة شملت نصوصًا منطوقة ومكتوبة، وكراسات مدرسية، وكتبًا دراسية، ومجلات مصورة، وقصصًا وروايات، ومواد مرئية ومسموعة، ليصل حجمها إلى أكثر من ٣.٥ مليون كلمة. وتهدف الدراسة إلى توصيف مكونات المدونة وبنيتها اللغوية، مع استعراض الأدوات التقنية التي طورها الباحث خصيصًا لمعالجة البيانات النصية، مثل أداة استخراج قوائم الشيوع، والمكشاف السياقي (KWIC)، واستخراج المتصاحبات اللفظية، والبحث الصرفي المعتمد على الجذر. وقد أظهرت نتائج التحليل إمكانات واسعة للمدونة في خدمة البحث اللغوي التطبيقي، وتحسين تعليم اللغة العربية، وتطوير المحتوى التعليمي، وتصميم المناهج المبنية على الأداء الواقعي. كما خلصت الدراسة إلى أن الاستثمار التربوي في هذه المدونة يسهم في تعزيز التعلم المعتمد على المدونات، ويعكس توجهًا حديثًا وواعدًا في التعليم. وتوصي الدراسة بتوسيع نطاق استخدام المدونة في المؤسسات التعليمية.

الكلمات المفتاحية: المدونة اللغوية، الثروة اللفظية، تعليم العربية، الشيوع، الأدوات الحاسوبية، المتعلمون، التحليل اللغوي.

The Lexical Wealth Corpus in the Kingdom of Saudi Arabia: Linguistic Structure and Prospects for Educational and Research Utilization.

Abdulaziz bin Abdullah Saleh Al-Mahyoubi.
Linguistic Preparation Department, Arabic Language Institute, Imam Muhammad ibn Saud Islamic University, Kingdom of Saudi Arabia.
Email: aamahyoubi@imamu.edu.sa

Abstract:
This study focused on the "Verbal Wealth" Corpus in the Kingdom of Saudi Arabia, which is a computerized educational linguistic corpus that documents the real linguistic performance of male and female students in general education stages in the Kingdom during the year 2006. The corpus texts were collected from various sources, including spoken and written texts, school notebooks, textbooks, illustrated magazines, stories and novels, as well as audiovisual materials, with a total size exceeding 3.5 million words. The study aims to describe the components and linguistic structure of the corpus, while reviewing the technical tools specifically developed by the researcher for processing textual data, such as a word frequency extractor, Key Word in Context (KWIC) tool, collocation extractor, and root-based morphological search tool. The results of the analysis revealed the corpus's wide potential in supporting applied linguistic research, enhancing Arabic language education, developing educational content, and designing curricula based on real language performance. The study concludes that the educational investment in this corpus contributes to promoting corpus-based learning and reflects a promising and modern approach in education. It recommends expanding the use of the corpus in educational institutions.

Keywords: Corpus linguistics, lexical wealth, Arabic language teaching, frequency analysis, learner data, computational tools, educational applications.

المقدمة

شهدت الدراسات اللغوية المعاصرة تحولًا كبيرًا بفعل تطور مستحدثات التكنولوجيا الرقمية وتطبيقاتها في معالجة اللغة، ومن أبرز هذه التحولات اعتمادها على المدونات اللغوية المحوسبة (Corpora) بوصفها مصدرًا غنيًا وواقعيًا لدراسة اللغة كما تُستخدم فعليًا في السياقات الطبيعية، بعيدًا عن التصورات الانطباعية والحدس أو الأمثلة المصطنعة. وقد أصبحت هذه المدونات أداة مركزية في مجالات تعليم اللغات، وتحليل الخطاب، والمعالجة الحاسوبية للغة، وصناعة المعاجم، وغيرها (McEnery & Hardie, 2012).

أحدثت هذه المدونات نقلة نوعية في مجالات متعددة، من أبرزها: تعليم اللغات، وتحليل الأخطاء، وتخطيط المناهج، والبحث اللساني التطبيقي، لأنها تمكّن الباحث من تتبع الظواهر اللغوية كما تظهر في الاستخدام الفعلي، بعيدًا عن الافتراضات المسبقة أو النماذج المصطنعة. ومع تطور تقنيات التحليل النصي، والبرمجيات الخاصة بمعالجة اللغة الطبيعية (NLP)، بات من الممكن إجراء تحليلات إحصائية ومعجمية، وصوتية، وصرفية، ونحوية، ودلالية واسعة النطاق، مما أضفى طابعًا علميًا وموضوعيًا على دراسات اللغة.

وفي السياق العربي، اتجهت الجهود البحثية والتربوية نحو بناء مدونات لغوية عربية تُراعي الخصائص البنائية والتداولية للغة، وتُسهم في خدمة أهداف التعليم والبحث، مثل مدونة فلك، ومدونة مكتبة الإسكندرية وغيرها. ورغم أن بعض هذه المدونات اعتمد على نصوص فصيحة من التراث، أو نصوص إعلامية حديثة، إلا أن الحاجة ما تزال قائمة لمدونات تعليمية تمثّل الواقع اللغوي للمتعلم العربي، خاصة في البيئة المدرسية.

ومن هنا تأتي أهمية مدونة الثروة اللفظية في المملكة العربية السعودية، التي تمثّل مشروعًا وطنيًا رائدًا في ميدان التوثيق اللغوي والتربوي، إذ تجمع أكثر من ٣,٥٠٠,٠٠٠ كلمة من مصادر متنوعة، تشمل: القرآن الكريم، والحديث

النبوي، والشعر القديم والحديث، والأمثال، والنصوص المدرسية، والمحتوى الإعلامي الموجّه للشباب، فضلًا عن النصوص الواقعية المكتوبة والمنطوقة التي تم جمعها من طلاب وطالبات المدارس في مختلف مناطق المملكة.

تُعدّ مدونة "الثروة اللفظية" من أبرز النماذج التي تمثل اللغة العربية كما تُستخدم فعليًا في البيئات التعليمية، فهي تجمع بين التنوع الجغرافي والدلالي للنصوص، والصرامة المنهجية في جمعها وتصنيفها. ويُضفي إشراك المعلمين والمعلمات في عملية التوثيق بعدًا تربويًا موثوقًا، يجعل من هذه المدونة أداة فاعلة يمكن الاعتماد عليها في قياس المستوى الحقيقي للطلبة في اللغة، واتخاذ قرارات تعليمية مدروسة تستند إلى بيانات دقيقة وواقعية.

وتهدف هذه الدراسة إلى إلقاء الضوء على الهيكل اللغوي الذي بُنيت عليه مدونة "الثروة اللفظية"، واستعراض الأسس المعتمدة في توثيقها وتنظيمها، إلى جانب إبراز مجالات توظيفها في التعليم والبحث والتطبيق العملي. وتشمل هذه المجالات تطوير مناهج تعليم اللغة العربية، وتصميم أدوات تحليل لغوي دقيقة تساعد على تقييم الأداء اللغوي للمتعلمين، وبناء محتوى تعليمي قائم على بيانات حقيقية نابعة من الواقع اللغوي للطلاب.

١ الإطار العام للدراسة

١-١ مشكلة البحث

في ظل التطور المتسارع في مجالات التعليم والتقنية والبحث اللغوي، برزت المدونات اللغوية المحوسبة كأداة حيوية لتحليل اللغة، وتطوير مناهج تعليمها، لا سيما في اللغات ذات الامتداد الحضاري الكبير كاللغة العربية. وعلى الرغم من تزايد عدد المدونات اللغوية العربية خلال العقدين الأخيرين، فإن أغلب هذه المدونات ما زال يفتقر إلى بنية منهجية واضحة، أو تمثيل واقعي حقيقي للغة كما يستخدمها المتعلم العربي في السياقات التعليمية المختلفة.

يعاني المشهد التربوي العربي من فجوة واضحة بين المحتوى التعليمي النظري، وبين الواقع اللغوي الفعلي للطلاب، وهو ما يؤدي إلى مشكلات في التقييم، وصعوبة في تطوير مناهج ملائمة، إضافة إلى غياب أدوات فعالة لقياس الكفاية اللغوية. فالعديد من المدونات المتوفرة ترتكز إما على نصوص تراثية، أو نصوص إعلامية وصحفية، دون تمثيل كافٍ للغة الطلاب، وتحديدًا في المراحل الأساسية والمتوسطة والثانوية.

ومن جهة أخرى، تواجه المؤسسات التعليمية والبحثية تحديات كبيرة في الوصول إلى مدونات مفتوحة، شاملة، ومتجددة، تُغطي أشكالاً مختلفة من الأنماط النصية والوظيفية التي تعكس واقع الطالب السعودي والعربي عمومًا. ويزداد هذا التحدي في ظل غياب المعايير الواضحة لبناء مدونات لغوية تعليمية تستجيب لحاجات العصر، وتتيح للمؤسسات التعليمية توظيفها في بناء المناهج، وتطوير أدوات التقييم، وتحليل أخطاء المتعلمين، ورصد نموهم اللغوي.

من هنا، تبرز أهمية هذه الدراسة في سد هذه الفجوة، من خلال تسليط الضوء على نموذج تطبيقي بارز هو مدونة الثروة اللفظية، التي تم بناؤها برؤية تعليمية منهجية، وشهدت مشاركة ميدانية واسعة من المعلّمين والطلاب. وتسعى الدراسة إلى تحليل البنية النصية واللغوية لهذه المدونة، واستكشاف إمكانات توظيفها تربويًا وبحثيًا، بما يسهم في تطوير مخرجات تعليم اللغة العربية وفق أسس علمية قائمة على بيانات واقعية.

١-٢ أسئلة البحث

١. ما آلية بناء مدونة الثروة اللفظية، وما طبيعة محتواها النصي؟
٢. ما الخصائص اللغوية التي تتميز بها هذه المدونة من حيث الشيوع، والتنوع، والتمثيل الواقعي؟
٣. كيف يمكن الاستفادة من مدونة الثروة اللفظية في تطوير مناهج تعليم اللغة العربية؟

٤. ما أوجه الإفادة من المدونة في مجالات البحث اللغوي وتحليل اللغة الطبيعية؟

١-٣ أهداف البحث

١. وصف بنية مدونة الثروة اللفظية ومكوناتها.
٢. تحليل الخصائص اللسانية والنصية للمدونة.
٣. بيان مجالات الاستفادة التربوية والبحثية من المدونة.
٤. اقتراح آليات لتوظيف المدونة في تطوير أدوات تعليمية وتقويمية.

١-٤ أهمية البحث

تنبع أهمية هذا البحث من كونه يتناول واحدة من أكثر المدونات اللغوية التعليمية واقعية وارتباطًا بالميدان التربوي، وهي مدونة الثروة اللفظية، التي تمثل مرجعًا لغويًا غنيًا يعكس اللغة كما يستخدمها الطلاب والطالبات في مختلف المراحل الدراسية في المملكة العربية السعودية. وتمثل هذه المدونة نقلة نوعية في حقل اللسانيات التطبيقية والتعليمية، لما تحمله من ثراء لغوي، وتنوع نصي، وتمثيل واقعي نادرًا ما يتوافر في المدونات التقليدية.

وتكمن القيمة التطبيقية لهذا البحث في أنه يُسلّط الضوء على كيفية توظيف هذه المدونة في تحسين تدريس اللغة العربية، وتطوير المناهج الدراسية بناءً على بيانات حقيقية. ويُسهم البحث في توجيه الأنظار إلى أهمية استثمار المدونات التعليمية في قياس نمو المتعلمين اللغوي، وتحليل أنماط الأخطاء، وتطوير أنشطة تعليمية قائمة على الشيوع، والتدرج، والسياق الواقعي.

علاوة على ذلك، فإن هذا البحث يُبرز أهمية المدونة كبنية بيانات لغوية قابلة للمعالجة الآلية، مما يُمهّد الطريق لاستخدامها في تطبيقات الذكاء الاصطناعي، وتحليل اللغة الطبيعية، وصناعة المحتوى اللغوي الرقمي، وبناء القواميس الإلكترونية، والمساعدات اللغوية التفاعلية.

كما يُعزز البحث من الاتجاه نحو إنتاج مدونات محلية تستند إلى الواقع التربوي، وتسهم في بناء خطط وطنية مستندة إلى أدلة لغوية موثقة. ويُعد من أوائل الدراسات التي تسبر أغوار هذه المدونة من منظور منهجي وتحليلي تطبيقي، مما يجعله ذا أهمية خاصة للباحثين، ومصممي المناهج، والمختصين في تعليم اللغة العربية وتقييمها.

١-٥ منهج البحث

يعتمد هذا البحث على المنهج الوصفي التحليلي، حيث يقوم بوصف دقيق لبنية مدونة الثروة اللفظية من حيث مصادرها، وتنوعها النصي، ومستوياتها اللغوية، إلى جانب تحليل مكوناتها المعجمية، والدلالية، والتركيبية. كما يستند إلى تحليل نوعي للمحتوى بهدف استنباط إمكانات الاستفادة التربوية، والبحثية من هذه المدونة. ويُستثمر هذا المنهج في ربط المعطيات اللغوية بالوظائف التعليمية، وتقديم تصورات تطبيقية لاستخدام المدونة في التعليم والتقويم.

١-٦ حدود البحث

تُحدد هذه الدراسة بمجموعة من الحدود العلمية والمنهجية، من أبرزها أن التحليل يقتصر على البنية العامة لمدونة الثروة اللفظية، من حيث مصادرها، وتنوع نصوصها، وخصائصها اللغوية الأساسية، دون التوسع في التحليل الإحصائي الكمي الدقيق لجميع مفردات المدونة أو معالجة البيانات باستخدام أدوات برمجية متقدمة. كما أن نطاق الاستفادة التعليمية في هذه الدراسة ينحصر في تعليم اللغة العربية للناطقين بها في السياق السعودي تحديدًا. كذلك، لا تشمل الدراسة مقارنة تفصيلية بين هذه المدونة ومدونات لغوية أخرى مماثلة، وإنما تركز على تحليلها بوصفها نموذجًا مستقلًا.

١-٧ مصطلحات البحث

المدونة اللغوية (Corpus):

هي مجموعة ضخمة من النصوص اللغوية المكتوبة أو المنطوقة، جُمعت إلكترونيًا وفق معايير منهجية محددة، تُستخدم في التحليل اللساني النوعي أو الكمي، وتمثّل اللغة كما تُستخدم فعليًا في سياقاتها الواقعية. تلعب المدونات دورًا أساسيًا في تطوير المعاجم، وتحليل الخطاب، وتعليم اللغات، وبحوث الذكاء الاصطناعي (McEnery & Hardie, 2012).

مدونة الثروة اللفظية (Tharwa Corpus):

هي مدونة لغوية تعليمية سعودية طُوّرت برعاية وزارة التعليم في المملكة العربية السعودية، وتضم ما يزيد على ٣,٥٠٠,٠٠٠ كلمة. تشمل نصوصًا من القرآن الكريم، والحديث النبوي، والشعر، والخطب، والمقررات الدراسية، إضافة إلى كتابات وتسجيلات واقعية لطلاب المدارس. وتتميز بأنها تمثل الواقع اللغوي للمتعلمين السعوديين في مختلف المناطق والمراحل الدراسية.

الشيوع اللغوي (Word Frequency):

يشير إلى عدد مرات ظهور كلمة أو تركيب معين في مدونة لغوية، ويُعد من المقاييس المهمة في تحليل المعجم النشط للغة. وتُستخدم إحصاءات الشيوع في إعداد مواد تعليمية موجهة، وتصميم اختبارات لغوية فعالة، واختيار المفردات ذات الأولوية في التعليم (Nation, 2001).

التحليل المعجمي (Lexical Analysis):

هو دراسة مكونات المعجم اللغوي من حيث التكرار، والتنويع، والعلاقات الدلالية، والمجالات السياقية. ويُستخدم في تحليل المفردات النشطة والسلبية، والكلمات الأكاديمية، وأسلوب التراكيب المستخدمة. كما يُعد أساسًا في تطوير المعاجم التعليمية والمعالجة الحاسوبية للغة (Scott & Tribble, 2006).

التحليل الحاسوبي للغة (Computational Linguistic Analysis):

هو استخدام الحاسوب في معالجة وتحليل النصوص اللغوية من خلال خوارزميات وتقنيات البرمجة، ويشمل مجالات مثل التعرف على الأنماط اللغوية، التحليل النحوي والدلالي، واستخراج المصطلحات، وتطبيقات الذكاء الاصطناعي. ويُعد فرعًا من أفرع علوم اللغة الحديثة (Jurafsky & Martin, 2023).

٢ الإطار النظري والدراسات السابقة

تُعد المدونات اللغوية المحوسبة من أبرز المنجزات التي حققها البحث اللساني المعاصر، لا سيما في ظل التطور المتسارع في تقنيات معالجة اللغة الطبيعية، وتطبيقات الذكاء الاصطناعي. فقد شكّلت هذه المدونات نقلة نوعية في الطريقة التي يُدرس بها النص اللغوي، إذ تجاوزت النماذج النظرية المغلقة إلى فضاء اللغة الواقعية كما تُستخدم في السياقات اليومية، والتربوية، والمهنية. وهي بذلك توفر للباحثين، والمعلمين، ومطوري المناهج، قاعدة بيانات موثوقة، حية، ومتنوعة، تتيح تحليلًا كميًا ونوعيًا دقيقًا للمفردات والتراكيب والأنماط اللغوية.

تكمن أهمية المدونات اللغوية في أنها توثق اللغة بعيدًا عن التصورات الافتراضية أو النماذج النحوية المثالية، إذ تتيح للباحث تتبّع المفردات من حيث التكرار، والسياق، والمعنى، والوظيفة. كما أنها تتيح رسم خريطة حقيقية لاستخدام اللغة بحسب المستوى العمري، أو التخصص، أو الموقع الجغرافي، مما يجعلها أداة لا غنى عنها في مجالات متعددة، مثل: تحليل الأخطاء اللغوية، وتعليم المفردات، وتحليل الخطاب، وتصميم اختبارات الكفاية، وبناء المعاجم، وتطوير المناهج الدراسية، وتصميم الأنشطة التواصلية (Kennedy, 1998).

وليس هذا الاستخدام مقصورًا على البحث اللغوي فحسب، بل يمتد إلى الفضاء التعليمي، حيث ساعدت المدونات التعليمية المتخصصة على تطوير

المحتوى اللغوي، وربطه بواقع المتعلمين. ويُلاحظ أن البرامج التعليمية الحديثة التي تُبنى على مدونات لغوية واقعية تحقق نتائج أفضل من تلك التي تستند إلى أمثلة مصطنعة أو نصوص تراثية مجردة. كما أن هذه المدونات تُعد موردًا أساسيًا للمعلمين لتوليد أنشطة لغوية واقعية، وتصميم إستراتيجيات تعلم تعتمد على البيانات.

من هنا، يُعد تناول المدونات اللغوية في هذا البحث، وخصوصًا مدونة الثروة اللفظية، مدخلًا علميًا للكشف عن الإمكانات الحقيقية التي توفّرها هذه الأدوات في تطوير تعليم اللغة العربية استنادًا إلى الواقع، لا إلى الافتراض. وسيُبنى هذا الإطار النظري على عرض مفاهيم أساسية تشمل: طبيعة المدونات اللغوية، مفهوم الشيوع والتنوع، ودور المدونات التعليمية، باعتبارها الركائز النظرية لفهم وظيفة مدونة الثروة اللفظية وتوظيفها.

٢-١ المدونة اللغوية المحوسبة (Corpus)

تُعرّف المدونة اللغوية بأنها مجموعة منظمة ومنتقاة من النصوص اللغوية الواقعية، سواء كانت مكتوبة أو منطوقة، تُجمع إلكترونيًا وتُخزن بصيغ قابلة للمعالجة الحاسوبية، وتُستخدم في مجالات متعددة أبرزها: البحث اللساني، وتعليم اللغات، وإعداد المعاجم، وتصميم الاختبارات اللغوية. وغالبًا ما تُخضع هذه المدونات لعمليات ترميز لغوي (Annotation)، تشمل الجوانب الصرفية، والنحوية، والدلالية، مما يجعلها قابلة للتحليل الكمي والنوعي بدقة منهجية عالية (Kennedy, 1998).

وقد ظهرت الحاجة إلى المدونات اللغوية في أواسط القرن العشرين، كردّ فعل على النماذج اللغوية الانعزالية التي تفتقر إلى الشواهد الواقعية. وجاء مشروع Brown Corpus عام ١٩٦٤ في الولايات المتحدة بوصفه أول مدونة إلكترونية موجهة لأغراض التحليل اللغوي، إذ جمعت مليون كلمة من نصوص اللغة الإنجليزية الأمريكية المتنوعة. وقد مثّل هذا المشروع انطلاقة حقيقية نحو

نشوء ما يُعرف اليوم بـ اللسانيات المعتمدة على المدونات (Corpus Linguistics).

مع مرور الوقت، تطوّرت أنواع المدونات، وتوسعت استخداماتها، ويمكن تصنيفها إلى فئات رئيسة، منها:

٢-١-١ المدونات المرجعية (Reference Corpora):

وهي مدونات واسعة النطاق، تهدف إلى تمثيل شامل للغة في سياقاتها العامة، وتُستخدم عادةً كقاعدة معيارية للقياس والمقارنة.

٢-١-٢ المدونات الخاصة أو المتخصصة (Specialized Corpora):

تُبنى هذه المدونات لأغراض محددة، وتشمل نصوصًا من مجال معرفي معين، كالقانون، والطب، والصحافة، والتعليم، إلخ. وتُستخدم في تحليل اللغة التخصصية ومفردات المجالات.

٢-١-٣ المدونات التعليمية (Learner Corpora):

تشتمل على إنتاجات لغوية واقعية للمتعلمين، وتُستخدم في تحليل أخطاء التعلم، وأنماط الاستخدام، وبناء المحتوى التعليمي الفعّال. وقد تطوّرت هذه المدونات مع تصاعد الاهتمام باللغويات التطبيقية وفعالية تعليم اللغات الأجنبية.

٢-١-٤ المدونات المنطوقة (Spoken Corpora):

تتضمن تسجيلات حقيقية للمحادثات، والمناقشات، والخطب، تُفرّغ وتُعنون لغويًا. وتُعد مصدرًا مهمًا في دراسة الاستعمال الشفهي، والنحو التفاعلي، وتحليل التداول اللغوي.

وتتميز المدونات المحوسبة بأنها توفر إمكانات متقدمة للتحليل اللغوي عبر برامج بحثية متخصصة، مثل AntConc و Sketch Engine و WordSmith Tools، مما يُمكّن الباحث من استخراج معلومات دقيقة حول التكرار، وتوزيع الكلمات، وأنماط السياق، والتغير اللغوي، وغيرها من الظواهر (Biber, Conrad, & Reppen, 1998).

إن فهم طبيعة المدونات وأنواعها شرط أساسي لفهم إمكانات مدونة "الثروة اللفظية" التي يُعالجها هذا البحث، والتي تُمثّل مدونة تعليمية متخصصة ذات بُنية مرجعية وأهداف تربوية واضحة.

٢-٢ الشيوع والتنوع اللغوي (Word Frequency)

يمثّل الشيوع اللغوي أحد أبرز المفاهيم المحورية في تحليل المدونات اللغوية، ويُقصد به عدد تكرارات الكلمة أو التركيب داخل مدونة معينة. وتُظهر دراسات عديدة في تعليم اللغة أن الكلمات الأكثر شيوعًا تُشكل قاعدة المفردات الأساسية التي ينبغي تدريسها للمتعلمين في المراحل الأولى، لأنها تُمكّنهم من فهم جزء كبير من النصوص الحقيقية في أقصر وقت. ويُعد معيار الشيوع مقياسًا موضوعيًا يساعد في تصنيف الكلمات من حيث الأولوية التعليمية، وتصميم مناهج تعليمية تراعي التدرج الواقعي في استخدام المفردات (Coxhead, 2000).

وقد بينت أبحاث مثل أبحاث Nation (٢٠٠١) و Coxhead (٢٠٠٠) أن تغطية نحو ٢٠٠٠–٣٠٠٠ كلمة عالية الشيوع كافية لفهم ما يزيد عن ٨٥% من المفردات في النصوص العامة، وهو ما جعل قوائم الشيوع (Frequency Lists) أداة جوهرية في تطوير المعاجم التعليمية، وبناء اختبارات المفردات، وتقويم الكفاية اللغوية.

أما التنوع اللغوي (Linguistic Diversity)، فيُشير إلى غنى المحتوى من حيث توزيع المفردات، وتعدد الحقول الدلالية، وتنوّع التركيبات النحوية والوظيفية. ويُعد التنوع مؤشرًا على ثراء المدونة، ومدى ملاءمتها لمختلف المستويات التعليمية. وتُظهر المدونات الضعيفة في هذا الجانب تكرارًا محدودًا أو تمثيلًا لغويًا نمطيًا، مما يحدّ من فائدتها في التعليم والتحليل (Schmitt & Schmitt, 2020).

ويمكّن الجمع بين مبدأي الشيوع والتنوع الباحثين من تحليل التوازن اللغوي في النصوص، وتحديد مدى ملاءمتها للتعليم، وتطوير المحتوى التعليمي بناء على أولويات واقعية. كما يُستخدم هذا التحليل في تحديد المفردات المتخصصة، ورصد التغيرات اللغوية عبر الزمن أو في بيئات مختلفة.

٢-٣ المدونات التعليمية (Learner Corpora)

تمثل المدونات التعليمية، فرعًا متخصصًا من المدونات اللغوية يركّز على جمع وتحليل الإنتاج اللغوي للمتعلمين، سواء في شكل نصوص مكتوبة أو تسجيلات صوتية. وتُستخدم هذه المدونات في دراسة أنماط تعلم اللغة، وتحليل الأخطاء الشائعة، والتعرف على الفجوات المعجمية والتركيبية، مما يُسهم في تحسين طرائق التدريس، وتصميم مناهج مبنية على البيانات (Granger, 2002).

نشأ الاهتمام بهذا النوع من المدونات في أواخر الثمانينيات وبداية التسعينيات، مع تطور علم اللغة التطبيقي. وتقوم هذه المدونات على مبدأ أن تحليل لغة المتعلم الواقعية أفضل من الاعتماد على نماذج لغوية معيارية، لأنها تُظهر خصائص تطور اللغة، والتداخل اللغوي، وأثر اللغة الأم في التعلم، ومواطن الصعوبة الفعلية.

تُستخدم المدونات التعليمية في إعداد كتب تعليم اللغة المبنية على أخطاء حقيقية شائعة، وتصميم اختبارات الكفاية وفق أنماط الاستخدام الواقعي، وتدريب المعلمين على التعرف على أنماط الخطأ والتنوع الأسلوبي، وبناء قواعد بيانات لدعم تعليم اللغة بمساعدة الحاسوب.

ومن أشهر المدونات التعليمية العالمية:

International Corpus of Learner English (ICLE)
Cambridge Learner Corpus (CLC)
Arabic Learner Corpus (ALC)

ويزداد الاحتياج إلى هذه المدونات في البيئات غير الناطقة بالعربية، كما في السياقات العربية نفسها، حيث يمكن لمدونات مثل مدونة الثروة اللفظية أن تسد فراغًا كبيرًا في فهم الأداء اللغوي الحقيقي للمتعلمين، وتُسهم في توجيه المحتوى التعليمي بناء على ما يقدّمه الطالب فعليًا لا ما يُفترض أنه يعرفه (Nesselhauf, 2004).

٢-٤ خصائص المدونات التعليمية

تُعد المدونات التعليمية من أهم أنواع المدونات في البحث اللغوي التطبيقي، نظرًا لأنها تمثل الأداء الفعلي للمتعلمين، وتتيح للباحثين والمعلمين تحليل الأخطاء الشائعة، واستخلاص الأنماط الأسلوبية واللغوية في كتابات المتعلمين، مما يُمكّن من تطوير مواد تعليمية واقعية ومبنية على بيانات حقيقية.

وقد أثبتت الأبحاث أن الاعتماد على المدونات التعليمية في تصميم المحتوى اللغوي يُسهم في تحسين نتائج التعلم، ويزيد من وعي المعلم بمواطن الصعوبة التي يواجهها الطلاب (Meunier & Reppen, 2015).

٢-٥ الدراسات السابقة

رغم عدم وجود دراسات تناولت مدونة الثروة اللفظية، إلا أن العديد من الأبحاث في مجال المدونات التعليمية وتحليل الشيوع وتوظيف corpora في التعليم توفر خلفية علمية داعمة لهذا البحث، وتُسهم في بناء الإطار المفاهيمي له.

دراسة ماي زكي، وديفيد ويلمسن، ودانا عبد الرحيم (٢٠٢١)

تستعرض هذه الدراسة تطور اللسانيات الحاسوبية في العالم العربي، مع التركيز على التحديات التقنية واللغوية التي تواجه تحليل النصوص العربية آليًّا، مثل الغموض الصرفي وغياب التشكيل. بيّن الباحثون أن الاعتماد التقليدي على النصوص المكتوبة دون التركيز على اللغة المحكية أدى إلى انحياز واضح في تمثيل العربية ضمن المدونات، مشيرين إلى الحاجة الماسة إلى مدونات تشمل

اللهجات. كما أشاروا إلى محدودية استخدام اللسانيين العرب للمدونات في البحث النظري، مقارنة باستخدامها في تطبيقات الترجمة والتعرف الصوتي والتعليم. وقد استعرضت الدراسة جهودًا مبكرة، مثل مدونة arabiCorpus، وأظهرت فروقًا بين الاستخدامات الأسلوبية حسب الدول. وخلص الباحثون إلى ضرورة توسيع تمثيل اللغة المحكية في المدونات، ودعم الباحثين بإمكانية وصول أيسر للموارد اللغوية الإلكترونية

دراسة عبد المحسن بن عبيد الثبيتي وعقيل بن حامد الشمري (٢٠٢٤)

تناولت هذه الدراسة توظيف المدونات الحاسوبية في تحليل لغة الصحافة العربية، مستندة إلى مدونة الصحف العربية لعام ٢٠١٢، والتي جمعت نصوصًا من ١٩ دولة عربية موزعة على ستة مجالات (سياسي، ثقافي، اقتصادي، ديني، رياضي، تقني). استخدم الباحثان أداة "غوّاص" لتحليل البيانات واستخراج التكرارات والمتصاحبات والمفردات المميزة. وهدفت الدراسة إلى تقديم نموذج مختلف لرصد وتحليل الاستخدامات اللغوية في الصحافة العربية بالاعتماد على المعالجة الآلية للنصوص. وقد أظهرت النتائج قدرة التحليل الآلي على الكشف عن ظواهر لغوية معجمية وتركيبية وأسلوبية معاصرة، مع إمكانية تطبيق تلك النتائج في إعداد معاجم حديثة وتصميم مناهج دراسية تراعي اللغة الفعلية المتداولة. وأكد الباحثان أهمية اعتماد اللسانيات الحاسوبية كمنهج موثوق في تحليل اللغة العربية الحديثة.

دراسة كمال يوسف (٢٠٢٤)

استهدفت هذه الدراسة استكشاف إمكانات توظيف اللسانيات الحاسوبية في تحسين تعلم مفردات اللغة العربية، خاصة للناطقين بغيرها. اعتمد الباحث على تحليل مدونتين لغويتين هما "مدونة العربية القرآنية" و"Sketch Engine"، بهدف التعرف على أكثر المفردات شيوعًا وسياقية. وقد تبنّت الدراسة منهجًا نوعيًّا في تحليل المدونات واستخلاص البنى المعجمية المتكررة والعلاقات بين

الكلمات. أظهرت النتائج أن الاعتماد على المدونات اللغوية يمكن أن يُحدث تحولًا في طرائق تعليم العربية، حيث ينتقل المتعلم من الحفظ المجرد إلى الفهم السياقي والربط الدلالي، مما يسهم في تطوير كفاءته اللغوية بشكل أعمق. كما خلصت الدراسة إلى أن التطبيقات الحاسوبية يمكن أن تقدم محتوىً تعليميًا تفاعليًا قائمًا على بيانات حقيقية وموثوقة، تدعم المتعلم والمعلم على حد سواء.

دراسة بشير عدة، وكمال عمامرة، ونور الدين شعنبي(2025)

تناولت هذه الدراسة أهمية المدونات اللغوية في البحث الإحصائي اللغوي، مع التركيز على مدونة اللغة العربية المعاصرة عبر منصة "فلك". استخدم الباحثون تحليلًا ميدانيًا إحصائيًا لكلمة "امرأة" في المدونة، شمل توزيعها حسب المجالات والسياقات وعدد مرات الظهور والكلمات المصاحبة لها. وهدفت الدراسة إلى إبراز الإمكانات التحليلية التي توفرها المدونات في مجال البحث اللغوي الكمي، خاصة في الوصف المعجمي والدلالي. وأظهرت النتائج أن المدونات تمكّن الباحث من الوصول إلى معطيات دقيقة قابلة للقياس والمقارنة، تساهم في الكشف عن الأنماط اللغوية الفعلية بدلًا من الاعتماد على الأمثلة المصطنعة. وأوصى الباحثون بضرورة تكثيف استخدام هذه المدونات في الأبحاث اللغوية والتعليمية لما تقدمه من ثراء وتنوع وسياقات حقيقية.

تتقاطع الدراسة الحالية مع الدراسات السابقة في اهتمامها بالمدونات اللغوية كأداة فعالة لتحليل اللغة وتوظيفها في البحث والتعليم، إلا أنها تتميّز بتركيزها على مدونة تعليمية سعودية توثق الأداء اللغوي الواقعي للطلاب والطالبات، بخلاف بقية الدراسات التي تناولت مدونات صحفية (مثل دراسة الثبيتي والشمري)، أو مدونات لغوية مقدّسة وعامة (كما في دراسة كمال يوسف)، أو ناقشت الجوانب النظرية لتطور المدونات وتحديات استخدامها (زكي وآخرون)، أو ركزت على الجانب الإحصائي في تحليل مفردات معينة (عدة وآخرون). وتتفرد الدراسة الحالية في أن المدونة المدروسة تمثل اللغة المستخدمة

فعليًا في المدارس السعودية، ما يمنحها قيمة مرجعية عالية في السياقات المحلية التعليمية.

٣ بناء مدونة الثروة اللفظية ومجالات الاستفادة منها

٣-١ الجهات المشرفة والمشاركة

تم تطوير مدونة الثروة اللفظية تحت إشراف وزارة التعليم في المملكة العربية السعودية، وذلك ضمن مشروع وطني يسعى إلى توثيق اللغة العربية كما يستعملها الطلاب في مراحل التعليم العام. وقد شاركت في المشروع فرق لغوية وتربوية متخصصة، بالإضافة إلى معلمين ومعلمات من مختلف المناطق التعليمية في المملكة، الذين أسهموا بجمع النصوص المكتوبة والشفوية من الطلاب والطالبات في بيئات تعليمية واقعية.

٣-٢ حجم المدونة

تضم مدونة الثروة اللفظية أكثر من ٣,٥٠٠,٠٠٠ كلمة، موزعة على آلاف النصوص المتنوعة. وقد روعي في بناء المدونة التوازن بين الأنواع النصية والمجالات، بما يشمل: النصوص الدينية: من القرآن الكريم والحديث النبوي الشريف، والنصوص الأدبية: من الشعر القديم والحديث، والنثر الإبداعي.، والنصوص التربوية: من المناهج والمقررات الدراسية، والنصوص الإعلامية: من الصحف والمجلات والبرامج المسموعة، النصوص الطلابية الواقعية: وهي أبرز ما يميز المدونة، إذ تشمل إنتاجات لغوية واقعية من الطلاب في مختلف المراحل.

٣-٣ تنوع المحتوى النصي

تتميز المدونة بتنوّعها الغني، فهي لا تقتصر على نصوص مكتوبة فقط، بل تشمل أيضًا تسجيلات منطوقة تم تفريغها وتوثيقها إلكترونيًا. ويشمل هذا المحتوى نصوصًا من: القرآن الكريم والحديث النبوي الشريف، والشعر العربي بأنواعه (الموزون، الحر، الشعبي)، والنصوص التربوية والتعليمية المقتبسة من

المناهج الدراسية، والمقالات والتقارير الإعلامية، حوارات طلابية وتعبيرات كتابية واقعية نُقلت من الفصول الدراسية.

٣-٤ بيانات مدونة الثروة اللفظية للمرحلة الابتدائية (طلابًا وطالبات)

تُعد هذه المدونة مصدرًا غنيًا لدراسة اللغة العربية لدى طلاب المرحلة الابتدائية، وقد جُمعت مادتها من سياقات تعليمية طبيعية على امتداد المملكة، وتمثّل تنوعًا في الأنواع النصية، والمستويات الصفية. وتشتمل المدونة على ما يزيد عن مليون ونصف المليون كلمة (١,٥٠٠,٠٠٠)، موزعة على مجموعة متنوعة من النصوص الشفوية والمكتوبة، وفقًا للتفصيل الآتي:

أولًا: بيانات الطالبات

- ٣,٦٣١ مادة منطوقة تم تسجيلها لطالبات المرحلة الابتدائية.
- ١,١٥٣ مادة مستكتبة تم جمعها عبر الإملاء بإشراف المعلمين.
- ١,٧٨٦ كراسة مدرسية، تشمل موضوعات إنشائية وأنشطة مدرسية مكتوبة بخط اليد.

ثانيًا: بيانات الطلاب

- ٢,٧٤٧ مادة منطوقة مسجلة لطلاب المرحلة الابتدائية.
- ٩٠١ مادة مستكتبة مكتوبة من قبل الطلاب.
- ٢,٤٠٩ كراسة مدرسية متضمنة لمواد كتابية متنوعة.

ثالثًا: المصادر التعليمية والنصوص المصاحبة

- ٨٠ كتابًا مدرسيًا معتمدًا للمرحلة الابتدائية صادرة في عام ٢٠٠٦م.
- ٢٢ عددًا من مجلة "الشبل" الموجهة للأطفال في المراحل الأولى من التعليم.
- مجموعة من القصص التربوية والتعليمية المخصصة لطلبة المرحلة الابتدائية.
- مواد تلفزيونية تعليمية منتجة لذات الفئة العمرية، ومتنوعة في أهدافها (لغوية، سلوكية، معرفية).

رابعًا: التغطية الجغرافية:

تغطي المدونة أكثر من ٣٦٠ مدرسة ابتدائية تمثل مختلف مناطق المملكة، ما يمنحها تنوعًا لغويًا وثقافيًا يعزز من قدرتها على دعم الدراسات المعجمية، والأسلوبية، واللغويات التطبيقية، وخصوصًا في تحليل اللغة المستخدمة لدى الفئات العمرية المبكرة.

٣-٥ بيانات مدونة الثروة اللفظية للمرحلتين المتوسطة والثانوية (طلابًا وطالبات)

تُعد هذه المدونة من أكبر المدونات التعليمية المتخصصة في تحليل اللغة المستخدمة لدى طلاب وطالبات المرحلتين المتوسطة والثانوية في المملكة العربية السعودية، وقد تم تطويرها لتكون أداة علمية فاعلة في دراسة الأداء اللغوي الكتابي للمتعلمين في هاتين المرحلتين، من خلال مصادر تعليمية أصلية ومتنوعة. وتحتوي المدونة على أكثر من مليوني كلمة (٢,٠٠٠,٠٠٠)، موزعة كما يلي:

أولًا: بيانات الطلاب والطالبات (نصوص مستكتبة)

- ١,٧٥٨ مادة مستكتبة تم جمعها من طالبات المرحلتين المتوسطة والثانوية، وتشمل موضوعات إنشائية، واختبارات، وتعبيرات ذاتية.
- ١,٤٧٦ مادة مستكتبة أُخذت من طلاب المرحلتين المتوسطة والثانوية، وتشمل نصوصًا تعليمية وتطبيقية متنوعة.

يوضح الجدول رقم (١) توزيع النصوص الشفوية والمكتوبة لطالبات المرحلة الابتدائية بحسب المناطق التعليمية في المملكة العربية السعودية

المنطقة	عدد المدارس (تقريبي)	نصوص منطوقة	نصوص مستكتبة	كراسات مدرسية	إجمالي النصوص
الباحة	٧	٣٥	٠	٣٥	٧٠
الجوف	٣	١٧	٠	١٧	٣٤
الحدود ش	٢	٩	٠	٩	١٨
الرياض	١٥	٧٥	٣٠	٧٥	١٨٠
الشرقية	٩	٥٠	٢٥	٥٠	١٢٥
القصيم	٨	٥٠	٢٠	٥٠	١٢٠
المدينة	٧	٤٥	٢٥	٤٥	١١٥
تبوك	٣	١٥	١٠	١٥	٤٠
جيزان	٩	٤٥	٢٥	٤٥	١١٥
حائل	٤	٢٠	٨	٢٠	٤٨
عسير	١٠	٥٠	٣٠	٥٠	١٣٠
مكة	٩	٤٥	٢٨	٤٥	١١٨
نجران	٥	٢٥	١٠	٢٥	٦٠

يوضح الجدول رقم (٢) توزيع النصوص الشفوية والمكتوبة لطلاب المرحلة الابتدائية بحسب المناطق التعليمية في المملكة العربية السعودية

المنطقة	عدد المدارس (تقريبي)	نصوص منطوقة	نصوص مستكتبة	كراسات مدرسية	إجمالي النصوص
الباحة	٧	٢٨	٠	٢٨	٥٦
الجوف	٣	١٥	٠	١٥	٣٠
الحدود ش	٢	٧	٠	٧	١٤
الرياض	١٥	٧٠	٢٥	٧٠	١٦٥
الشرقية	٩	٤٨	٢٠	٤٨	١١٦
القصيم	٨	٤٧	١٨	٤٧	١١٢
المدينة	٧	٤٣	٢٢	٤٣	١٠٨
تبوك	٣	١٣	٩	١٣	٣٥
جيزان	٩	٤٠	٢٢	٤٠	١٠٢
حائل	٤	١٨	٧	١٨	٤٣
عسير	١٠	٤٨	٢٧	٤٨	١٢٣
مكة	٩	٤٢	٢٦	٤٢	١١٠
نجران	٥	٢٣	٩	٢٣	٥٥

يوضح الجدول رقم (٣) توزيع النصوص المستكتبة للطلاب والطالبات في المدارس المتوسط والثانوية حسب المناطق التعليمية في المملكة العربية السعودية

المنطقة	عدد المدارس (تقريبي)	نصوص مستكتبة (طالبات)	نصوص مستكتبة (طلاب)	إجمالي النصوص
الرياض	٢٠	٢٥٠	٢٣٠	٤٨٠
الشرقية	١٢	١٨٠	١٧٥	٣٥٥
مكة	١٥	٢١٠	٢٠٠	٤١٠
القصيم	٨	١٣٠	١٢٥	٢٥٥
عسير	١٠	١٦٠	١٥٠	٣١٠
المدينة	٧	١٢٠	١١٠	٢٣٠
جيزان	٦	١٠٠	٩٥	١٩٥
نجران	٥	٩٠	٨٥	١٧٥

ثانيًا: المصادر التعليمية والمرافقة

- ١٦٦ كتابًا مدرسيًا معتمدًا في المرحلتين المتوسطة والثانوية، صادرة عن وزارة التعليم في عام ٢٠٠٦م، تغطي جميع المواد الدراسية ومجالاتها.
- ٣٥ عددًا من الصحف والمجلات التي تُعنى بلغة طلاب هذه المرحلة، تم اختيارها بعناية لتمثل الخطاب الإعلامي الموجه للشباب.
- مجموعة من القصص الموجهة لفئة المراهقين، بما يسهم في تحليل البنية السردية واللغوية في سياقات غير رسمية.
- مواد تلفزيونية تعليمية وتثقيفية موجهة للطلاب في هذه المرحلة، تتنوع في موضوعاتها وأشكالها التعبيرية.

ثالثًا: التغطية الجغرافية:

شملت عملية جمع البيانات أكثر من ١٨٠ مدرسة متوسطة وثانوية موزعة على مختلف مناطق المملكة، ما يضمن تمثيلًا لغويًا وثقافيًا واسع النطاق يعكس تنوع البيئات التعليمية والاجتماعية للمتعلمين.

٣-٦ منهجيات الجمع والتوثيق

اتُّبعت منهجيتان رئيسيتان في بناء المدونة:

الجمع الكتابي: بجمع النصوص من المقررات والمراجع والمقالات المنشورة.

الجمع الشفهي: من خلال تسجيل محادثات الطلاب، ثم تفريغها ومعالجتها إلكترونيًا.

وقد خضعت النصوص لعمليات تحرير ومراجعة لغوية مبدئية للحفاظ على الاتساق، دون المساس بالطابع الواقعي للنصوص الطلابية.

٣-٧ آليات الترميز والتصنيف

تم استخدام آليات ترميز مبدئية تعتمد على تصنيف النصوص وفق:

- مصدر النص (ديني، تعليمي، إعلامي، طلابي).
- جنس صاحب النص (طالب، طالبة).
- الفئة العمرية للمتكلمين (ابتدائي، متوسط، ثانوي).
- المنطقة الجغرافية (المنطقة، المدينة، القرية)
- المدارس (طلاب، طالبات)

وقد أُعدت هذه التصنيفات لتسهيل التحليل الإحصائي والدلالي لاحقًا، وجعل المدونة قابلة للمعالجة الآلية المتقدمة.

٣-٨ أدوات المعالجة الحاسوبية في مدونة الثروة اللفظية

تُعد مدونة الثروة اللفظية إحدى أبرز المدونات العربية التعليمية المحوسبة، وقد تم تطويرها ضمن مشروع "لسان"، وهي متاحة عبر الرابط: https://corpus.lisan1.com

يوضح الشكل (١) واجهة المستخدم في منصة مدونة الثروة اللفظية التعليمية

وتجدر الإشارة إلى أن الباحث هو من قام ببناء وتطوير الأدوات الحاسوبية الخاصة بتحليل المدونة، وذلك في إطار عمله على تيسير استنطاق محتوى المدونة، واستخلاص المؤشرات اللغوية والمعجمية منها، بما يخدم أهداف البحث العلمي في مجال تعليم اللغة العربية وتحليل أدائها لدى المتعلمين.

وتشمل الأدوات التي قام الباحث بتطويرها ما يلي:

- أداة استخراج قوائم الشيوع (Frequency Lists): لتوليد جداول بالكلمات الأكثر تكرارًا، سواء على مستوى المدونة كاملة أو ضمن فئات محددة.

يوضح الشكل (٢) قائمة الكلمات الأكثر شيوعًا في مدونة الثروة اللفظية: تحليل تكراري لأكثر المفردات استخدامًا

- أداة المكشاف السياقي (KWIC): لعرض الكلمات ضمن سياقاتها الأصلية، مما يتيح التحليل الدلالي والنحوي في مواضعها الواقعية.

يوضح الشكل (٣) أمثلة سياقية لاستخدام مفردة (الصلاة) في نصوص طالبات الصف الخامس – مدونة الثروة اللفظية

- أداة تحليل المتصاحبات اللفظية (Collocations): لاستكشاف العلاقات المعجمية بين الكلمات، والتعرف على التركيبات اللغوية الشائعة.
- أداة البحث بالجذر الصرفي: لاسترجاع جميع المشتقات الفعلية والاسمية المرتبطة بجذر معين، ما يُسهم في تحليل البنية الصرفية، والكشف عن أنماط الاستخدام.

وقد صُممت هذه الأدوات بلغة حاسوبية مخصصة لتتوافق مع طبيعة النصوص التعليمية وتنوعها، وتمثل إضافة علمية داعمة للباحثين في مجالات التحليل المعجمي، والتركيبي، والدلالي، والتربوي

٣-٩ خصائص مدونة الثروة اللفظية

أولاً: التمثيل اللغوي الواقعي للمتعلمين السعوديين

تمثل مدونة الثروة اللفظية أحد أبرز المشاريع العربية في توثيق اللغة كما تُستخدم واقعيًا في البيئات التعليمية، وليس كما يُفترض أن تكون. وتُعد هذه السمة من أبرز خصائص المدونات التعليمية، إذ تساعد الباحثين على تتبع الأنماط اللغوية الفعلية للمتعلمين، والكشف عن مواطن الصعوبة والأخطاء الشائعة في استخدام المفردات والتراكيب (Granger, 2002). وتتيح المدونة تحليل الأداء اللغوي للمتعلمين في السياق السعودي، مما يدعم تطوير مناهج تعليم اللغة العربية بناء على شواهد واقعية.

ثانيًا: التغطية الجغرافية والديموغرافية

تشمل المدونة نصوصًا من مختلف مناطق المملكة، جُمعت بمشاركة مئات المدارس والمعلمين، مما يجعلها مرآة للتنوع اللغوي والجغرافي. ويمثل هذا النوع من التنوع قيمة مضافة في الدراسات اللغوية، إذ يُتيح إمكانية دراسة الفروق بين الجنسين، والمستويات الدراسية، والمناطق الجغرافية في الأداء اللغوي، (Tagliamonte, 2006).

ثالثًا: التوازن بين الفصحى الكلاسيكية والفصحى الحديثة

على الرغم من أن المدونة تُبنى في إطار اللغة العربية الفصيحة، إلا أنها تعكس تفاوتًا في درجات الاستخدام بين الفصحى الكلاسيكية والفصحى الحديثة، والتي تُستخدم غالبًا في الكتابة الطلابية والإعلام التعليمي. ويُعد هذا التوازن أمرًا ضروريًا في المدونات التعليمية، لأنه يُحاكي بيئة التعلم الفعلية. (Selinker, 1972).

رابعًا: الخصائص الصوتية والدلالية والصرفية

رغم أن المدونة لم تُفهرس بعد بمستويات كاملة من التعليق الصرفي أو الصوتي، إلا أن وجود النصوص المنطوقة المُفرغة يفتح المجال أمام تحليل النطق، والبنية الصوتية. أما من الناحية الصرفية والدلالية، فتُظهر البيانات الأولية إمكانات كبيرة لتحليل الشيوع، الحقول المعجمية، الأنماط الصرفية، والتركيب اللغوي، مما يجعلها أساسًا جيدًا لمعالجة اللغة الطبيعية في السياق العربي (Jurafsky & Martin, 2023).

٣-١٠ مجالات الاستفادة من المدونة على المستوى التعليمي

أولاً: إعداد المواد التعليمية وفق مستويات لغوية واقعية

تُعد المدونات التعليمية مصدرًا رئيسًا لبناء محتوى لغوي حقيقي موجه للمتعلمين، بحيث تُنتقى المفردات والتراكيب بناءً على درجة شيوعها واستخدامها الفعلي في البيئات التعليمية، لا على أساس تصورات تقليدية للمحتوى. ويؤكد الباحثون في تعليم اللغات أن التعلم القائم على المدونات يُعزز من وعي المتعلم بخصائص اللغة الواقعية، ويساعد في بناء كفاءة استخدام حقيقية (Gilquin, Granger, & Paquot, 2007).

ثانيًا: بناء اختبارات الكفاية اللغوية

تمثل المدونات اللغوية مصدرًا موثوقًا لبناء اختبارات معيارية، خاصة إذا كانت المدونة تمثل بيئة المتعلمين المستهدفة. فمن خلال تحليل الشيوع والتنوع

اللغوي في المدونة، يمكن تحديد مستويات المفردات والتراكيب المناسبة لكل مستوى دراسي، وبناء مهام اختبارية واقعية (Alderson, 2000).

ثالثًا: تصميم أنشطة تعليمية في المفردات والتراكيب

توفر المدونة أمثلة طبيعية على استخدام المفردات في السياق، مما يتيح للمعلمين تطوير تمارين تستند إلى أمثلة واقعية لا مصطنعة. ويمكن استخدام برامج مثل AntConc[1] لاستخراج السياقات النصية، وتوليد تمارين فراغات، تكميل، مطابقة، وتصنيف (Flowerdew, 2012).

رابعًا: تحليل أخطاء الطلاب اللغوية

نظرًا لاحتواء مدونة الثروة اللفظية على نصوص طلابية حقيقية، فإنها تتيح فرصة ثمينة لتوثيق، وتصنيف أخطاء التعلم، سواء كانت نحوية أو معجمية أو دلالية، بما يمكّن من تطوير محتوى علاجي مخصص لكل فئة عمرية أو تعليمية (Granger, 2008).

٣-١١ مجالات الاستفادة من المدونة على المستوى البحثي

أولا: دراسات شيوع الكلمات

توفر المدونة إمكانية إجراء دراسات كمية على تكرار الكلمات، والتراكيب، واستخراج القوائم الشائعة الأكثر استخدامًا، وهو ما يساعد في إعداد مناهج تعليمية، ومعاجم مدرسية (Nation, 2001).

ثانيًا: تحليل التراكيب الأكثر استخدامًا

من خلال تحليل التوافقات السياقية، يمكن تحديد التركيبات النمطية الشائعة (مثل: "ذهب إلى المدرسة"، "أريد أن")، ودراستها من حيث التكرار

---

١ برنامج AntConc هو أداة مجانية ومفتوحة المصدر لتحليل المدونات النصية (corpus analysis)، طوّرها البروفيسور لورنس أنتوني (Laurence Anthony)، ويُستخدم على نطاق واسع في اللسانيات الحاسوبية وتحليل الخطاب وتعليم اللغات

والوظيفة التداولية، وهو مجال متطور في تحليل النحو التداولي (Hunston, 2002).

ثالثًا: تحليل اللغة حسب الجنس والمستوى والمنطقة

تُمكّن المدونة من إجراء دراسات لغوية اجتماعية، لتحليل أنماط اللغة بين الذكور والإناث، أو بين طلاب المناطق المختلفة، وهو ما يُعزّز من فهم الفروق الفردية والجماعية في تعلم اللغة (Biber & Gray, 2016).

رابعًا: التطبيقات الحاسوبية (معالجة اللغات الطبيعية والذكاء الاصطناعي)

بما أن المدونة تتضمن نصوصًا مفرّغة ومرمّزة جزئيًا، فهي تمثّل موردًا مهمًا لبناء أدوات لغوية حاسوبية، كأنظمة التصحيح الآلي، تحليل المشاعر، أو تصنيف النصوص التعليمية (Jurafsky & Martin, 2023).

٣-١٢ مجالات الاستفادة من المدونة على المستوى التقويمي والمؤسسي

أولا: رصد التغيرات اللغوية

تمكّن المدونة من تتبع تطور الاستعمال اللغوي للمتعلمين على مدى سنوات، ومقارنة أنماط اللغة في الفئات العمرية المختلفة، مما يتيح بناء خريطة لغوية زمنية لتطور الأداء (Leech et al., 2009).

ثانيًا: تحسين المناهج الدراسية

من خلال ربط بيانات المدونة بمخرجات التعليم، يمكن تطوير محتوى المناهج بحيث يكون أكثر توافقًا مع الواقع اللغوي للمتعلمين، كما هو مطبق في تجارب تعليم اللغات الأوروبية (Römer, 2008).

ثالثًا: تطوير برامج إعداد المعلمين

توفر المدونة محتوى واقعيًا يمكن استخدامه في تدريب المعلمين على تحليل أداء الطلاب، وتصميم المحتوى اللغوي القائم على بيانات، وتوظيف أدوات البحث اللغوي في غرفة الصف (Reppen, 2010).

٣-١٣ التطبيقات المقترحة للاستفادة من المدونة

أولا: أمثلة على استخدام المدونة في تحليل نصوص واقعية

تمثل مدونة الثروة اللفظية أداة مثالية لتقديم أنشطة لغوية تعتمد على بيانات واقعية، وهو توجه بات شائعًا في تعليم اللغات وفق ما يُعرف بـ التعليم القائم على البيانات. ويشير Johns (١٩٩١) إلى أن تمكين الطالب من "أن يصبح لغويًا صغيرًا" من خلال تفاعله المباشر مع المدونة يساعده في تنمية الوعي اللغوي والاكتشاف الذاتي للقواعد والاستخدامات.

ومن الأمثلة التطبيقية: استخدام المدونة لاستخراج مقاطع تحتوي على تراكيب لغوية معينة (مثل: الحال، الشرط، المفعول المطلق) وتوظيفها في أنشطة استقراء لغوي. وتحليل كتابات الطلاب في المدونة لتحديد أخطاء شائعة في استخدام أدوات الربط أو تصريف الأفعال، وتصميم أنشطة علاجية تستند إلى تلك الأخطاء (Granger et al., 2015). وإجراء تمارين مطابقة، وتكميل، وتصنيف مفردات ضمن حقول دلالية، استنادًا إلى أمثلة فعلية من المدونة.

ثانيًا: أدوات إلكترونية مبنية على المدونة

في ضوء تزايد الاهتمام بتقنيات التعليم المدعوم بالحاسوب، يمكن توظيف محتوى مدونة الثروة اللفظية في تطوير أدوات رقمية تفاعلية تستفيد من البيانات الموثقة داخلها. ومن بين هذه الأدوات:

ثالثًا: مولد تمارين تلقائي

يمكن تطوير أداة إلكترونية تتصل مباشرة بقاعدة بيانات المدونة، وتقوم تلقائيًا باستخراج مقاطع تحتوي على مفردات أو تراكيب معينة، وتحويلها إلى أسئلة تفاعلية، مثل أسئلة اختيار من متعدد، أو ملء الفراغات، أو أسئلة تصحيح.

رابعًا: أداة تحليل لغوي للمتعلمين

وهي أداة تسمح للمعلم أو الباحث برفع نص طالب معين، ليتم مقارنته بنصوص مماثلة في المدونة، وتقديم تقرير فوري عن مدى الشيوع، الأخطاء اللغوية المحتملة، مستوى المفردات، وبنية الجمل. (Gablasova et al., 2019).

خامسًا: منصات تعليمية تفاعلية قائمة على المدونات

مثل منصات SKELL, FLAX، و Lextutor، التي تسمح للطلاب باستكشاف الاستخدامات الحقيقية للكلمات من خلال جمل وسياقات حقيقية مستخلصة من مدونات، ويمكن برمجة واجهة مماثلة للغة العربية تعتمد على مدونة الثروة اللفظية (Römer, 2011).

سادسًا: دمج المدونة في أنظمة الذكاء الاصطناعي

تُعد المدونة مصدرًا غنيًا يمكن توظيفه في تدريب أنظمة تعلم الآلة الخاصة بتصحيح الأخطاء، أو تحليل الأسلوب، أو اقتراح بدائل لغوية، كما هو الحال في تطبيقات التدقيق اللغوي وتوليد النصوص (Jurafsky & Martin, 2023).

٣-١٤ التحديات التي تواجه مدونة الثروة اللفظية

على الرغم من القيمة الكبيرة التي تقدمها مدونة الثروة اللفظية، فإن هناك عددًا من التحديات التقنية والمنهجية التي قد تُعيق استثمارها الكامل في المجالات التعليمية والبحثية. ومن أبرز هذه التحديات:

أولا: نقص التحديث الدوري للبيانات

إذا لم تُحدّث المدونة بانتظام لإضافة نصوص جديدة تمثل التطور اللغوي، فإنها ستفقد تدريجيًا قيمتها التمثيلية. ويشير McEnery & Hardie (٢٠١٢) إلى أن المدونات الديناميكية أكثر فعالية في تتبع التغيرات اللغوية، ومناسبة لتحليل اللغة المعاصرة.

ثانيًا: مشكلات الترميز والتحليل النحوي والدلالي

تفتقر العديد من المدونات التعليمية إلى نظم ترميز موحدة؛ مما يصعب عمليات التحليل الآلي، ويُقلل من قابليتها للمعالجة الحاسوبية الدقيقة (Leech, 2004). ويُعد غياب الترميز عائقًا أمام توظيف المدونة في الذكاء الاصطناعي أو بناء تطبيقات تعتمد على البيانات اللغوية.

ثالثًا: الحاجة إلى توثيق الاستخدام والنتائج

نحن في حاجة إلى نشر تقارير دورية توضح طرق استخدام المدونة في بيئات التعليم والبحث، وتقديم دراسات حالة، تعزز من الثقة المجتمعية والعلمية في فاعليتها.

نتائج الدراسة

تُبرز هذه الدراسة القيمة العلمية والتربوية الهامة لمدونة الثروة اللفظية بوصفها أول مدونة تعليمية سعودية واسعة النطاق تُعنى بتوثيق الاستعمال الواقعي للغة العربية لدى طلاب التعليم العام. وقد كشفت النتائج أن المدونة تتميز بعدة خصائص تجعلها مؤهلة لتكون ركيزة معرفية ومنهجية في تطوير تعليم اللغة العربية، من أبرزها: تنوع مصادرها النصية، وشمولها للمستوى اللغوي الوسيط، واحتواؤها على إنتاجات لغوية واقعية من بيئات متعددة داخل المملكة.

وقد بيّنت التحليلات أن المدونة قادرة على دعم تطوير المحتوى التعليمي، والأنشطة الصفية التفاعلية، فضلًا عن دورها في تحليل الأخطاء اللغوية الشائعة، واكتشاف الحقول الدلالية المهيمنة في لغة المتعلم. كما أثبتت أنها تصلح كبنية بيانات قوية لدعم تطبيقات المعالجة الآلية للغة، وإثراء أدوات الذكاء الاصطناعي المهتمة باللغة العربية.

تمثل مدونة الثروة اللفظية نواة حقيقية لمشروع وطني يمكن أن يمتد إلى بناء منصة لغوية تعليمية شاملة توثق تطور اللغة العربية في المملكة، وتُسهم في نشر اللغة العربية الفصحى بأسلوب حديث يستند إلى بيانات واقعية، ويعكس اهتمامات المتعلم العربي، ويضع الأسس لتحوّل رقمي حقيقي في تعليم اللغة العربية وتقييمها.

التوصيات

استنادًا إلى ما سبق، تقترح هذه الدراسة جملة من التوصيات التي يمكن أن تُسهم في تطوير مدونة الثروة اللفظية وزيادة فاعليتها:

١. توفير دليل استخدام تفاعلي يوضح كيفية البحث والتحليل.
٢. تحديث المدونة دوريًا بإضافة نصوص جديدة من بيئات متنوعة، تمثل أنماط اللغة المتغيرة عند المتعلمين في المملكة.
٣. اعتماد أنظمة ترميز معيارية وفق النماذج العالمية مثل Universal POS Tags و W3C Linguistic Annotation Framework، مما يُسهم في تحسين دقة المعالجة.
٤. إنشاء منصة إلكترونية تعليمية تدمج بين المدونة وأدوات تعليم اللغة، على غرار أنظمة FLAX و SKELL، مما يُساعد في الانتقال من التحليل النظري إلى التطبيقات الصفية المباشرة.
٥. تعزيز التعاون بين الجامعات والمراكز اللغوية لاستخدام المدونات في مشاريع بحثية، وتقديم ورش تدريبية للمعلمين في كيفية توظيف المدونات في التعليم (Reppen, 2010).

المراجع العربية:

- الثبيتي، عبد المحسن بن عبيد، & الشمري، عقيل بن حامد. (٢٠٢٤). استخدام المدونات الحاسوبية في التحليل اللغوي: نماذج من لغة الصحافة العربية. مجلة المعرفة، ٢٠٧، ٨٩–١٣٧.
- عدة، بشير، عمامرة، كمال، & شعنبي، نور الدين. (٢٠٢٥). أهمية المدونات اللغوية في البحث الإحصائي اللغوي: مدونة اللغة العربية المعاصرة نموذجًا. مجلة جسور المعرفة، ١١(١)، ٥٢–٧٩.

المراجع الأجنبية:

- Alderson, J. C. (2000). Assessing Reading. Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34(2), 213–238.
- Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development, description and application. International Journal of Learner Corpus Research, 5(2), 126–157.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. Journal of English for Academic Purposes, 6(4), 319–335.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (pp. 3–33). John Benjamins.
- Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (Eds.), Corpus linguistics: An international handbook (Vol. 1, pp. 259–275). De Gruyter Mouton.

- Granger, S., Gilquin, G., & Meunier, F. (2015). The Cambridge Handbook of Learner Corpus Research. Cambridge University Press.
- Hunston, S. (2002). Colligation, lexis, pattern and text. In Pattern Grammar. John Benjamins.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. ELR Journal, 4, 1–16.
- Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed., draft). Stanford University. https://web.stanford.edu/~jurafsky/slp3
- Kennedy, G. (1998). An Introduction to Corpus Linguistics. Longman.
- Leech, G. (2004). Adding linguistic annotation. In Developing linguistic corpora: A guide to good practice (pp. 17–29). Oxbow Books.
- Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). Change in Contemporary English: A Grammatical Study. Cambridge University Press.
- McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge University Press.
- Meunier, F., & Reppen, R. (Eds.). (2015). Assessing learner performance in L2: Corpora and applications. Routledge.
- Nation, I. S. P. (2001). Learning Vocabulary in Another Language. Cambridge University Press.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), How to Use Corpora in Language Teaching (pp. 125–152). John Benjamins.
- Reppen, R. (2010). Using Corpora in the Language Classroom. Cambridge University Press.
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for

future developments. Zeitschrift für Anglistik und Amerikanistik, 54(2), 121–134.

- Römer, U. (2008). Corpora and language teaching. In A. Lüdeling & M. Kytö (Eds.), Corpus linguistics: An international handbook (Vol. 1, pp. 112–130). De Gruyter Mouton.
- Römer, U. (2011). Corpus research applications in second language teaching. Annual Review of Applied Linguistics, 31, 205–225.
- Schmitt, N., & Schmitt, D. (2020). Vocabulary in Language Teaching (2nd ed.). Cambridge University Press.
- Scott, M., & Tribble, C. (2006). Textual Patterns: Key Words and Corpus Analysis in Language Education. John Benjamins.
- Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics in Language Teaching, 10(1–4), 209–231.
- Tagliamonte, S. A. (2006). Analysing Sociolinguistic Variation. Cambridge University Press.
- Yusuf, Kamal. (2024). Leveraging corpus linguistics for effective Arabic vocabulary learning. Nazar: Journal of Arabic Language Teaching, Linguistics and Literature, 13(1), 347–351.
- Zaki, Mai, Wilmsen, David, & Abdulrahim, Dana. (2021). The utility of Arabic corpus linguistics. In Arabic Linguistics (Chapter 20). Cambridge University Press.