دراساتٌ لغويّة:

إسهامات المدونات النصية العربية في الدراسات اللغوية وتعليم اللغة العربية

The Contribution of Arabic Corpus in language studies and Arabic teaching

Sumbangan Korpus Bahasa Arab kepada kajian bahasa dan pengajaran bahasa Arab

حسلينا حسان*

خديجة خليلة عبد الرشيد **

ملخص البحث:

تهدف هذه الدراسة إلى الكشف عن إسهامات المدوَّنات النصيَّة في الدراسات اللغويَّة وبيان دورها في تعليم اللغة العربيَّة وتعلُّمها، ستقوم هذه الدراسة بتحليل مدوَّنتين نصيَّتين في ماليزيا، هما: مُدوَّنة الكتب العربيَّة المُقرَّرة والمُدونات النصيَّة العربيَّة للجامعة الإسلاميَّة العالميَّة بماليزيا؛ إذ تتمثَّل مدوَّنة الكتب العربيَّة في الكتب المُقرَّرة لمادّة اللغة العربيَّة بالمدارس الثانويَّة بماليزيا ابتداء من السنة الأولى حتى السنة الخامسة؛ أمَّا المدوَّنات الأخرى فتحتوي على النصوص الأكاديميَّة العربيَّة التابعة للجامعة الإسلاميَّة العالميَّة بماليزيا مثل الأطروحات الجامعيَّة، وأوراق المؤتمرات والمقالات المنشورة في المجلات المُحكَّمة. ونتج عن هذه الدراسة قوائم كلمات، وكشافات سياقيَّة اعتماداً على كلتا المُدوَّنتين سعياً إلى الكشف عن الكلمات الشائعة ووجوه استخدامها في السياقات المختلفة، وأوضحت الدراسة بأنَّه لا يمكن الاستغناء عن المُدوَّنات النصيَّة لمن لديهم اهتمام في تطوير لغتهم وتعبيرهم مثل الطلاب والمُعلِّمين ومؤلفي الكتب المقرَّرة، وكتب تمارين اللغة العربيَّة والمعاجم؛ حيث تُعتبر قائمة الكلمات مرجعاً أساسيّاً ومهمّاً في الدراسات اللغويَّة وبالأخصّ في تعليم اللغة وتعلُّمها.

* أستاذة مشاركة، قسم اللغة العربية وآدابها، كلية معارف الوحي والعلوم الإنسانية، الجامعة الإسلامية العالمية بماليزيا.
haslina.h@iium.edu.my

** طالبة دكتوراه، قسم اللغة العربية وآدابها، كلية معارف الوحي والعلوم الإنسانية، الجامعة الإسلامية العالمية بماليزيا.
khadijahkhalilah07@gmail.com

أرسل البحث بتاريخ: ٢٠١٩/٣/١٢م، وقبل بتاريخ: ٢٠١٩/٩/٢٢م.

الكلمات المفتاحية: المدوَّنات النصيَّة–قائمة الكلمات–الكشافات السياقيَّة–تكرار الكلمات–الكتب العربية المقررة.

Abstract:

This study aims to discover the contribution of Arabic corpus in language studies, and to explain it's role in learning and teaching Arabic language. This study analyses two Malaysian corpora which are Arabic Textbooks Corpus and Arabic Corpus of International Islamic University Malaysia. The first corpus is on Arabic textbooks used in Malaysian secondary schools by the first until the fifth years of the secondary schools. The second corpus contains academic texts for example from theses, conference papers and journal articles. The result of this study shows the word list and concordance for both corpuses. This study proves that the corpus is important to develop the language among learners, teachers, writers of textbooks, exercise books and dictionaries. The word list is regarded as the basic reference in language study as well as in learning and teaching of language.

Keywords: Corpus, word list, concordance, word frequency, arabic textbooks

Abstrak:

Kajian ini bertujuan mengenal pasti sumbangan korpus berbahasa Arab dalam pengajian bahasa, di samping menghuraikan peranannya dalam pembelajaran dan pengajaran bahasa Arab. Kajian ini menganalisis dua korpus dari Malaysia iaitu Korpus Buku Teks Bahasa Arab dan Korpus Bahasa Arab dari Universiti Islam Antarabangsa Malaysia. Sumber korpus yang pertama daripada buku teks bahasa Arab bagi tingkatan satu hingga tingkatan lima yang digunakan di sekolah menengah di Malaysia. Manakala korpus yang kedua mengandungi teks akademik seperti tesis, kertas persidangan dan artikel jurnal. Hasil bagi kajian ini ditunjukkan dalam bentuk senarai kata dan konkordans bagi kedua-dua korpus. Kajian ini membuktikan bahawa korpus penting untuk pengajian bahasa dalam kalagan pelajar, tenaga pengajar, serta penulis buku teks, buku latihan dan kamus terutamanya hasil daripada senarai kata yang menjadi rujukan asas dalam kajian bahasa, serta pembelajaran dan pengajaran bahasa.

Kata kunci: Kekerapan kata kerja, data korpus-buku teks-kekerapan, senarai kata.

مقدمة:

إنَّ لسانيَّات المدوَّنات النصيَّة من مجالات اللسانيَّات الحاسوبيَّة، وتدخل في مجال اللسانيَّات التطبيقيَّة، فالمدوّنات النصيَّة ليست معجماً إلكترونياً بل هي مجموعة من النصوص المكتوبة أو المنطوقة التي يمكن تحليلها أو تخزينها حاسوبيّاً ولها أهميَّة في تعليم اللغة وتعلُّمها. وتتشعَّب الاستفادة من المدوَّنات النصيَّة في عدَّة مجالات من الدراسات اللغويَّة مثل النحو، والصرف، وعلم الأصوات، وعلم المعجم وعلم الدلالة.

تختلف مزايا المدوَّنات النصيَّة للغة العربيَّة بين الماضي والحاضر، وبطبيعة الحال تُبنى المدوَّنات في الحاضر بمساعدة الحاسوب والتقنية الحديثة مما لم يكن متاحاً في الماضي. بدأت عمليَّة التدوين للغة العربيَّة في القرن الأول والثاني للهجرة؛ حيث كانت جهودهم حينئذ بمثابة أول ظهور لفكرة المدوّنات في الماضي،[١] فمن المصادر الأساسيَّة للمدوّنات النصيَّة في الماضي نصوص القرآن الكريم والنصوص الشعريَّة. لم تتضح مصادر المُدوَّنات النصيَّة للغة العربيَّة في الحاضر، وقد تنوَّعت محتوياتها بين نصوص من اللغة المكتوبة مثل الكتب، والمجلّات، والصحف، ونصوص أخرى مكتوبة في المواقع الإلكترونيَّة، وبين بيانات من اللغة المنطوقة مثل الكلام في الحوارات، والمناقشات والمؤتمرات، أو الكلام المنقول عبر التقنيات الحديثة مثل الهاتف، المذياع، والتلفاز.[٢]

من خلال هذه الدراسة يظهر أنّ من أسباب معاناة معلمي اللغة العربيّة ومتعلِّميها في الاستفادة الكاملة من التقنيات الحديثة التي تستعين بالتكنولوجيا- والتي تتمثل بالدراسات اللغويّة الحاسوبيَّة- هو عدم توفُّر المدوّنات النصيَّة بشكل شامل يحقِّق أهداف الدراسات اللغويُّة في عمليَّة التعلّم والتعليم، ومن أجل الإسهام في تقديم حلٍّ لهذه المشكلة، تسعى الدراسة لتحقيق هدفين رئيسين هما: الأول الكشف عن إسهامات المدوَّنات النصيَّة في الدراسات اللغويَّة، والثاني وبيان دور المدوَّنات النصيَّة في تعليم اللغة العربيَّة وتعلُّمها.

وتكمن أهميَّة الدراسة في أنَّها تُسهم في تقديم اقتراحات للاستفادة من إسهامات المدوَّنات النصيَّة، بهدف تعليم اللغة العربيَّة وتعلُّمها؛ إذ تُنتِج المُدوَّنات قائمة من الكلمات الشائعة مع تكرارها وكذلك الكشافات السياقيَّة، وتفيد المدوَّنات النصيَّة محبي اللغة العربيَّة من المتعلِّمين، والمعلِّمين، ومؤلفي الكتب، ومصممي أوراق الامتحانات، وتحدِّد الدراسة مدوَّنة الكتب العربيَّة المُقرَّرة والمدوَّنات النصيَّة العربيَّة التابعة للجامعة الإسلاميَّة العالميَّة بماليزيا، بوصفهما مواد مستعملة للدراسة وذلك لعرض إسهامات المدوَّنات النصيَّة للغة العربيَّة في الدراسات اللغويَّة.

وبالنسبة إلى نتائج الدراسة، فتختار الدراسة تطبيقين اثنين فقط وهما قائمة الكلمات والكشافات السياقيَّة في كشف مدى فائدة استخدام المدوَّنات النصيَّة في دراسة اللغة العربيَّة. يُقصد بقائمة الكلمات مجموعة الكلمات المرتَّبة حسب شيوعها، وتعرف الكشافات السياقيَّة بأنّها بناء نصوص المدوّنة اللغويّة لتظهر في صورة منظَّمة، يسهل التعامل معها آليًا، سواء على مستوى المفردات، أم على مستوى الجمل والتراكيب.[٣]

أولاً: تعريف المدونات النصية وأهميتها

جاءت كلمة "مدوَّنة" في اللغة العربيَّة من الجذر "دُونَ"، ويدل على معنى نقيض لـ "فوق" ويكون ظرفاً، وبمعنى أمام، ووراء، وضد، وبمعنى غير.[٤] ولا يشتقّ منه فعل، وبعض النحاة يقول منه دانَ يدُونُ دَوناً، وأُدِينَ إدَانةً، ويروي قول عدّي: "أنسل الذرعان غرب حذم### وعلا الربرب أزم لم يُدَنْ"، وغيره يرويه: لم يُدنّ على ما لم يسم فاعله، من دَنّى يُدَنِّي، أي: ضعف.[٥]

إنَّ أقرب كلمة واردة لمدوَّنة اصطلاحاً هي "ديوان" وهي كلمة فارسيَّة معرَّبة، ولم ترد كلمة "مدوَّنة" بالضبط في المعاجم العربيّة مثل **لسان العرب**، والصحاح، و**معجم النفائس الوسيط**، و**القاموس المحيط**، و**الكليَّات**. والديوان أصله "دوَّان"، فعوض من إحدى الواوين ياء؛ لأنَّه يجمع على دواوين ولو كانت الياء أصليَّة لقالوا "دياوين"، وقد دوَّنت الدواوين.[٦] قال ابن بري وحكى ابن دريد وابن جني أنّه يقال "دياوين" وفي الحديث لا يجمعهم ديوان حافظ، قال ابن الأثير هو الدفتر الذي تكتب فيه أسماء الجيش وأهل العطاء، وأول من دَوَّنَ الديوان هو عمر رضي الله عنه.[٧] ويمكن القول بأنَّ مصطلح "دَوَّن الكتب" بمعنى جمعها؛ لأنّ جمع الأشياء إدناء بعضها من بعض، كما يرد معنى "الديوان" في معجم النفائس الوسيط بإنشاء الكتب، وجمعها، وترتيبها.[٨]

ترد كلمة "مدوَّنة" في **معجم الغني الزاهر** نصاً مدوّناً في وثيقة.[٩] ويبيّن قاموس المصطلحات اللغويَّة والأدبيَّة كلمة "مدوَّنة" كـ"مجموعة محدودة من الأقوال المحكيّة أو المكتوبة تكوّن مادة تخضع للدراسة اللغويّة والوصف اللساني".[١٠]

وبناء على ما تقدَّم، لم يظهر فرق كبير بين تعريف مصطلح "مدوَّنة" باللغة العربيَّ؛. أما تعريفها في غير العربيَّة، فكما تعرِّف سلوى حمادة كلمة "corpus" باللغة الإنجليزيَّة بأنّها مُشتقَّة من كلمة لاتينيَّة بمعنى "الجسد"، وهي عبارة عن مجموعة من النصوص المتاحة للاستخدام الحاسوبي.[١١] ويرادف مصطلح "مدوَّنة" بالعربيَّة مصطلح "korpus" في اللغة الملايويَّة كما يشير قاموس ديوان إلى تعريفه بوصفه مجموعة من النصوص (نصوص مكتوبة) تتناول موضوعات معيَّنة أو بيانات لأجل دراسات بحثيَّة.[١٢]

١. المدونات النصية للغة العربية

تؤدي المدوَّنات النصيَّة أدواراً مهمة في الدراسة اللغويّة والدراسة الحاسوبيّة،[١٣] وتهتمّ هذه الدراسة بأهميّة المدوَّنات النصيَّة في الدراسة اللغويَّة. إنَّ مجال هذه المدوّنات النصيّة لا يعدُّ مجالاً جديداً في الدراسات العربيّة، بل من الملاحظ توفّر عدّة دراسات متعلّقة بالدراسة الحاسوبيّة ودراسة المدوَّنات النصيَّة للغة العربيَّة.

وتُطبِّق دراسة ليلي حنيفاريزان تقنية من تقنيات المدوَّنات النصيَّة في الدراسة الصرفيَّة؛ حيث تعرض هذه الدراسة نتائج التحليل الصرفي الحاسوبي من النصوص المبثوثة على موقع محيط "شبكة الإعلام العربيّة"، وتقتصر الدراسة على صفحات الأخبار الثقافيّة، والاقتصاديّة، والرياضيّة. وتقوم هذه الدراسة بتحليلها تحليلاً صرفياً إلى ما يعرف باللواصق، وتقسّمها إلى نوعين وهما: السوابق واللواحق للكلمات الواردة في النصوص المتوفّرة. وتتمّ دراستها مستعينة ببرنامج "فصيح" وهو برنامج خاص لتحليل النصوص العربيّة وإنتاج مدوّنات نصيّة مهيكلة ومعنونة بمعلومات التحليل الصرفي العربي بشكل آلي تماماً. أشارت نتائج الدراسة إلى أنّ هناك ١٢٩ سابقة مقسّمة إلى ١١ سابقة مفردة، و٤٤ سابقة مركّبة ثنائيّة، و٥٤ سابقة مركّبة ثلاثيّة، ٢٠ سابقة مركّبة رباعيّة؛ أما اللواحق فهناك ٢١٤ لاحقة مقسّمة إلى ٢٤ لاحقة مفردة، و٩٨ لاحقة مركّبة ثنائيّة، و١٠٤ لاحقة مركّبة ثلاثيّة، و١٥ لاحقة مركّبة رباعيّة.[١٤]

تستعين دراسة زين الرجال بتحليل تكرار الكلمات الواردة في قسم الشؤون الدوليّة وقسم الرياضة في وسائل الإعلام العربيّة: الصحف العربيّة نموذجاً، عن طريق برنامج حاسوبي وردسميث ٥،٠ (Wordsmith 5.0) لإعداد قائمة كلمات مستمدّة من المدوّنات النصيّة للصحف العربيَّة. وتهدف هذه الدراسة إلى كشف أسلوب اللغة المستخدمة في الصحف العربيَّة، ويتم ذلك بتحليل تكراريَّة الكلمات الواردة في قسم الشؤون الدوليّة وقسم الرياضة في الصحف المعتمدة، وتتكوّن هذه القائمة من ثلاثين مقالة مأخوذة من سبع صحف عربيّة من مختلف البلدان، ويقوم الباحث بتحليل تكراريّة الكلمات لكل قسمين حسب الأنواع الثلاثة الرئيسة وهي: الأسماء، والأفعال، والحروف اعتماداً على الصحف بفروعها المختلفة.[١٥]

وبالنسبة إلى دراسة الأفعال العربيّة التي لها علاقة بالمدوّنات النصيّة، توصّلت نور عاشقين في دراستها إلى أنّ الأفعال الماضية هي الأكثر تكراراً في قائمة الأفعال الشائعة بحسب الأزمنة الثلاثة الواردة في مدوّنة (Arabic World News)، فهذه المدوّنة تحتوي على كثير من الجمل الماضية، ولم تستخدم أفعال الأمر في النصوص الإخباريَّة.[١٦]

أما دراسة ماشطة فتهدف إلى الكشف عن الكلمات الشائعة المتوفرة في المدونة النصية الأكاديمية العربية التي تسمى بالمنقب العربي، وكذلك عرض تلك الكلمات إلى أقسامها الثلاثة وهي اسم، وفعل،

وحرف. إن هذه المدونة النصية الأكاديمية لها اسهامات في معاونة باحثي اللغة العربية على استخدام الكلمات في تكوين العبارات والجمل الصحيحة من خلال ملاحظة توظيف هذه الكلمات في سياقاتها المختلفة، فضلاً عن مساعدة الطلبة الذين يدرسون العربية لأغراض أكاديمية على الرجوع إلى الألفاظ الشائعة في مجال تخصصهم.[١٧]

وبناء على ما تقدم، تفيد المدوّنات النصيّة المتعلّمين والمعلّمين لجميع اللغات، وعلى سبيل المثال، يرشد المدرّس طلابه إلى مدى ضرورة معرفة الكلمات الشائعة في تعليم اللغة العربيّة، وذلك عن طريق الرجوع إلى قائمة الكلمات الشائعة مع الإشارة إلى عدد التكرار الوارد في نصوص المدوّنة. فحينها سيتمكّن الطلبة من التركيز على دراسة الكلمات الشائعة في اللغة العربيّة، ولا يقع اهتمامهم كثيراً على الكلمات غير الشائعة والتي تستعمل بشكل نادر، لأنّهم لن يستعينوا بها في الاتصال اللغوي مع متكلمي اللغة العربيّة.

هناك عدّة أمثلة للمدوّنات العربيّة التي تحتوي على بيانات مختلفة، وهي كالآتي مع بيانها بإيجاز:

أ. الذخائر اللغويّة: مدوّنة مكوّنة من ثلاثة أرباع مليون كلمة عربيّة متوازنة المصادر لصالح مشروع نملار بواسطة أداتها فصيح؛ حيث إنّ هذه المدوّنة قد احتوت على جميع مجالات الدراسات اللغوية أي الصرفيّة، والنحويّة، والصوتيّة، والدلاليّة المعجميّة. وقد تمّت مراجعة عناوين هذه المدوّنة يدويًّا بالكامل، وهي نتيجة مجهودات الباحثين من الشركة الهنديّة لتطوير النظم الرقميّة بمصر، ولمزيد من التفاصيل، يمكن مراجعة وثيقة المواصفات لمدونة الذخائر اللغويّة.[١٨]

ب. المدونة العربية: تحتوي هذه المدوّنة العربيّة على قرابة ١٧٣ مليون كلمة، وتشمل النصوص المكتوبة باللغة العربيّة على الجرائد، والروايات، والكتب المختارة في اللغويّات والأدبيّات، والآيات القرآنية، وهي عبارة عن جهود مبذولة من المحاضر ديلورت باركنسن بجامعة برنجام يانج[١٩] في الولايات المتحدة، ومن مزايا المدوّنة سهولة البحث فيها، ويتاح البحث بالألفاظ المستقلّة أو بالتراكيب متعدّدة الألفاظ كاسم شخص "صلاح الدين الأيوبي"، أو كلام مركّب "الشعر الحر"، أو اسم خاص "حركة الإخوان المسلمين" دون اللجوء إلى استخدام محدّدات البحث.

ج. قاعدة بيانات صوتيّة باللغة العربيّة الفصحى المنطوقة في الجزائر (ALGASD) :

إنّ أحد أهداف هذه المدوّنة الصوتيّة هو تمثيل اللهجات الإقليميّة للغة العربيّة الفصحى الحديثة المتداولة في الجزائر، كما تحتوي على معلومات مفيدة عن المتكلّمين مثل العمر والجنس ومستوى التعليم، وعيّنتُها تتكون من ٣٠٠ متكلم جزائري تم اختيارهم من ١١ منطقة في الجزائر، وهي غير متوفّرة في الموقع الإلكتروني. وهذه القاعدة البيانيّة هي من عمل عدّة باحثين من دولة الجزائر وكندا مثل حمداني، وبدرة،

وسلواني، وقد تمّ دراستها في سنة ٢٠٠٩م.[٢٠] إنّ هذا المجهود يمكّن الباحثين من استخدامه أساساً لأنظمة الاستكشاف الآلي للكلام المخصّص باللغة العربيّة.

ثانياً: منهجية الدراسة

تعتمد هذه الدراسة على مدوّنتين نصيّتين بوصفهما مادّتين للدراسة، أوّلهما مدوّنة الكتب العربيّة المقرّرة وثانيهما المدوّنات النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا.

فمدونة الكتب العربيّة المقرّرة هي مجموعة من الكتب المقرّرة لمادّة اللغة العربيّة من السنة الدراسيّة الأولى حتى السنة الخامسة والمستخدمة في المدارس الثانويّة الحكوميّة بماليزيا في سنة ٢٠١٢م، والتي تمّ تجميعها وإعدادها بالاستعانة ببرنامج "سكيش انجين" وهو برنامج حاسوبي متوفّر على شبكة الإنترنت، وتحتوي هذه المدوّنة على ١٨٠،٠٠٠ كلمة.

أما المدوّنة النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا، فهي مدوّنة أكاديميّة عربيّة تحتوي على واحد وأربعين مليون كلمة، وتتضمن مجموعة من النصوص الأكاديميّة المكتوبة باللغة العربيّة مثل الرسائل الأكاديميّة، وأوراق أعمال المؤتمرات، والأوراق المنشورة في المجلات المحكّمة. وقد تأسّست هذه المدوّنة في سنة ٢٠١٠م على يد محاضِرَين من قسم اللغة العربيّة وآدابها في الجامعة الإسلاميّة العالميّة بماليزيا، وهما الدكتورة حسلينا حسّان والدكتور محمد فهام محمد غالب.

تقوم الدراسة بتحليل حاسوبي للمدوّنتين السابقتين عن طريق تطبيقين اثنين وهما قائمة الكلمات مع تكرار شيوعها والكشّافات السياقيّة، ويتمّ ذلك بمساعدة برنامج "سكيش انجين" وبرنامج "المنقب العربي". يقدّم البرنامج الأول قائمة الكلمات الشائعة لمدوّنة الكتب العربيّة المقرّرة مع بيان عدد تكرارها، وستشتمل القائمة على جميع الكلمات الواردة في الكتب المقرّرة لمادة اللغة العربيّة للمرحلة الثانويّة بماليزيا، كما ستتضمن قائمة الكلمات ثلاث قوائم مقسّمة بحسب أقسام الكلمة الثلاثة وهي قائمة الأسماء، وقائمة الأفعال وقائمة الحروف؛ أما البرنامج الثاني فسيعرض الكشافات السياقيّة وكذلك قائمة الكلمات الشائعة الخاصّة بالمدوّنات النصيّة العربيّة التابعة للجامعة الإسلاميّة العالميّة بماليزيا والتي تدخل فيها جميع نصوص المدوّنة المكتوبة باللغة العربيّة.

نتائج الدراسة ومناقشتها

تناقش هذه الدراسة عاملين من العوامل التي تدفع الدارسين لاستعمال المدوّنات في تعليم اللغة العربيّة وتعلُّمها:

أ) قائمة الكلمات

المدوّنات النصيّة والدراسة المعجميّة

ليست المدوّنات النصيّة معجماً إلكترونًا في أصلها، بل هي مادّة مهمّة لإعداد المعاجم والقواميس وتشكّل المدوّنات النصيّة فائدة مهمّة في الدراسة المعجميّة، ويمكن أيضاً أن يستفاد من المدونات النصيّة الخاصة بالمعاجم والقواميس في مجالات مختلفة كبياناتها؛ حيث تتكون المعاجم والقواميس من ملايين الكلمات، ويمكنها أن تفيد مستخدمي اللغة في البحث عن المفردات وإعداد التعبيرات الخاصة بهم في مجالهم، كما وتُسهم المدوّنات النصيّة في الدراسة المعجميّة للغة العربيّة؛ إذ تسهم في إعداد مجموعات من الكلمات مأخوذة من عدّة مصادر بيانات بطريقة سريعة وحديثة؛ ما ينتفع به المهتمّون باللغة العربيّة في أداء وإنجاز بعض الأغراض الدراسيّة.

الصورة (١): قائمة الكلمات الشائعة الواردة في المدوّنات النصيّة العربية للجامعة الإسلاميّة العالميّة بماليزيا مع عدد تكرارها

قائمة الكلمات
عدد الكلمات : 270,889

إذهب إلى الصفحة: 1 من 2709

الكلمة	عدد المرات	الكلمة	عدد المرات
قال	5224	في	118902
غير	5204	من	81990
بعض	5138	على	46519
–	5124	أن	39698
إذا	4832	إلى	35240
يكون	4755	،	22530
فيها	4717	عن	21763
أي	4652	التي	20430
بعد	4614	ما	19803
تعالى	4548	لا	18053
وفي	4527	أو	18021
فإن	4478	هذا	17005
أنه	4466	الله	14030
بها	4330	هذه	13874
دار	4297	بين	13161
الإسلام	4093	الذي	12693
وهي	4088	ذلك	11829

تعرض صورة (١) قائمة الكلمات المُستقاة من المدوّنات النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا، وتبيّن القائمة عدد مرات تكرار الكلمات، حيث تحتوي على ٢٧٠،٨٨٩ كلمة غير مكرّرة. إنّ عدد تكرار الكلمات في هذه القائمة له دلالة إحصائيّة تفيد معرفة مدى كثرة الاستعمال لكل الكلمات المأخوذة من النصوص الأكاديميّة في هذه المدوّنة، وكلمة "في" هي الأكثر استعمالاً؛ حيث تمّ إحصاء مرات تكرارها بـ ١١٨،٩٠٢ مرّة، وكلمة "من" تكررت ٨١،٩٩٠ مرّة، وكلمة "على" أُحصيت بـ ٤٦،٥١٩ مرّة.

من خلال مدوّنة الكتب العربيّة المقرّرة يتمّ إعداد قائمة الكلمات مع عدد مرات تكرارها وتتكوّن القائمة من جميع أقسام الكلمة الأسماء والأفعال والحروف، وكذلك يتمّ إحصاء الأرقام، وعلامات الترقيم مثل النقط والفواصل. ومن أجل إعداد قائمة تشمل الأقسام الثلاثة، فتحذف الحروف والأفعال لإعداد قائمة الأسماء، وتحذف الأسماء والحروف لإعداد قائمة الأفعال، وتحذف الأسماء والأفعال لإعداد قائمة الحروف.[٢١] وتعرض الجداول الآتية ثلاث قوائم مقسّمة بحسب أقسام الكلام الثلاثة.

الجدول (١): الأسماء الواردة في قائمة الكلمات من مدوّنة الكتب العربيّة المُقرّرة

الرقم الترتيبي	الكلمة	التكرار
١	الآتية	٦٣٨
٢	ما	٦٠١
٣	ماليزيا	٤٧٧
٤	التي	٤١٤
٥	هذا	٣٧٤
٦	الله	٣٤١
٧	الكلمات	٣١٩
٨	مائة	٣٠٠
٩	المثال	٢٩٣
١٠	بين	٢٦٢

يظهر الجدول (١) بعض الأسماء الواردة في قائمة الأسماء المُستمدّة من مدوّنة الكتب العربيّة المقرّرة، ويظهر أنّ أكثرها تكراراً هو اسم "الآتية"؛ حيث تكرر ٦٣٨ مرّة، ثم يليه اسم "ما" وتكرر ٦٠١ مرّة، ويليهما اسم "ماليزيا"؛ حيث تكرر ٤٧٧ مرّة. إنّ هذه القائمة ستساعد المعلمين في الكشف عن الكلمات المشهورة وتقديمها للطلبة، وكذلك ستتيح للمعلمين القيام بأنشطة لغويّة خاصة باللغة العربيّة لاكتساب المفردات العربيّة الشائعة .

الجدول (٢): الأفعال الواردة في قائمة الكلمات المُستقاة من مدوّنة الكتب العربيّة المقرّرة

الرقم الترتيبي	الكلمة	التكرار
١	اقرأ	٢٤٤
٢	استمع	١٨٠
٣	هات	١٧٠

٤	كان	١٦٥
٥	اكتب	١٥٥
٦	املأ	١٤٩
٧	أجب	١٤٦
٨	يأتي	١٤٥
٩	قال	١٣١
١٠	اذكر	١٠٣

يعرض جدول (٢) بعض الأفعال الواردة في قائمة الأفعال المُستمدّة من مدوّنة الكتب العربيّة المقرّرة، ويظهر أنّ أكثرها تكراراً هو الفعل "اقرأ"؛ حيث ورد ٢٤٤ مرّة، والفعل "استمع" تكرّر وروده ١٨٠ مرة، والفعل "هات" تكرر وروده ١٧٠ مرّة. ويمكن للطلبة الذين يتعلّمون اللغة العربيّة أن ينتفعوا من هذه القائمة في بناء الجمل العربيّة المفيدة والصحيحة المكوّنة من الأفعال التي يكثر تكرارها في كتابات المتخصّصين، وذلك من أجل ممارسة الأفعال الشائعة والتركيز عليها دون غيرها.

الجدول (٣): الحروف الواردة في قائمة الكلمات المُستقاة من مدوّنة الكتب العربيّة المقرّرة

الرقم الترتيبي	الكلمة	التكرار
١	في	٣٥٨٧
٢	من	٢١٤٩
٣	إلى	١٠٨٢
٤	على	١٠١٠
٥	أن	٥٩٨
٦	عن	٥٢٥
٧	مع	٤٤٦
٨	ثم	٣٦٩
٩	لا	٣٥٦
١٠	يا	٢٥٥

يُقدِّم جدول (٣) بعض الحروف الواردة في قائمة الحروف للكتب العربيّة المُقرّرة، والأكثر تكراراً من بينها هو الحرف "في" فنجده قد تكرّر ٣٥٨٧ مرّة، وحرف "من" تكرّر ٢١٤٩ مرّة، وحرف "إلى" تكرّر ١٠٨٢ مرّة.

المدوّنات النصيّة والدراسة الصرفيّة

تفيد قائمة الأفعال من قائمة الكلمات متعلمي اللغة العربية ومعلّميها في إعداد القائمة الخاصة بالدراسة الصرفيّة. تعرض قائمة الأفعال تحليلاً لكل الأفعال الواردة فيها على عدّة أقسام، فعلى سبيل المثال من حيث الزمن إلى ماض، ومضارع، وأمر، ومن حيث التجريد والزيادة، ومن حيث نوع الحروف التي يتكون منها الفعل إما صحيحاً أو معتلاً؛ حيث قسّم علماء اللغة العربيّة الحروف إلى حروف صحيحة وحروف علّة، فالألف والواو والياء هي حروف علّة، ويستفيد الطلبة من هذا التقسيم لفهم بنية الكلمة العربيّة بشكل صحيح وفعّال. إنّ قائمة الأفعال العربيّة تظهر أهميّتها في تيسير تصريف الأفعال العربيّة لغير المتخصّصين في اللغة العربيّة من الطلبة والمبتدئين، ليكون التصريف العربي بواسطتها أكثر وضوحا وسهولة.

الجدول (٤): قائمة الأفعال الواردة في قائمة الكلمات المستمدّة من نصوص مدوّنة الكتب العربيّة المقرّرة

الرقم	الأفعال	نوعيّة الأفعال (مضارع، ماض، أمر)	نوعيّة حروف الأفعال (صحيح، معتل)	نوعية الأفعال (مزيد، مجرّد)	أوزان الأفعال	التكرار
١	اقرأ	أمر	صحيح	مجرد	افْعَل	٢٤٤
٢	استمع	أمر	صحيح	مزيد	افْتعِل	١٨٠
٣	كان	ماض	معتل	مجرد	فعَلَ	١٦٥
٤	اكتب	أمر	صحيح	مجرد	أُفعُل	١٥٥
٥	املأ	أمر	صحيح	مجرد	افعَل	١٤٩
٦	أجب	أمر	معتل	مزيد	أفِل	١٤٦
٧	يأتي	مضارع	معتل	مجرد	يفعِل	١٤٥
٨	قال	ماض	معتل	مجرد	فعَلَ	١٣١
٩	اذكر	أمر	صحيح	مجرد	أُفعُل	١٠٣
١٠	يكون	مضارع	معتل	مجرد	يَفعُل	٩٣

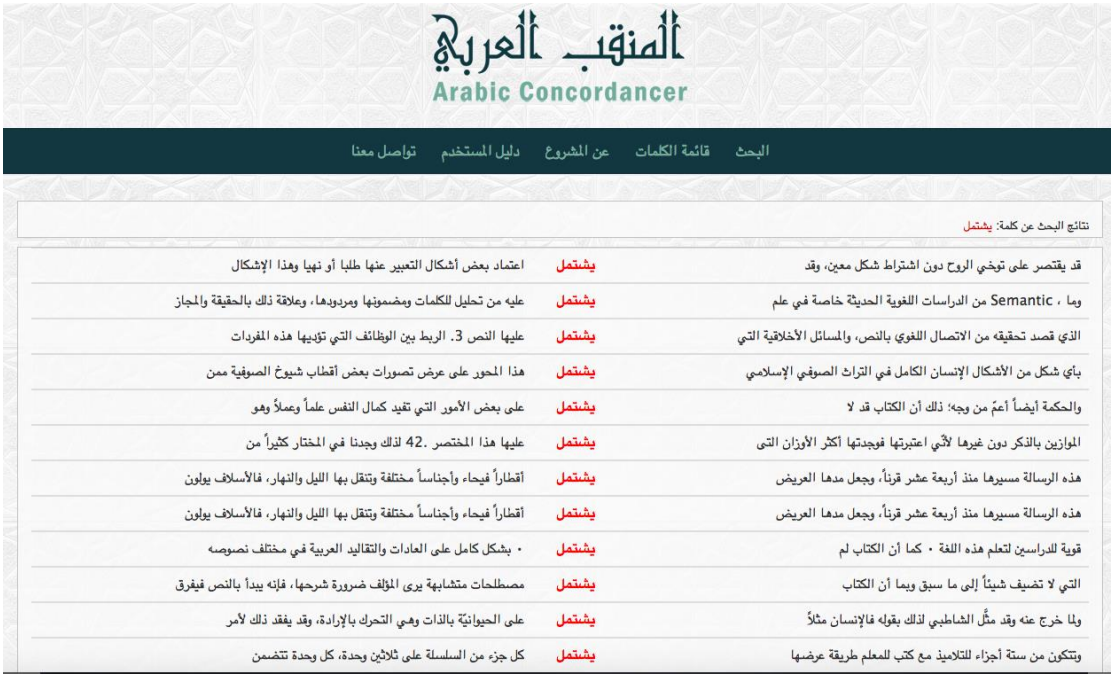
يعرض جدول (٤) بعض الأفعال الواردة في قائمة الكلمات المستمدّة من مدوّنة الكتب العربيّة المقرّرة مع عدد تكرارها، وتبيّن الدراسة نوعيّة الأفعال لجميع الأفعال الواردة في نصوص المدوّنة، مثل الفعل "اقرأ" فعل أمر، وهو فعل صحيح مجرّد ويبنى على وزن "افعَل."

ب) الكشافات السياقيّة

المدوّنات النصيّة والدراسة النحويّة والتركيبيّة

تؤدي المدوّنات أدواراً كبيرة في خدمة الدراسات النحويّة والتركيبيّة؛[٢٢] لأنّ جميع اللغات لها نظامها وقواعدها التي تحتاج إلى عدّة مصادر ومراجع في تعليمها وتعلُّمها. ويمكن أن يُكتَسب النظام النحوي للغة العربيّة عند المتعلّمين والمعلّمين بطريقة حديثة وسهلة باستخدام الكشافات السياقيّة المستمدّة من المدوّنات .

الصورة (٢): كيفيّة استخدام كلمة "يشتمل" في الكشّافات السياقيّة المستمدّة من المدوّنات النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا



تعرض صورة (٢) نتيجة البحث عن استخدام كلمة "يشتمل" في الكشافات السياقيّة من المدوّنات النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا، وتقدّم نتيجة البحث أمثلة لجمل نموذجيّة من النصوص الأكاديميّة التي تمّ استعمال الكلمة فيها. ويستفيد متعلمو اللغة العربيّة من هذه الخدمة خصوصاً في المرحلة الجامعيّة، وذلك لممارسة الكتابة الأكاديميّة بشكل صحيح وسليم.

المدوّنات النصيّة والمهارات اللغوية

إنّ التعبير هو القدرة على الإفصاح عن المشاعر والآراء والأحاسيس بوضوح وتسلسل وبلغة سليمة بهدف إيصالها إلى الآخرين، فإذا أتقن الطلبة فنّ التعبير وحلّ الأخطاء الإملائيّة، فسيمكّنهم ذلك من نقل أفكارهم وعواطفهم ومشاعرهم، ونقل وجهات نظرهم إلى غيرهم، والإبانة عما في نفوسهم سواء بطريقة

المشافهة أو الكتابة،[٢٣] وتساعد المدوّنات على حلّ مشكلات الإملاء في اللغة، وكذلك على تحسين تعبيرات اللغة حيث يواجه الطلبة غالباً في عمليّة التعلّم مشكلات في الإملاء، والكتابة، والتعبير، وتقدّم الكشافات السياقيّة حلولاً لها.

الصورة (٣): كيفيّة استخدام كلمة "ادخرت" في الكشافات السياقيّة المستمدّة من المدوّنات النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا

توضّح صورة (٣) فعلاً غير شائع وهو "ادخرت"، فهذا الفعل غير شائع في اللغة العربيّة ولا يستخدم كثيراً في الكتابة العاديّة ولا في المحادثات اليوميّة، وبالنظر إلى عدد تكراره في قائمة الكلمات من هذه المدوّنة نجد أنّه قد تكرّر عشرة مرات فقط في المدوّنات النصيّة العربيّة للجامعة الإسلاميّة العالميّة بماليزيا. وقد قدّمت الكشافات السياقيّة من هذه المدوّنة أمثلة لتعبيرات الجمل المستخدمة لهذا الفعل حيث تساعد متعلمي اللغة العربيّة في إنجاز جمل مفيدة وصحيحة باستخدام الكلمات غير الشائعة.

ومن جانب آخر، تساعد الكشافات السياقية في بناء الثروة اللغوية اعتمادا على المدونات النصية التي تحتوي على عدد كبير من الكلمات أي تعد الثروة اللغوية السياقات التركيبية للكلمات المتوفرة في قائمتها كما تعرض دراسة ماشطة قيمة الكلمات الملتبسة بالثروة اللغوية في المدونة النصية الأكاديمية.[٢٤] تأخذ هذه الدراسة مثالا من دراستها لكلمة "إذ" التي تأتي بوجهين؛ اسما وحرفا كما في الجدول الآتي:

الجدول (٥): الثروة اللغوية لكلمة "إذ" الواردة في المدونة النصية الأكاديمية المنقب العربي

"إذ" الاسمية	"إذ" الحرفية
وأهمية هذا التمييز تظهر حالياً في واقعنا الحالي **إذ** انتشرت الكتابة تعلماً	ومثله مثل أي قضية فرعية في متائل الفقه، **إذ** إنه لا تكاد توجد قضية في الفقه ليس حولها خلاف

وممارسة.	بين العلماء.

الخاتمة:

ومن خلال ما تقدّم تتّضح النتائج على النحو الآتي:

١. إنّ دراسة المدوّنات النصيّة الإلكترونيّة محور مهم في مجال اللسانيّات الحاسوبيّة، وتفاعل بين اللغة والحاسوب، وهي مهمّة للدراسات اللغويّة كالدراسات النحويّة، والصرفيّة، والمعجميّة، وتبرزها في إطار جديد باستخدام البرامج الحاسوبيّة.

٢. تستخدم المدوّنات النصيّة في تعليم اللغة العربيّة وتعلّمها وبالأخص في الدراسات اللغويّة.

٣. خلصت نتائج المدوّنات النصيّة إلى تطبيقين أساسيين وهما: الكشافات السياقيّة وقائمة الكلمات مع عدد مرات تكرارها.

هوامش البحث:

[1] Ditters, Everhard, *Arabic corpus linguistics in past and present in Studies in the History of Arabic Grammar II*, Amsterdam: John Benjamins Publishing Company, 1990, p. 130.

[٢] عبد الرشيد، خديجة خليلة، **الكلمات الواردة في الكتب المقرَّرة لمادة اللغة العربيَّة للمدارس الثانويَّة بماليزيا: دراسة لغويَّة حاسوبيَّة**، رسالة ماجستير غير منشورة، (كوالالمبور: الجامعة الإسلامية العالمية بماليزيا، ٢٠١٤م)، ص١٧.

[3] Sinclair, John, *Corpus, Concordance, Collocation*, Oxford: Oxford University Press, 1991, p. 32.

[٤] الفيروز آبادي، مجد الدين محمد بن يعقوب، **القاموس المحيط**، ط٥، (بيروت: مؤسسة الرسالة، ١٩٩٦م)، ص١٠٥٤.

[٥] الجوهري، إسماعيل بن حماد، **الصحاح: تاج اللغة وصحاح العربيَّة**، ط١، (القاهرة: دار الكتاب العربي، ١٩٩٩م)، ج٥، ص٢١١٥.

[٦] المرجع السابق، ج٥، ص٢١١٧.

[٧] ابن منظور، جمال الدين محمد بن مكرم، **لسان العرب**، ط١، (بيروت: دار صادر، ١٩٩٠م)، ج١٣، ص١٦٦.

[٨] أبو حاقة، أحمد، **معجم النفائس الوسيط**، ط١، (بيروت: دار النفائس، ٢٠٠٧م)، ص٤٠٣.

[٩] أبي العزم، عبد الغني، **معجم الغني الزاهر**، موقع إلكتروني: http://www.almaany.com.

[١٠] يعقوب، اميل وبركة، بسّام وشيخاني، مي، **قاموس المصطلحات اللغويّة والأدبيّة: عربي–إنكليزي–فرنسي**، ط١، (بيروت: دار العلم للملايين، ١٩٨٧م)، ص٣٥١.

[١١] حمادة، سلوى، **المدوّنات النصيّة ودور اللغة العربيّة في التعامل معها**، موقع إلكتروني: www.thomala.com.

[١٢] الموقع الإلكتروني: http://prpm.dbp.gov.my.

[13] McEnery, Tony and Wilson, Andrew, *Corpus Linguistics*, Edinburgh: Edinburgh University Press, 1996, p. 87.

[١٤] حسب الله، ليلي حنيفاريزن، **التحليل الصرفي في مجال الإعلام المكتوب باللغة العربيّة بين النظريّة والتطبيق: دراسة لغويّة حاسوبيّة**، (رسالة ماجستير غير منشورة، الجامعة الإسلامية العالمية بماليزيا، ٢٠١٣م).

[15] Abdul Razak, Zainur Rijal, *Modern Media Arabic: A Study of Word Frequency Analysis in World Affairs and Sports Sections in Arabic Newspaper*, PhD's Dissertation, University of Birmingham, 2011, p. 58.

[١٦] ياشيم، نور عاشقين، **الأفعال الشائعة المقترنة بالحرف في وكالة أنباء برناما بالعربيّة: دراسة لغويّة حاسوبيّة في المدوّنات الإلكترونيّة**، رسالة ماجستر غير منشورة، (كوالالمبور: الجامعة الإسلاميّة العالميّة بماليزيا، ٢٠١٣م)، ص٨٠-٨٢.

[١٧] عبد العزيز، ماشطة، **الكلمات الشائعة في المدونات النصية الأكاديمية الخاصة بالنصوص العربية في الجامعة الإسلامية العالمية بماليزيا: برنامج المنقب العربي نموذجا**، (رسالة ماجستير غير منشورة، كوالالمبور: الجامعة الإسلامية العالمية بماليزيا، ٢٠١٥م).

[18] Attia, Mohammed, A. Rashwan, Mohsen A., and Al-Badrashiny, Mohamed A. S. A. A., Fassieh®, *A Semi-Automatic Visual Interactive Tool dor Morphological, PoS-Tags, Phonetic, and Semantic Annotation of Arabic Text Corpora*, IEEE Transactions on Audio, Speech and Language Processing, 2009, vol. 17, no. (5).

[19] Brigham Young

[20] Hamdani, Ghania-Droua, Selouani, Sid Ahmed and Boudraa, Malika, *Algerian Arabic Speech Database (ALGASD): Corpus Design And Automatic Speech Recognition Application*, The Arabian Journal of Science and Engineering, vol. 5, no. 2C, Saudi Arabia: King Fahd University of Petroleum and Minerals, December 2010), p. 157-159.

[٢١] عبد الرشيد، خديجة خليلة، **الكلمات الشائعة في المدونات النصية الأكاديمية الخاصة بالنصوص العربية في الجامعة الإسلامية العالمية بماليزيا: برنامج المنقب العربي نموذجا**، ص٥٥.

[22] McEnery, Tony and Wilson, Andrew, *Corpus Linguistics*, p. 93.

[٢٣] الخولي، أحمد عبد الكريم، **التعبير الكتابي وأساليب تدريسه**، ط١، (عمان: دار الفلاح للنشر والتوزيع، ٢٠٠٤م)، ص١٨.

[٢٤] عبد العزيز، ماشطة، **الكلمات الشائعة في المدونات النصية الأكاديمية الخاصة بالنصوص العربية في الجامعة الإسلامية العالمية بماليزيا: برنامج المنقب العربي نموذجا**، ص٣٢.

References المراجع

‘Alī, Nabīl, *al-Lughah al-‘arabiyyah Wa al-Hāsūb*, (without city: MaTābi‘ al-KhaT, 1988).

Abd Aziz, Masitah, *al-Kalimāt al-Sha’i‘ah Fī al-Mudawwanāt al-Akādimiyyah al-Khāssah Bi al-Nusūs al-‘Arabiyyah Fī al-Jāmi‘ah al-Islāmiyyah al-‘ālamiyyah Bi Mālīziya: Barnāmij al-Munaqqib al-‘arabiy Namūzajan*, unpublished master thesis, (Kuala Lumpur: International Islamic University Malaysia, 2015).

Abdul Rashid, Khadijah Khalilah, *al-Kalimāt al-Wāridah Fī al-Kutub al-Muqarrarah Li Mādah al-Lughah al-‘Arabiyyah Li al-Madāris al-Thānawiyyah Bi Mālīziyā: Dirāsah Lughawiyyah Hāsūbiyyah*, unpublished master thesis, (Kuala Lumpur: International Islamic University Malaysia, 2014).

Abdul Razak, Zainur Rijal, *Modern Media Arabic: A Study of Word Frequency Analysis in World Affairs and Sports Sections in Arabic Newspaper*, unpublished PhD’s dissertation, (Birmingham, University of Birmingham, 2011).

Al-Khawlī, Ahmad ‘Abd al-Karīm, *al-Ta‘bīr al-Kitābī Wa Asālīb Tadrisuh*, 1st Edition, (Jordan: Dār al-Falāh Li al-Nashr Wa al-Tawzī‘, 2004).

Attia, Mohammed, A. Rashwan, Mohsen A., and al-Badrashiny, Mohamed A. S. A. A., Fassieh®, “a Semi-Automatic Visual Interactive Tool dor Morphological, PoS-Tags, Phonetic, and Semantic Annotation of Arabic Text Corpora”. *IEEE Transactions on Audio, Speech and Language Processing*, vol. (17), no. (5), 2009.

Ditters, Everhard, *Arabic Corpus Linguistics in Past and Present in Studies* in the History of Arabic Grammar II, ed. Kees Versteegh and Michael G. Carter (Amsterdam, John Benjamins Publishing Company, 1990).

Flowerdew, Lynne, “Applying Corpus Linguistics to Pedagogy: A Critical Evaluation”, *International Journal of Corpus Linguistics*, The Netherlands: John Benjamin Publishing Company, vol. (14), no. (3), 2009.

Hamādah, Salwā, *al-Mu‘ālajah al-āliyah Li al-Lughah al-‘Arabiyyah: al-Mashākil Wa al-Hulūl*, (Cairo: Dār Gharīb, 2009).

Hasbullah, Lily Hanefarezan, *al-Tahlīl al-Sarfī fī Majāl al-I‘lām al-Maktūb bi al-Lughah al-‘Arabiyyah Bayna al-Nazariyyah Wa al-Tatbīq: Dirasah Lughawiyyah Hāsubiyyah*, unpublished master thesis, (Kuala Lumpur: International Islamic University Malaysia, 2013).

Hassan, Haslina and Mohd Ghalib, Mohd Feham, “Mashru‘ Jam‘ al-Mudawwanāt al-Nassiyyah al-Khāssah Bi al-Nusūs al-Akādimiyyah Fi al-Lughah al-‘Arabiyyah”, *Majallah Majma‘ al-Lughah al-‘Arabiyyah al-Urdunī*, al-‘adad (85), 2013.

McEnery, Tony and Wilson, Andrew, *Corpus Linguistics*, (Edinburgh, Edinburgh University Press, 1996).

O'Keeffe, Anne; McCarthy, Michael and Carter, Ronald, *From Corpus to Classroom*, (Cambridge, Cambridge University Press, 2007).

Sinclair, John, *Corpus, Concordance, Collocation*, (Oxford, Oxford University Press, 1991).

Tribble, Chris and Jones, Glyn, *Concordances in the Classroom*, (England, Longman, 1990).

Yasyim, Norasyikin, *al-Afʿāl al-Shāʾiʿah al-Muqtaranah Bi al-Harf fi Wakālah Anbāʾ Bernama Bi al-ʿarabiyyah: Dirāsah Lughawiyyah Hāsubiyyah*, unpublished master thesis, (Kuala Lumpur: International Islamic University Malaysia, 2013).