



كلية الآداب
قسم اللغة العربية وآدابها



جامعة عين شمس

المَدُونَاتُ الحَاسُوبِيَّةُ لِأَرْصَدَةِ اللُّغَةِ العَرَبِيَّةِ دراسة وصفية تحليلية

"أُطْرُوحَةُ مُقَدِّمَةِ الحُصُولِ عَلَى دَرَجَةِ المَاجِسْتِيرِ فِي عِلْمِ اللُّغَةِ"

إعداد الباحث

محمد رمضان علي الغليص

إشراف

الأستاذ الدكتور

محسن عبد الرازق رشوان

أستاذ هندسة الإلكترونيات
والاتصالات

كلية الهندسة- جامعة القاهرة

الأستاذ الدكتور

علي محمد هندراوي

الأستاذ المتفرغ بقسم اللغة العربية
وآدابها

كلية الآداب- جامعة عين شمس

١٤٣٩هـ / ٢٠١٨م

إهداء

أبي العظيم وأمي الغالية إخواني وأخواتي، زوجتي
الحبيبة والأميرات، وإلى أرواح الشهداء،،،
إليكم جميعاً أهدي هذه الثمرة - نرجوها طيبة -
تعبيراً عن عرفان بالجميل، وشفيعي في ذلك نية
حسنة أرادت الخير والإحسان.

شُكْرٌ وَتَقْدِيرٌ

قال رسول الله ﷺ

"مَنْ لَا يَشْكُرُ النَّاسَ لَا يَشْكُرُ اللَّهَ"

صدق رسول الله ﷺ

أتقدّم بخالص الشكر والتقدير إلى أستاذي الكريمين:

الأستاذ الدكتور/ علي محمد هندawi

الذي غمرني بكرمه وحسن توجيهه بقبوله الإشراف على هذه الرسالة؛ فله خالص الشكر وعظيم التقدير.

والأستاذ الدكتور/ محسن رشوان

الذي أشرف على الجانب الحاسوبي من الرسالة، وأفاض على صاحبها بالمساندة العملية والعلمية؛ فله جزيل الشكر والود.

كما أتقدّم بالشكر الجزيل إلى أستاذي الكريمين:

الأستاذ الدكتور/ حسن محمود نصر السيد

و

الأستاذ الدكتور/ خالد فهمي إبراهيم

على تفضلهما بقبول مناقشة هذه الرسالة، وتحمل مشقة قراءتها وتفحص ما فيها لمناقشتها وتصويب ما فيها.

وأني لأتقدم بخالص الشكر للأستاذ الدكتور/ فرانتشيك أوندراش

والسيدة الدكتورة/ حسنة عبد السميع

الذين تعهدا صاحب الرسالة في مهده الأكاديمي وتبني فكرة الدراسة في مهدها؛ فلهما أجزل الشكر وأوفاه.

كما أتقدم بالشكر الجزيل للأستاذ الدكتور/ بيتر زيمانك على ما بذل من وقته الثمين في توجيه صاحب الرسالة.

كما أتوجه بخالص الشكر والامتنان والتقدير

للأستاذ الدكتور/ سامح الأنصاري

والأستاذ الدكتور/ سلطان بن ناصر المجبول

والأستاذ الدكتور/ عبدالمحسن الثبيتي

على مساعدتهم الإيجابية والفعالة في تعزيز مدونة الصوامت العربية

فهرس الموضوعات

الصفحة	الموضوعات
١	المقدمة.
٣	طبيعة الدراسة.
٣	أهمية الدراسة.
٤	مشكلة الدراسة.
٤	أهداف الدراسة.
٤	منهج الدراسة.
٤	الدراسات السابقة.
٦	حدود الدراسة.
٧	هيكل الرسالة.
٩	الفصل الأول: أنواع المدونات اللغوية الحاسوبية.
٩١	الفصل الثاني: دراسة نظرية لبعض نماذج توظيف التقنيات الحديثة في إنتاج بعض التطبيقات اللغوية المنبثقة من المدونات الحاسوبية.
١٠٩	الفصل الثالث: بناء مدونة للصوامت الصعبة في اللغة العربية.
١٥٦	الخاتمة.
١٥٧	١. النتائج.
١٥٩	- جوانب القصور.
١٦٠	٢. التوصيات.
١٦١	قائمة الملاحق.
١٧٦	قائمة المراجع.
١٧٧	أولاً- المراجع العربية.
١٨٢	ثانياً- المراجع الأجنبية.
١٨٤	ثالثاً- المواقع الإلكترونية والبرمجيات.
١٨٨	ABSTRACT

فهرس الجداول

رقم الجدول	عنوان الجدول	الصفحة
١-١	مصادر مدونة كلارا CLARA	٦٩
١-٢	اختبارات قياس الكفاءة اللغوية في مجال اللغة العربية للناطقين بغيرها.	٩٧
٢-٢	المقياس المتدرج العام لاختبار اللسان العربي (اسع).	١٠٤
٣-٢	التوصيفات التي وضعها واضعوا الإطار المرجعي الأوربي المشترك للمهارات.	١٠٦
٤-٢	ارتباط درجات امتحان (اسع) بالإطار الأوروبي المرجعي الموحد للغات من واقع درجات اختبار (IELTS).	١٠٧
١-٣	بيان الأصوات الصامتة.	١١٠
٢-٣	بيان أكثر الألسنة من حيث عدد الناطقين بها غير اللغة العربية.	١١٢
٣-٣	البلاد التي تتحدث تلك اللغة غالبًا.	١١٣
٤-٣	الصوامت العربية ولغة الماندرين الصينية.	١١٥
٥-٣	صوامت لغة الماندرين الصينية بالمقارنة بالصوامت العربية.	١١٦
٦-٣	الصوامت العربية واللغة الأسبانية.	١١٧
٧-٣	صوامت اللغة الأسبانية بالمقارنة بالصوامت العربية.	١١٧
٨-٣	الصوامت العربية واللغة الإنجليزية.	١١٨
٩-٣	صوامت اللغة الإنجليزية بالمقارنة بالصوامت العربية.	١١٩
١٠-٤	الفروق الصوتية بين اللغتين العربية والتركية.	١٢٢
١١-٣	صوامت الألفبائية الصوتية العربية الدولية.	١٢٣
١٢-٣	صوامت الألفبائية الدولية.	١٢٤
١٣-٣	مقارنة بين رموز IPA للغات المختلفة واللغة الإنجليزية.	١٢٧
١٤-٣	الأصوات الصامتة والحركات كما ذكرها د. كمال بشر في كتاب "علم الأصوات".	١٢٨
١٥-٣	صفات الصوامت العربية (١).	١٤٣
	صفات الصوامت العربية (٢).	١٤٤

فهرس الأشكال

رقم الشكل	عنوان الشكل	الصفحة
١-١	مثالاً لخطأ شائع من "معجم لونغمان للاختبارات".	١٧
٢-١	تحليل الأخطاء اللغوية لدارسي اللغة العربية - طلبة كوريين - .	١٧
٣-١	أربعة مصادر أساسية للمدونة اللغوية العربية العالمية.	٢٠
٤-١	فئات المدونة اللغوية العربية العالمية الإحدى عشرة.	٢٠
٥-١	مصادر المدونة من داخل وخارج الوطن العربي.	٢١
٦-١	صفحة تسجيل الدخول للمدونة اللغوية العربية العالمية.	٢٢
٧-١	نموذج من نتائج استخدام محركات بحث المدونة.	٢٤
٨-١	واجهة موقع مدونة فيينا للغة العربية المنطوقة.	٣١
٩-١	واجهة صفحة بحث موقع مدونة فيينا للغة العربية المنطوقة.	٣١
١٠-١	نموذج من نتائج استخدام محركات بحث المدونة.	٣٢
١١-١	نموذج من شروطاً مبتكرة لبناء الجملة القرآنية باستخدام الرسوم البيانية.	٣٦
١٢-١	ترجمة رموز الوظائف النحوية في مخطط وصف المدونة العربية للغة القرآن الكريم.	٣٧
١٣-١	صفحة التسجيل في موقع المدونة العربية للغة القرآن الكريم.	٣٨
١٤-١	واجهة صفحة البحث في موقع المدونة العربية للغة القرآن الكريم.	٣٩
١٥-١	نموذج من شاشة لشرح الكلمة على مستوى قواعد اللغة العربية نحواً وصرفاً، وكذلك الترجمة الإنجليزية للآيات من موقع المدونة العربية للغة القرآن الكريم.	٤١
١٦-١	صفحة تسجيل الدخول للمدونة اللغوية لمتعلمي اللغة العربية.	٤٥
١٧-١	نموذج من نتائج استخدام محركات بحث المدونة اللغوية لمتعلمي اللغة العربية.	٤٧
١٨-١	صفحة تسجيل موقع الكوربوس العربي.	٥١
١٩-١	نموذج من نتائج استخدام محركات بحث موقع الكوربوس العربي.	٥٣
٢٠-١	واجهة صفحة البحث في موقع المدونة اللغوية العربية لمدينة الملك عبدالعزيز.	٦٤
٢١-١	نموذج من نتائج استخدام محركات بحث المدونة.	٦٥
٢٢-١	صفحة تسجيل موقع المدونات المتوازية لوثائق الأمم المتحدة.	٧٢
٢٣-١	نموذج من نتائج استخدام محركات بحث مدونة نظام الأمم المتحدة للمعلومات الببليوغرافية.	٧٥
٢٤-١	نموذج من نتائج استخدام محركات بحث المدونة اللغوية المتوازية عربي - إنجليزي.	٨١

٨٦	نموذج من نتائج استخدام محركات بحث المدونات النصية العربية للجامعة الإسلامية العالمية بماليزيا (المُنْقَب العربي).	٢٥-١
١٠٧	ارتباط درجات امتحان (IELTS) بالإطار المرجعي الموحد للغات.	١-٢
١٢٤	صوائت الألفبائية الدولية.	١-٣
١٢٥	حركة اللسان مع الصوائت الأساسية.	٢-٣
١٢٨	الحركات العربية (الصوائت) والحركات المعيارية.	٣-٣
١٢٩	بيان الصوامت العربية من حيث طريقة النطق.	٤-٣
١٤٧	أداة خافض اللسان.	٥-٣

مَسْرَدُ المَصْطَلَحَات

Arabic Lingua Test (ALT)	اختبار اللسان العربي "اسع"
International Phonetic Alphabet (IPA)	الألفبائية الصوتية الدولية
Arabic Lexicography	الصناعة المعجمية العربية
Concordancer	الكشاف السياقي
Unique Words	الكلمات الفريدة/غير المكررة
Computational Linguistics	اللسانيات الحاسوبية
Spoken language Corpus	المدونات اللغوية المنطوقة
Corpus	المدونة اللغوية/متن/مكتز
part-of-speech tagging	توسيم نحوي
Language Computing	حوسبة اللغة
Phonetics	علم الأصوات
Phoneme	فونيم "وحدة صوتية"
Lists of common words	قوائم الكلمات الشائعة
Corpus Linguistics	لسانيات المدونة
Search Engine	محرك البحث
Parallel corpus	مدونة متوازية
Frequency Information	معلومات التكرار
Linguistic Data Consortium	مؤسسة البيانات اللغوية
Buckwalter transliteration system	نظام باك والتر للنقل الكتابي
DT transliteration system	نظام دي تي للنقل الكتابي
DARPA	وكالة مشاريع البحوث المتطورة الدفاعية التابعة لوزارة الدفاع الأمريكية

المقدمة

لا يخفى على أحد الإمكانيات المتعددة التي أتاحها الحاسب الآلي والتي ولدت الرغبة في تسخير علومه لخدمة نشاطات عدة، من أهمها درس اللغوي؛ إذ يعد دخول الحاسب الآلي في مجال اللغة بما عُرف بـ"حوسبة اللغة" خطوة كبيرة في مجالات اللغات الإنسانية بصفة عامة، ولاسيما اللغة العربية ومجالاتها، عن طريق المعالجة الحاسوبية للأرصدة اللغوية المكتوبة والمنطوقة؛ لتصبح مدونة حاسوبية يمكن التعامل معها، ومن ثم إنتاج العديد من التطبيقات التي تخدم اللغة وتقرّبها إلى الناطقين بها وبغيرها.

وقد اهتمت جهات عدة بجمع هذه الأرصدة اللغوية، وحشدت لها الجهود والفرق البحثية للعمل عليها والإفادة منها في الخروج بتطبيقات لغوية مثلت طفرة كبيرة في صناعة القواميس، وتعليم اللغات للناطقين بغيرها، والبرامج التي تخدم البحث الأكاديمي وتيسر مادته على الباحثين في التخصصات المتعددة مُدَلِّلة لهم الحدود الجغرافية وما تُمثِّلُه من عوائق وصعوبات.

ومن تلك الجهات داخل مصر والعالم العربي، على سبيل المثال:

١- الجامعة الأمريكية بالقاهرة ومساهمتها في مبادرة "أرشيف المجموعات العربية على شبكة المعلومات العالمية"^(١) عام ٢٠١٥م.

٢- "المدونات النصية المتوازية من الإنجليزية للعربية" التي قامت برصدها جامعة الكويت عام ٢٠٠٣م^(٢).

٣- مشروع "الرصيد اللغوي الوظيفي" الذي تقرر إنجازه عن طريق بلدان ثلاثة هي تونس والجزائر والمغرب عام ١٩٦٧م. وانتهى العمل في ١٩٧٢م، وأقر وزراء تعليم هذه الدول إدراجه في التعليم ابتداء من عام ١٩٧٥م^(٣).

٤- مشروع بناء "المعجم الكبير"، الذي تبناه مجمع اللغة العربية بالقاهرة، صدرت منه -حتى الآن- سبعة أجزاء بين عامي ١٩٥٦ و ٢٠١٢م، لم يكتمل بعد^(٤).

كما عمل على تكوين مثل هذه الأرصدة الإلكترونية للغة العربية من خارج العالم العربي فرق عمل بحثية وأقسام للغة العربية، على سبيل المثال:

(1) <http://aucegypt.edu/ar/news/stories/-العربية-المجموعات-أرشيف-إلى-مبادرة-أرشيف-المجموعات-العربية-> على الإنترنت

(2) E-A Parallel Corpus (2003), University of Kuwait, Publications from Kuwait National Council.

(٣) "الرصيد اللغوي للطفل العربي وأهمية الاهتمام بمدى استجابته لحاجاته في العصر الحاضر، الممارسات اللغوية" عبد الرحمن الحاج صالح، ٢٠١٠م، (١)، ٩-٢١.

(٤) "مدونة معجم تاريخي للغة العربية، معالجة لغوية حاسوبية"، أطروحة مقدمة للحصول على درجة الدكتوراه في علم اللغة، المعتز بالله السعيد، ٢٠١٠م.